

Στατιστική Μοντελοποίηση Δεδομένων με την
Ασύμμετρη Π-σιγμοειδή Κατανομή

Η Μεταπτυχιακή Εργασία Εξειδίκευσης

υποβάλλεται στην ορισθείσα

από τη Γενική Συνέλευση Ειδικής Σύνθεσης
του Τμήματος Μηχανικών Η/Υ και Πληροφορικής
Εξεταστική Επιτροπή

από τον

Ευστάθιο Μπαλίκα

ως μέρος των υποχρεώσεων για την απόκτηση του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ

ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ

Πανεπιστήμιο Ιωαννίνων

Σεπτέμβριος 2017

ΑΦΙΕΡΩΣΗ

Σε όλους όσους πάντα εμπνέονται και προσπαθούν, να απολαύσουν το ταξίδι κρατώντας στο νου τον προορισμό.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή κ. Αριστείδη Λύκα για την ουσιαστική βοήθεια και την υπομονή του, κατά την διάρκεια εκπόνησης της μεταπτυχιακής μου εργασίας. Επίσης θα ήθελα να ευχαριστήσω όλους τους συναδέλφους μου με τους οποίους συνεργάστηκα στα πλαίσια του μεταπτυχιακού προγράμματος, καθώς και την οικογένεια μου για την αμέριστη υποστήριξη που επέδειξε σε όλη την διάρκεια της φοίτησης μου στο Πανεπιστήμιο Ιωαννίνων.

ΠΕΡΙΕΧΟΜΕΝΑ

	Σελ
ΑΦΙΕΡΩΣΗ	i
ΕΥΧΑΡΙΣΤΙΕΣ	iii
ΠΕΡΙΕΧΟΜΕΝΑ	iv
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ	vi
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ	vii
ΠΕΡΙΛΗΨΗ	xii
EXTENDED ABSTRACT IN ENGLISH	xiv
ΚΕΦΑΛΑΙΟ 1. Εισαγωγή	1
1.1. Στατιστική μοντελοποίηση - Μοντέλα Π-σιγμοειδών μικτών κατανομών	1
1.2. Πώς ορίζεται το πρόβλημα της εκτίμησης πυκνότητας πιθανότητας;	2
1.3. Η κανονική κατανομή ως παραμετρικό μοντέλο	4
1.4. Μικτές κατανομές	7
1.5. Μικτές κανονικές κατανομές	10
1.6. Μέθοδοι εκτίμησης παραμέτρων	11
1.6.1. Μέγιστη Πιθανοφάνεια	11
1.6.2. Εφαρμογή της μέγιστης πιθανοφάνειας στην κανονική κατανομή	13
1.6.3. Εφαρμογή σε μικτές κατανομές	14
1.7. Βελτιστοποίηση μέσω του αλγορίθμου Expectation-Maximization (EM)	16
1.7.1. Ορισμός του αλγορίθμου - Εφαρμογή στα μοντέλα μικτών κατανομών	16
1.7.2. Σύγκλιση του αλγορίθμου EM	20
1.7.3. Ο αλγόριθμος Generalized Expectation Maximization (GEM)	21
1.7.4. Εφαρμογή του αλγορίθμου EM σε μικτές κανονικές κατανομές	23
1.7.5. Σχολιασμός του αλγορίθμου EM	25
1.8. Το κριτήριο dip	26
1.9. Διάρθρωση και συνεισφορά της εργασίας	29
ΚΕΦΑΛΑΙΟ 2. Η κατανομή Π-sigmoid	31
2.1. Γενικά	31
2.2. Η σιγμοειδής συνάρτηση	33
2.3. Ορισμός της συνάρτησης πυκνότητας πιθανότητας Π-sigmoid	36
2.4. Ιδιότητες της Π-sigmoid	40
2.5. Η πολυδιάστατη Π-sigmoid κατανομή	41
2.6. Μικτά μοντέλα Π-sigmoid κατανομών (PsMM)	42
2.7. Εκτίμηση του αριθμού των συνιστωσών κατανομών σε ένα PsMM	43
ΚΕΦΑΛΑΙΟ 3. Η ασύμμετρη Π-sigmoid κατανομή	48
3.1. Γενικά	48
3.2. Ορισμός της συνάρτησης πυκνότητας πιθανότητας asymmetric Π-sigmoid	51
3.3. Ιδιότητες της asymmetric Π-sigmoid	56
3.4. Η πολυδιάστατη asymmetric Π-sigmoid κατανομή	59

3.5. Εκπαίδευση της asymmetric Π-sigmoid κατανομής	60
3.6. Μικτά μοντέλα asymmetric Π-sigmoid κατανομών	73
3.7. Εκπαίδευση ενός AsPisMM μέσω του GEM αλγορίθμου	75
3.7.1. Αρχικοποίηση του GEM αλγορίθμου	76
3.7.2. Ε-βήμα (Expectation step)	78
3.7.3. Μ-βήμα (Maximization step)	78
3.8. Αντιμετώπιση θορύβου	88
3.9. Εκτίμηση του πλήθους των συνιστωσών κατανομών σε ένα μοντέλο μικτών asymmetric Π-sigmoid κατανομών.	89
ΚΕΦΑΛΑΙΟ 4. Πειραματική αξιολόγηση	93
4.1. Εισαγωγή	93
4.2. Τεχνητά σύνολα παρατηρήσεων	94
4.3. Πειραματική αξιολόγηση τεχνητών συνόλων παρατηρήσεων	96
4.3.1. Ομοιογενή σύνολα μονοδιάστατων παρατηρήσεων	97
4.3.2. Σύνθετα σύνολα μονοδιάστατων παρατηρήσεων	101
4.3.3. Ομοιογενή σύνολα πολυδιάστατων παρατηρήσεων	103
4.3.4. Σύνθετα σύνολα πολυδιάστατων παρατηρήσεων	104
4.4. Συμπεράσματα	106
4.5. Πειραματική αξιολόγηση πραγματικών συνόλων παρατηρήσεων	107
ΚΕΦΑΛΑΙΟ 5. Συμπεράσματα και μελλοντική έρευνα	118
5.1. Συμπεράσματα	118
5.2. Μελλοντική έρευνα	119
ΑΝΑΦΟΡΕΣ	120
ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ	121

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας	Σελ
Πίνακας 4.1 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	97
Πίνακας 4.2 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	97
Πίνακας 4.3 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	97
Πίνακας 4.4 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM. Dataset: Mixed, D=1, K=3, Ntest set=4500	101
Πίνακας 4.5 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	101
Πίνακας 4.6 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	103
Πίνακας 4.7 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	103
Πίνακας 4.8 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	104
Πίνακας 4.9 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	104
Πίνακας 4.10 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	104
Πίνακας 4.11 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	104
Πίνακας 4.12 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	105
Πίνακας 4.13 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	105
Πίνακας 4.14 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	105
Πίνακας 4.15 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	105
Πίνακας 4.16 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	105
Πίνακας 4.17 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.	106
Πίνακας 4.18 Σύγκριση των μοντέλων GMM, PsMM και AsPsMM σε gray-scale εικόνες. Ισχύει: Ntrain = Ntest = 14700	117

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Σχήμα	Σελ
Σχήμα 1.1: Η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής για διάφορες τιμές των παραμέτρων της.	5
Σχήμα 1.2: Γραφικό μοντέλο αναπαράστασης ενός μοντέλου μικτών κατανομών. Η z είναι κρυμμένη μεταβλητή (αναπαριστά την συνιστώσα κατανομή που δημιουργεί το x) και η από κοινού συνάρτηση πυκνότητας πιθανότητας έχει την μορφή $p(x, z) = p(z)p(x z) = \pi_k f(x; \theta_k)$. Μέσω περιθωριοποίησης ως προς την μεταβλητή z προκύπτει ο ορισμός του μοντέλου μικτών κατανομών.	8
Σχήμα 1.3: Γραφικό μοντέλο μικτών κανονικών κατανομών, που αναπαριστά την διαδικασία της δημιουργίας των παρατηρήσεων μέσω εξαρτήσεων τυχαίων μεταβλητών. Με τα μικρά τετράγωνα συμβολίζονται οι παράμετροι του μοντέλου, με κύκλους (κόμβους) συμβολίζονται οι τυχαίες μεταβλητές και οι κατανομές τους, ενώ οι ακμές μεταξύ δυο κύκλων (κόμβων) δηλώνουν την αλληλεπίδραση των τυχαίων μεταβλητών. Τέλος, τα γεμισμένα σχήματα υποδηλώνουν τις παρατηρούμενες τιμές, ενώ ο δείκτης $[K]$, υποδηλώνει διάλυσμα μεγέθους K .	10
Σχήμα 1.4 Απεικόνιση ενός βήματος του EM. Η συνάρτηση G , αποτελεί το κάτω φράγμα της συνάρτησης πιθανοφάνειας. Οι δύο συναρτήσεις είναι ίσες στο σημείο $\theta(t)$. Στο M -βήμα το $\theta(t+1)$ ορίζεται ως η τιμή του θ που μεγιστοποιεί το G .	22
Σχήμα 1.5 Ο τρόπος λειτουργίας του αλγόριθμου GEM για μια επανάληψη. Παρατηρούμε ότι παρόλο που δεν υπολογίζεται το μέγιστο του κάτω φράγματος G , η πιθανοφάνεια αυξάνεται.	22
Σχήμα 2.1 Η σιγμοειδής συνάρτηση με κέντρο το μηδέν και κλίση μονάδα.	33
Σχήμα 2.2 Η γραφική παράσταση της σιγμοειδούς συνάρτησης με κέντρο $a=0$ και κλίση $\lambda=5$.	34
Σχήμα 2.3 Η γραφική παράσταση της σιγμοειδούς συνάρτησης με κέντρο $a=0$ και κλίση $\lambda=100$. Φαίνεται καθαρά η απότομη μεταβολή της συνάρτησης από το μηδέν στην μονάδα.	35
Σχήμα 2.4 Παράθεση δύο διαφορετικών σιγμοειδών συναρτήσεων με κέντρα $a=-4$ και $b=4$ και κλίση $\lambda=1.5$. Το αποτέλεσμα της διαφοράς τους θα είναι μια καμπανοειδής συνάρτηση ενώ για $\lambda \gg 1$ θα τείνει σε σχήμα Π .	37
Σχήμα 2.5 Η κατανομή Π -sigmoid με παραμέτρους $a=-4$, $b=4$ και κλίση $\lambda=1$. Να σημειωθεί ότι το σχήμα αυτής της κατανομής προκύπτει από την διαφορά των σιγμοειδών συναρτήσεων αντίστοιχων με το Σχήμα 2.4.	37
Σχήμα 2.6 Η κατανομή Π -sigmoid με παραμέτρους $a=-4$, $b=4$ και κλίση $\lambda=5$. Μεγαλώνοντας την τιμή του λ , το γεωμετρικό σχήμα της κατανομής αρχίζει να προσεγγίζει το σχήμα Π .	38

- Σχήμα 2.7 Η κατανομή Π-sigmoid με παραμέτρους $a=-4$, $b=4$ και κλίση $\lambda=50$. Μεγαλώνοντας περισσότερο την τιμή του λ , η κατανομή προσεγγίζει με πολύ ικανοποιητικό τρόπο την ομοιόμορφη. 38
- Σχήμα 2.8 Η κατανομή Π-sigmoid με παραμέτρους $a=-4$, $b=4$ και κλίση $\lambda=0.1$. Η συνάρτηση πλατειάζει σημαντικά και για αυτό ανοίγουμε το διάστημα των τιμών από το $[-10\ 10]$ στο $[-60\ 60]$ για να είναι ορατό το σχήμα της. 39
- Σχήμα 2.9 Η κατανομή Π-sigmoid με παραμέτρους $a=-0.001$, $b=0.001$ και κλίση $\lambda=5$. Η συνάρτηση γίνεται απότομη, όταν η σχετική απόσταση του a με το b γίνει μικρή. 39
- Σχήμα 3.1 Επικαλυπτόμενες ομάδες κανονικών και ομοιόμορφων παρατηρήσεων. Το αποτέλεσμα μπορεί να θεωρηθεί ως μια ενιαία ομάδα. 50
- Σχήμα 3.2 Μη αποδοτική περιγραφή ενός μη-συμμετρικού συνόλου παρατηρήσεων από την συμμετρική κατανομή Π-sigmoid. 50
- Σχήμα 3.3 Δύο Π-sigmoid κατανομές με κοινό κέντρο $b=0$ και ίσες τιμές κλίσεων. Το αποτέλεσμα που θα προκύψει από το ημίθροισμά τους είναι μια συνεχής και συμμετρική συνάρτηση AsΠ-sigmoid 53
- Σχήμα 3.4 Δύο Π-sigmoid κατανομές με κοινό κέντρο $b=0$ και διαφορετικές τιμές κλίσεων. Το αποτέλεσμα που θα προκύψει από το ημίθροισμά τους είναι μια μη-συμμετρική συνάρτηση AsΠ-sigmoid. 54
- Σχήμα 3.5 Η συνάρτηση πυκνότητας πιθανότητας AsΠ-sigmoid. Παρατηρούμε τη γενική μη-συμμετρική της εκδοχή. Η «ανώμαλη» συμπεριφορά της συνάρτησης στο κέντρο της, είναι αποτέλεσμα του αθροίσματος των δύο Π-sigmoid κατανομών με διαφορετικές κλίσεις. 54
- Σχήμα 3.6 Η συνάρτηση πυκνότητας πιθανότητας AsΠ-sigmoid. Παρατηρούμε την δυνατότητα της κατανομής να πάρει συμμετρικό σχήμα που προσεγγίζει το σχήμα της κανονικής κατανομής, όταν οι δύο Π-sigmoid κατανομές έχουν ίσες κλίσεις με τιμές κοντά στη μονάδα. 55
- Σχήμα 3.7 Η συνάρτηση πυκνότητας πιθανότητας AsΠ-sigmoid. Όσο οι κοινές τιμές στις δύο κλίσεις αυξάνονται, τόσο η μετάβολή των τιμών της συνάρτησης στα κέντρα a, c γίνεται πιο απότομη. 55
- Σχήμα 3.8 Η συνάρτηση πυκνότητας πιθανότητας AsΠ-sigmoid. Με μια μεγάλη αύξηση στις κοινές τιμές των δύο κλίσεων η AsΠ-sigmoid κατανομή προσεγγίζει πλέον το σχήμα της ομοιόμορφης κατανομής. 56
- Σχήμα 3.9 Ένα περίγραμμα της κανονικής κατανομής στις δύο διαστάσεις η οποία χαρακτηρίζεται από το κέντρο μ και πίνακα συσχετιστικότητας του οποίου τα ιδιοδιανύσματα είναι u_1 και u_2 , με αντίστοιχες ιδιοτιμές L_1 και L_2 . 69
- Σχήμα 3.10 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε μονοδιάστατο σύνολο κανονικών παρατηρήσεων. 70
- Σχήμα 3.11 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε δισδιάστατο σύνολο κανονικών παρατηρήσεων. 70
- Σχήμα 3.12 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε μονοδιάστατο σύνολο ομοιόμορφων παρατηρήσεων. 71
- Σχήμα 3.13 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε δισδιάστατο σύνολο ομοιόμορφων παρατηρήσεων. 71
- Σχήμα 3.14 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε μονοδιάστατο σύνολο truncated κανονικών παρατηρήσεων. 72
- Σχήμα 3.15 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε δισδιάστατο σύνολο truncated κανονικών παρατηρήσεων. 72

- Σχήμα 3.16 Ανεπιτυχής περιγραφή μέσω μιας AsΠ-sigmoid δύο ομάδων παρατηρήσεων 73
- Σχήμα 3.17 Ανεπιτυχής περιγραφή μέσω μιας AsΠ-sigmoid δύο ομάδων παρατηρήσεων. 73
- Σχήμα 3.18 Αποτέλεσμα της εκπαίδευσης ενός AsΠsMM σε μονοδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων. 84
- Σχήμα 3.19 Αποτέλεσμα της εκπαίδευσης ενός AsΠsMM σε διδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων. 84
- Σχήμα 3.20 Αποτέλεσμα της εκπαίδευσης ενός AsΠsMM σε μονοδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων και μια ομάδα ομοιόμορφων παρατηρήσεων. 85
- Σχήμα 3.21 Αποτέλεσμα της εκπαίδευσης ενός AsΠsMM σε διδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων. 85
- Σχήμα 3.22 Αποτέλεσμα της εκπαίδευσης ενός AsΠsMM σε μονοδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα κανονικών παρατηρήσεων και μια ομάδα ομοιόμορφων παρατηρήσεων. 86
- Σχήμα 3.23 Αποτέλεσμα της εκπαίδευσης ενός AsΠsMM σε διδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων. 86
- Σχήμα 3.24 Αποτέλεσμα της εκπαίδευσης ενός AsΠsMM σε μονοδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων, μια ομάδα ομοιόμορφων παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων. 87
- Σχήμα 3.25 Αποτέλεσμα της εκπαίδευσης ενός AsΠsMM σε διδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων, μια ομάδα ομοιόμορφων παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων. 87
- Σχήμα 4.1 Μικτό σύνολο που αποτελείται από εννέα ομάδες διδιάστατων παρατηρήσεων. Περιλαμβάνει τρεις Gaussian με ελλειπτικό σχήμα, τρεις uniform με ορθογώνιο σχήμα και τρεις truncated Gaussian ως τμήματα ελλείψεων. Η αναλογία των κατηγοριών των ομάδων είναι σταθερή και οι ομάδες περιέχουν ίσο πλήθος παρατηρήσεων. Το αρχικό σύνολο διαμερίζεται τυχαία σε ένα σύνολο εκπαίδευσης και σε ένα σύνολο ελέγχου. Μετά το τέλος της εκπαίδευσης προκύπτουν οι εκτιμήσεις για το σύνολο των παραμέτρων του κάθε μοντέλου καθώς και του πλήθους των συνιστωσών κατανομών για τα AsΠsMM, ΠsMM. Τέλος, υπολογίζεται η τιμή της αρνητικής λογαριθμικής πιθανοφάνειας για κάθε μοντέλο. Το μοντέλο με την ελάχιστη τιμή, πετυχαίνει την βέλτιστη περιγραφή του συνόλου παρατηρήσεων. 96
- Σχήμα 4.2 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία gaussian ομάδα μέσω της κατανομής Π-sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης. 98
- Σχήμα 4.3 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία gaussian ομάδα μέσω της κατανομής AsΠ-sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης. 98

- Σχήμα 4.4 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία uniform ομάδα μέσω της κατανομής Π-sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης. 99
- Σχήμα 4.5 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία uniform ομάδα μέσω της κατανομής AsΠ-sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης. 99
- Σχήμα 4.6 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία truncated gaussian ομάδα μέσω της κατανομής Π-sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης. 100
- Σχήμα 4.7 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία truncated gaussian ομάδα μέσω της κατανομής AsΠ-sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης. 100
- Σχήμα 4.8 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από τρεις ομάδες παρατηρήσεων μέσω τριών Π-sigmoid κατανομών. Κάθε κατανομή μοντελοποιεί μια ενιαία ομάδα παρατηρήσεων. 101
- Σχήμα 4.9 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από τρεις ομάδες παρατηρήσεων μέσω του PsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου GEM στο σύνολο εκπαίδευσης. 101
- Σχήμα 4.10 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από τρεις ομάδες παρατηρήσεων μέσω τριών AsΠ-sigmoid κατανομών. Κάθε κατανομή μοντελοποιεί μια ενιαία ομάδα παρατηρήσεων. 102
- Σχήμα 4.11 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από τρεις ομάδες παρατηρήσεων μέσω του AsΠsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου GEM στο σύνολο εκπαίδευσης. 102
- Σχήμα 4.12 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από έξι ομάδες παρατηρήσεων μέσω έξι Π-sigmoid κατανομών. Κάθε κατανομή μοντελοποιεί μια ενιαία ομάδα παρατηρήσεων. 102
- Σχήμα 4.13 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από έξι ομάδες παρατηρήσεων μέσω του PsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου GEM στο σύνολο εκπαίδευσης. 102
- Σχήμα 4.14 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από έξι ομάδες παρατηρήσεων μέσω έξι AsΠ-sigmoid κατανομών. Κάθε κατανομή μοντελοποιεί μια ενιαία ομάδα παρατηρήσεων. 103
- Σχήμα 4.15 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από έξι ομάδες παρατηρήσεων μέσω του AsΠsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου GEM στο σύνολο εκπαίδευσης. 103
- Σχήμα 4.16 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “elephants.jpg” μέσω του GMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 109

- Σχήμα 4.17 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “elephants.jpg” μέσω του PsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 109
- Σχήμα 4.18 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “elephants.jpg” μέσω του AsPsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 109
- Σχήμα 4.19 Το αποτέλεσμα της κατάτμησης της αρχικής εικόνας “elephants.jpg” μέσω των τριών μοντέλων: GMM, PsMM, AsPsMM. Κάθε μοντέλο αποτελείται από τέσσερις συνιστώσες κατανομές (ομάδες). 110
- Σχήμα 4.20 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “clouds.jpg” μέσω του GMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 111
- Σχήμα 4.21 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “clouds.jpg” μέσω του PsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 111
- Σχήμα 4.22 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “clouds.jpg” μέσω του AsPsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 111
- Σχήμα 4.23 Το αποτέλεσμα της κατάτμησης της αρχικής εικόνας “clouds.jpg” μέσω των τριών μοντέλων: GMM, PsMM, AsPsMM. Κάθε μοντέλο αποτελείται από τέσσερις συνιστώσες κατανομές (ομάδες). 112
- Σχήμα 4.24 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rocks.jpg” μέσω του GMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 113
- Σχήμα 4.25 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rocks.jpg” μέσω του PsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 113
- Σχήμα 4.26 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rocks.jpg” μέσω του AsPsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 113
- Σχήμα 4.27 Το αποτέλεσμα της κατάτμησης της αρχικής εικόνας “rocks.jpg” μέσω των τριών μοντέλων: GMM, PsMM, AsPsMM. Κάθε μοντέλο αποτελείται από τέσσερις συνιστώσες κατανομές (ομάδες). 114
- Σχήμα 4.28 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rock_and_tree.jpg” μέσω του GMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 115
- Σχήμα 4.29 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rock_and_tree.jpg” μέσω του AsPsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 115
- Σχήμα 4.30 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rock_and_tree.jpg” μέσω του AsPsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης. 115
- Σχήμα 4.31 Το αποτέλεσμα της κατάτμησης της αρχικής εικόνας “rock_and_tree.jpg” μέσω των τριών μοντέλων: GMM, PsMM, AsPsMM. Κάθε μοντέλο αποτελείται από τέσσερις συνιστώσες κατανομές (ομάδες). 116

ΠΕΡΙΛΗΨΗ

Ευστάθιος Μπαλίκας του Ανδρέα και της Όλγας. MSc, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Σεπτέμβριος 2017. Στατιστική μοντελοποίηση δεδομένων με την ασύμμετρη Π-σιγμοειδή κατανομή. Επιβλέποντας: Αριστείδης Λύκας.

Δοθέντος ενός συνόλου παρατηρήσεων, το πρόβλημα της στατιστικής μοντελοποίησης συνίσταται στην εκτίμηση της συνάρτησης πυκνότητας πιθανότητας των παρατηρήσεων αυτών. Η Π-σιγμοειδής κατανομή ορίζεται ως η διαφορά δύο μετατοπισμένων λογιστικών σιγμοειδών συναρτήσεων που έχουν την ίδια κλίση, και έχει προταθεί ως μια συνεχής προσέγγιση της ομοιόμορφης κατανομής η οποία είναι ευκολότερο να εκπαιδευτεί σε σχέση με την ομοιόμορφη κατανομή. Ένα μειονέκτημα της κατανομής αυτής, όπως και της κανονικής κατανομής, είναι η συμμετρία που εμφανίζουν, γεγονός που τις καθιστά λιγότερο αποδοτικές για μοντελοποίηση παρατηρήσεων που δεν προέρχονται από κάποια συμμετρική κατανομή.

Στην παρούσα εργασία προτείνεται μια επέκταση της Π-σιγμοειδούς κατανομής που ονομάζεται ασύμμετρη Π-σιγμοειδής κατανομή και ορίζεται ως ένα μικτό μοντέλο από δύο Π-σιγμοειδείς κατανομές με διαφορετική κλίση η καθεμιά. Μελετώνται οι ιδιότητες της κατανομής αυτής και προτείνεται η εκπαίδευσή της δοθέντος ενός συνόλου παρατηρήσεων με χρήση του αλγορίθμου GEM (Generalized EM). Στη συνέχεια ορίζεται ένα μικτό μοντέλο από ασύμμετρες Π-σιγμοειδείς κατανομές και προτείνεται ένας αλγόριθμος εκπαίδευσης βασισμένος στον αλγόριθμο GEM. Επιπλέον προτείνεται και μια μεθοδολογία καθορισμού του πραγματικού αριθμού των συνιστωσών του μικτού μοντέλου με χρήση του στατιστικού κριτηρίου dip το οποίο εκτιμά την μονοτροπικότητα ενός συνόλου παρατηρήσεων.

Η επίδοση των προτεινόμενων στατιστικών μοντέλων και των μεθόδων εκπαίδευσής τους εξετάστηκε πειραματικά σε τεχνητά σύνολα παρατηρήσεων που

δημιουργήθηκαν από κανονικές, ομοιόμορφες και ασύμμετρες κατανομές, όπως για παράδειγμα η αποκομμένη (truncated) κανονική κατανομή. Από τα πειράματα προκύπτει ότι η χρήση ενός μικτού μοντέλου ασύμμετρων Π-σιγμοειδών κατανομών οδηγεί σε βελτιωμένες κατά μέσο όρο επιδόσεις μοντελοποίησης και σε ικανοποιητικές εκτιμήσεις του πραγματικού αριθμού των συνιστωσών του μικτού μοντέλου. Η πειραματική αξιολόγηση επεκτάθηκε και σε πραγματικά σύνολα παρατηρήσεων και συγκεκριμένα σε εικόνες. Στόχος της πειραματικής διαδικασίας που ακολουθήθηκε, ήταν αφενός η μοντελοποίηση των εικονοστοιχείων (pixels) της εικόνας και αφετέρου η κατάτμηση της εικόνας σε προκαθορισμένο πλήθος περιοχών (ομάδων). Από τα πειράματα προκύπτει πως οι επιδόσεις μοντελοποίησης ενός μικτού μοντέλου ασύμμετρων Π-σιγμοειδών κατανομών είναι βελτιωμένες και επιπλέον, οι τελικές περιοχές (ομάδες) που προκύπτουν από την διαδικασία της κατάτμησης, αναπαριστούν με μεγαλύτερη αξιοπιστία τις πραγματικές περιοχές της εικόνας.

Συμπερασματικά, ένα μικτό μοντέλο ασύμμετρων Π-σιγμοειδών κατανομών αποτελεί μια ιδιαίτερα αποδοτική μέθοδο μοντελοποίησης συνόλων παρατηρήσεων. Επιπλέον, η ενίσχυση της διαδικασίας εκπαίδευσης μέσω του κριτηρίου dip , προσδίδει στο μοντέλο μια σημαντική αυτονομία και το καθιστά ως ένα ισχυρό αλγόριθμο για τη στατιστική μοντελοποίηση παρατηρήσεων.

EXTENDED ABSTRACT IN ENGLISH

Balikas, Efstathios, MSc, Department of Computer Science and Engineering, University of Ioannina, Greece. September, 2017. Statistical data modeling using the asymmetric P-sigmoid distribution. Thesis Supervisor: Aristidis Likas.

Given a set of observations, statistical modeling is defined as the problem of estimating the probability distribution function that best explains these observations. The Π -sigmoid distribution, which has been proposed as a continuous approximation of the uniform distribution, is the difference between two shifted logistic sigmoid functions that have the same slope. A disadvantage of both the uniform distribution as well as the Π -sigmoid distribution is their symmetry. Because of that, they are inefficient when modeling asymmetric data.

In this thesis, we propose an extension of Π -sigmoid distribution, namely the asymmetric Π -sigmoid distribution, which is defined as a mixture model of two Π -sigmoid distributions with different slopes. We first study the theoretical properties of the asymmetric Π -sigmoid distribution. Then, given a set of observations, we propose a training algorithm which is based on Generalized Expectation-Maximization (GEM) algorithm to estimate its parameters. Further, we extend the proposed asymmetric Π -sigmoid distribution to a mixture model of asymmetric Π -sigmoid distributions and describe the associated GEM training algorithm. To identify the number of components of the mixture model, we explore the use of the dip statistical criterion, that estimates the unimodality of the empirical cdf (ecdf), given a set of observations. We empirically validate the efficiency of the proposed statistical models as well as their training algorithms in synthetic datasets created by sampling from Gaussian, uniform and asymmetric distributions such as the truncated Gaussian distribution. Our results suggest that using a mixture model of asymmetric Π -sigmoid distributions

results in better data modeling performance and satisfactory estimation of the number of mixture components.

We extend our evaluation in modeling the histogram of images. Our goal, is twofold: on one hand to model the image pixel values and on the other hand to evaluate the corresponding segmentation of the image on a predefined number of clusters. Our results suggest that using asymmetric Π -sigmoid distributions clearly benefits the performance on the task. Furthermore, the obtained clusters better reflect the ground truth segments of the images.

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

- 1.1. Στατιστική μοντελοποίηση - Μοντέλα Π-σιγμοειδών μικτών κατανομών
 - 1.2. Πώς ορίζεται το πρόβλημα της εκτίμησης πυκνότητας πιθανότητας;
 - 1.3. Η κανονική κατανομή ως παραμετρικό μοντέλο
 - 1.4. Μικτές κατανομές
 - 1.5. Μικτές κανονικές κατανομές
 - 1.6. Μέθοδοι εκτίμησης παραμέτρων
 - 1.6.1. Μέγιστη Πιθανοφάνεια
 - 1.6.2. Εφαρμογή της μέγιστης πιθανοφάνειας στην κανονική κατανομή
 - 1.6.3. Εφαρμογή σε μικτές κατανομές
 - 1.7. Βελτιστοποίηση μέσω του αλγορίθμου Expectation-Maximization (EM)
 - 1.7.1. Ορισμός του αλγορίθμου - Εφαρμογή στα μοντέλα μικτών κατανομών
 - 1.7.2. Σύγκλιση του αλγορίθμου EM
 - 1.7.3. Ο αλγόριθμος Generalized Expectation Maximization (GEM)
 - 1.7.4. Εφαρμογή του αλγορίθμου EM σε μικτές κανονικές κατανομές
 - 1.7.5. Σχολιασμός του αλγορίθμου EM
 - 1.8. Το dip κριτήριο
 - 1.9. Διάρθρωση και συνεισφορά της διατριβής
-

1.1. Στατιστική μοντελοποίηση - Μοντέλα Π-σιγμοειδών μικτών κατανομών

Η στατιστική μοντελοποίηση αποτελεί μια ξεχωριστή κατηγορία της γενικότερης, μαθηματικής μοντελοποίησης. Εμπεριέχει ένα σύνολο υποθέσεων που σχετίζονται με τον τρόπο γένεσης του συνόλου των παρατηρήσεων που έχουμε στην διάθεσή μας,

αλλά και παρόμοιων παρατηρήσεων που αντιστοιχούν στον πληθυσμό. Οι υποθέσεις αυτές ουσιαστικά περιγράφουν ένα σύνολο από κατανομές πιθανοτήτων. Οι κατανομές λοιπόν, είναι άρρηκτα συνδεδεμένες με τα στατιστικά μοντέλα και είναι αυτές που τα διαχωρίζουν από τα υπόλοιπα, μαθηματικά μοντέλα. Ένα στατιστικό μοντέλο αναπαριστά, σε μεγάλο βαθμό, την διαδικασία γένεσης των παρατηρήσεων και περιγράφεται συνήθως από μαθηματικές εξισώσεις οι οποίες συνδέουν μια ή περισσότερες τυχαίες μεταβλητές αλλά και άλλες, άγνωστες μεταβλητές.

Συμπερασματικά, ο κύριος σκοπός της στατιστικής μοντελοποίησης είναι η δημιουργία ενός πιθανοτικού μοντέλου ικανού να προσδιορίζει τα μοτίβα (patterns) των παρατηρήσεων, να εντοπίζει τις κατανομές τους και να περιγράφει τις πηγαιές στατιστικές τους ιδιότητες.

Πιο συγκεκριμένα, τα μοντέλα μικτών κατανομών, αποτελούν ευέλικτα και ισχυρά πιθανοτικά μοντέλα για την ανάλυση παρατηρήσεων. Η βασική προσέγγιση τους, έγκειται στην υπόθεση ότι οι παρατηρήσεις προκύπτουν από μια δειγματοληψία από παραμετρικές κατανομές. Υπό την προϋπόθεση ότι οι παρατηρήσεις είναι ανεξάρτητες, θα μπορούσαμε να υποθέσουμε διάφορα παραμετρικά στατιστικά μοντέλα για να περιγράψουμε τις στατιστικές ιδιότητες του συνόλου των παρατηρήσεων. Στην παρούσα εργασία προτείνουμε ένα ασύμμετρο μοντέλο Π-σιγμοειδών μικτών κατανομών.

Όπως θα δείξουμε σε επόμενα κεφάλαια, σε σύγκριση με το κυρίαρχο και ευρέως διαδεδομένο στατιστικό μοντέλο των μικτών κανονικών κατανομών, τα ασύμμετρα μοντέλα Π-σιγμοειδών μικτών κατανομών μπορούν να πετύχουν καλύτερη επίδοση σε συγκεκριμένες κατηγορίες δεδομένων (ομοιόμορφα, μικτά ομοιόμορφα-κανονικά κ.α.). Αυτό οφείλεται κυρίως στην ευελιξία της σιγμοειδούς συνάρτησης. Συγκεκριμένα, όπως και θα αναλυθεί στην συνέχεια, η σιγμοειδής κατανομή μπορεί να πάρει διαφορετικά σχήματα ικανά να παρέχουν μια προσαρμογή με μεγάλη ακρίβεια στις κατανομές των υπό μελέτη δεδομένων.

1.2. Πώς ορίζεται το πρόβλημα της εκτίμησης πυκνότητας πιθανότητας;

Η συνάρτηση πυκνότητας πιθανότητας - ή συνάρτηση μάζας πιθανότητας στην διακριτή περίπτωση - είναι μια θεμελιώδης έννοια στην θεωρία πιθανοτήτων και στη

στατιστική. Έστω μια τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας $p(x)$. Προσδιορίζοντας την συνάρτηση $p(x)$, έχουμε μια άμεση περιγραφή της κατανομής της τυχαίας μεταβλητής X και έναν τρόπο υπολογισμού πιθανοτήτων που σχετίζονται με την X από την σχέση:

$$P(a < X < b) = \int_a^b p(x) dx \quad \forall a < b \quad (1.1)$$

Έστω ένα σύνολο N παρατηρήσεων, διάστασης D :

$$X = \{x_j^i\}, \quad i = 1 \dots N, \quad j = 1 \dots D$$

Υποθέτουμε ότι οι παρατηρήσεις προήλθαν μέσω μιας δειγματοληψίας από μια άγνωστη συνάρτηση πυκνότητας πιθανότητας. Το πρόβλημα λοιπόν ανάγεται στην εκτίμηση της συνάρτησης πυκνότητας πιθανότητας βάση των παρατηρήσεων που έχουμε στην διάθεση μας.

Μια από τις διαδεδομένες μεθόδους εκτίμησης συνάρτησης πυκνότητας πιθανότητας είναι η παραμετρική. Υποθέτουμε ότι οι παρατηρήσεις έχουν προκύψει μέσω δειγματοληψίας από μια γνωστή παραμετρική οικογένεια κατανομών (για παράδειγμα από την κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2). Η συνάρτηση πυκνότητας πιθανότητας που αναπαριστά την κατανομή των παρατηρήσεων στην περίπτωση αυτή, μπορεί να εκτιμηθεί, υπολογίζοντας την μέση τιμή και τη διακύμανση από τις παρατηρήσεις και έπειτα αντικαθιστώντας αυτές, στην παραμετρική συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής. Με αυτόν τον τρόπο, έχει υπολογιστεί μια εκτίμηση της κατανομής των παρατηρήσεων.

Στην γενική περίπτωση, υποθέτουμε ότι η άγνωστη κατανομή των παρατηρήσεων περιγράφεται από μια συνάρτηση πυκνότητας πιθανότητας εξαρτώμενη από ένα διάνυσμα παραμέτρων. Η διαδικασία της εκτίμησης λοιπόν, ανάγεται στη βελτιστοποίηση των παραμέτρων αυτών, έτσι ώστε να επιτευχθεί η καλύτερη δυνατή περιγραφή της κατανομής των παρατηρήσεων από την παραμετρική συνάρτηση πυκνότητας πιθανότητας. Θα συμβολίζουμε την συνάρτηση πυκνότητας πιθανότητας η οποία εξαρτάται από το διάνυσμα παραμέτρων ως $p(x; \theta)$.

1.3. Η κανονική κατανομή ως παραμετρικό μοντέλο

Ένα από τα πλέον διαδεδομένα παραμετρικά μοντέλα συνεχών τυχαίων μεταβλητών είναι η κανονική (Gaussian) κατανομή, η οποία χαρακτηρίζεται από καλές στατιστικές και αναλυτικές ιδιότητες. Στην μονοδιάστατη εκδοχή της η κανονική κατανομή έχει την μορφή:

$$N(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} \quad (1.2)$$

όπου μ είναι η μέση τιμή και σ^2 η διακύμανση. Η σταθερά κανονικοποίησης:

$$\frac{1}{(2\pi\sigma^2)^{1/2}} \quad (1.3)$$

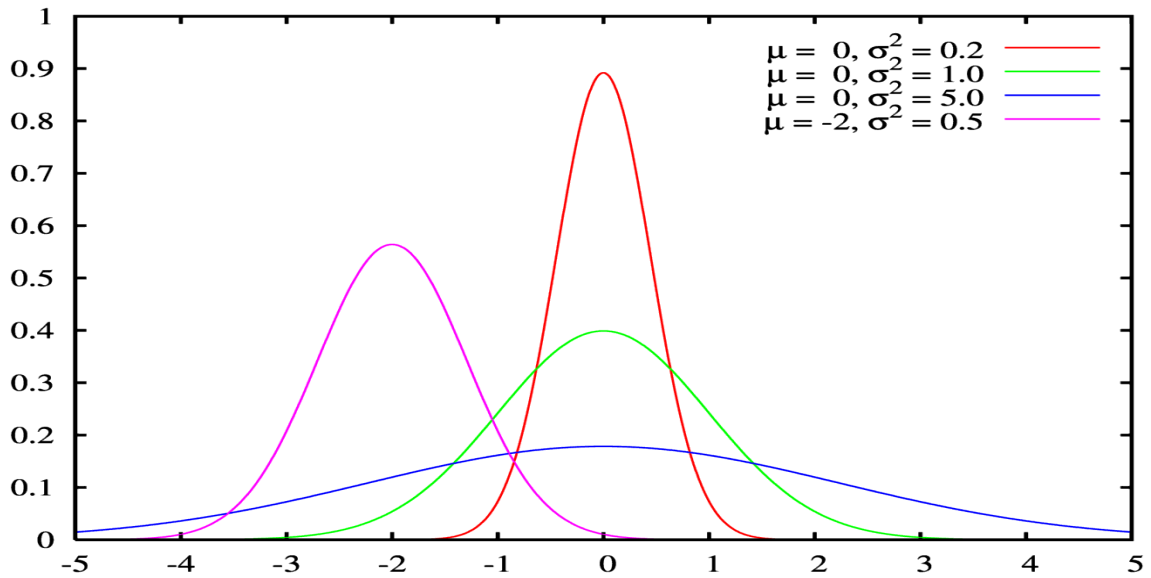
μας εξασφαλίζει ότι ισχύει η θεμελιώδης σχέση-ορισμός μιας συνάρτησης πυκνότητας πιθανότητας:

$$\int_{-\infty}^{+\infty} N(x|\mu, \sigma^2) dx = 1 \quad (1.4)$$

Αναφορικά με την μέση τιμή και την διακύμανση της κανονικής κατανομής, ισχύουν οι παρακάτω σχέσεις:

$$\mu = E(x) = \int_{-\infty}^{+\infty} xp(x) dx \quad (1.5)$$

$$\sigma^2 = E[(x - \mu)^2] = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx \quad (1.6)$$



Σχήμα 1.1: Η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής για διάφορες τιμές των παραμέτρων της.

Γενικότερα, η πολυδιάστατη κανονική κατανομή στις D -διαστάσεις έχει την ακόλουθη μορφή:

$$N(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right\} \quad (1.7)$$

όπου το μ είναι πλέον ένα διάνυσμα μέσης τιμής διάστασης D , ενώ ο Σ είναι ένας $D \times D$ πίνακας συμμεταβλητότητας και $|\Sigma|$ είναι η ορίζουσά του.

Ο παράγοντας μπροστά από το εκθετικό μέρος της συνάρτησης, όπως και στην μονοδιάστατη εκδοχή, μας εξασφαλίζει ότι θα ισχύει:

$$\int_{-\infty}^{+\infty} N(x|\mu, \Sigma) dx = 1 \quad (1.8)$$

Αναφορικά με τον συμβολισμό, θα γράφουμε $X \sim N(x; \mu, \Sigma)$, όταν το X είναι μια τυχαία μεταβλητή που ακολουθεί την κανονική κατανομή με μέση τιμή μ και πίνακα συμμεταβλητότητας Σ . Είναι φανερό ότι η συνάρτηση πυκνότητας $N(x; \mu, \Sigma)$ χαρακτηρίζεται από τις παραμέτρους μ και Σ οι οποίες ικανοποιούν τις παρακάτω σχέσεις:

$$\mu = E[X] \quad (1.9)$$

$$\Sigma = E[(x - \mu)(x - \mu)^T] \quad (1.10)$$

Από την τελευταία σχέση συμπεραίνουμε ότι ο πίνακας Σ είναι συμμετρικός και θετικά ημι-ορισμένος. Η ιδιότητα της συμμετρίας μας οδηγεί στο συμπέρασμα ότι ο Σ αποτελείται από $D(D + 1)/2$ ανεξάρτητες παραμέτρους. Εάν συνυπολογίσουμε και τα D στοιχεία από τα οποία αποτελείται το διάνυσμα της μέσης τιμής, έχουμε συνολικά $D(D + 3)/2$ ανεξάρτητες παραμέτρους οι οποίες χαρακτηρίζουν την κανονική κατανομή.

Σε αυτό το σημείο θα παρουσιάσουμε την γεωμετρική μορφή της κανονικής κατανομής. Η συναρτησιακή εξάρτηση της κατανομής αυτής από το x , διατυπώνεται μέσω της παρακάτω τετραγωνικής μορφής:

$$\Delta^2 = (x - \mu)^T \Sigma^{-1} (x - \mu) \quad (1.11)$$

Η ποσότητα Δ ονομάζεται απόσταση mahalanobis μεταξύ των x και μ . Παρακάτω, για τιμές της απόστασης mahalanobis, αλλά και του πίνακα συμμεταβλητότητας Σ , παρουσιάζονται τρεις περιπτώσεις με γεωμετρικό ενδιαφέρον:

Σταθερή απόσταση Δ^2

Σε αυτή την περίπτωση όλες οι παρατηρήσεις x ανήκουν σε μια υπερ-ελλειψοειδή επιφάνεια με κέντρο μ και σχήμα που καθορίζεται από τον πίνακα Σ . Είναι προφανές ότι η τιμή της πολυδιάστατης συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής, για όλες τις παρατηρήσεις είναι σταθερή και αυτό μας οδηγεί στο συμπέρασμα ότι οι παρατηρήσεις x που παράγονται από αυτή, σχηματίζουν στις D διαστάσεις υπέρ-ελλειψοειδείς πυρήνες.

Διαγώνιος πίνακας συμμεταβλητότητας Σ

Σε αρκετές περιπτώσεις είναι ιδιαίτερα βολικό να θεωρήσουμε ότι οι παρατηρήσεις είναι ανεξάρτητες ανά διάσταση, δηλαδή για κάθε παρατήρηση x^i ισχύει:

$$\sigma_{ab} = \text{cov}(x_k^i, x_m^i) = 0 \quad \forall i = 1 \dots N, \quad \forall a, b = 1 \dots D, a \neq b \quad (1.12)$$

Αυτό πρακτικά σημαίνει ότι ο πίνακας Σ εκφυλίζεται σε διαγώνιο πίνακα μορφής

$$\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_D^2)$$

Η παραπάνω διατύπωση περιλαμβάνει $2D$ ανεξάρτητες παραμέτρους, και ουσιαστικά τα διανύσματα για τα οποία η $N(x; \mu, \Sigma)$ δίνει ίδιες τιμές, ορίζουν στις D διαστάσεις μια υπερ-έλλειψη που έχει τους άξονες της ευθυγραμμισμένους με τους κύριους άξονες.

Πίνακας Σ της μορφής $\sigma^2 I_D$

Σε αυτή την απλουστευμένη περίπτωση, ο πίνακας Σ είναι διαγώνιος και υποθέτουμε ότι η διακύμανση σε κάθε διάσταση είναι σταθερή, δηλαδή ισχύει $\Sigma = \text{diag}(\sigma^2 \dots \sigma^2) = \sigma^2 I_D$, όπου I_D είναι ένας $D \times D$ μοναδιαίος πίνακας. Οι συνολικές παράμετροι σε αυτή τη περίπτωση, μειώνονται σε $D + 1$. Επιπλέον, τα διανύσματα που προκαλούν την ίδια τιμή στην κανονική κατανομή, σχηματίζουν πλέον στις D διαστάσεις μια υπερ-σφαίρα με κέντρο μ και ακτίνα σ .

1.4. Μικτές κατανομές

Μια τυχαία μεταβλητή X ακολουθεί μια μικτή κατανομή, όταν η συνάρτηση πυκνότητας πιθανότητας της, ορίζεται ως ένας γραμμικός συνδυασμός γνωστών συναρτήσεων πυκνότητας πιθανότητας σταθμισμένων με βάρη. Στην ειδική περίπτωση που οι συναρτήσεις αυτές ανήκουν σε κάποια γνωστή οικογένεια κατανομών, προσδιορίζεται ένα μικτό μοντέλο της οικογένειας αυτής.

Ο ορισμός ενός μικτού μοντέλου με K το πλήθος συνιστώσες κατανομές περιγράφεται από την παρακάτω σχέση:

$$p(x; \theta) = \sum_{k=1}^K \pi_k f(x; \theta_k) \quad (1.13)$$

Όπου με $f(x; \theta_k)$ συμβολίζεται μια συνιστώσα κατανομή του μικτού μοντέλου και με θ_k το διάνυσμα των παραμέτρων της.

Τα π_k ονομάζονται παράμετροι μίξης και πρέπει να ικανοποιούν τους παρακάτω περιορισμούς:

$$0 \leq \pi_k \leq 1, \quad \forall k = 1 \dots K \quad (1.14)$$

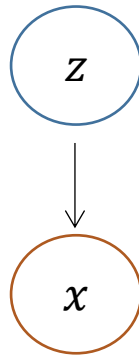
$$\sum_{k=1}^K \pi_k = 1 \quad (1.15)$$

Οι παραπάνω περιορισμοί είναι ταυτόσημοι με αυτούς που πρέπει να πληρούνται έτσι ώστε μια ποσότητα να θεωρείται πιθανότητα. Ακόμη, οι παράμετροι του μικτού μοντέλου ορίζονται ως:

$$\theta = \{(\pi_k, \theta_k), \quad k = 1 \dots K\} \quad (1.16)$$

Ποιά είναι όμως η διαδικασία με την οποία γεννάται μια παρατήρηση σύμφωνα με ένα μοντέλο μικτών κατανομών;

Η διαδικασία γένεσης μιας παρατήρησης, αποτελείται από δύο στάδια. Αρχικά επιλέγεται με πιθανότητα π_k μια συνιστώσα κατανομή. Στην συνέχεια, γίνεται δειγματοληψία από την $f(x; \theta_k)$ και προκύπτει η παρατήρηση. Συμπεραίνουμε λοιπόν πως τα βάρη π_k , μπορούν να ερμηνευτούν ως η εκ των προτέρων πιθανότητα, σύμφωνα με την οποία μια παρατήρηση γεννάται από την k -οστή συνιστώσα κατανομή του μικτού μοντέλου (βλέπε Σχήμα 1.2).



Σχήμα 1.2: Γραφικό μοντέλο αναπαράστασης ενός μοντέλου μικτών κατανομών. Η z είναι κρυμμένη μεταβλητή (αναπαριστά την συνιστώσα κατανομή που δημιουργεί το x) και η από κοινού συνάρτηση πυκνότητας πιθανότητας έχει την μορφή $p(x, z) = p(z)p(x|z) = \pi_k f(x; \theta_k)$. Μέσω περιθωριοποίησης ως προς την μεταβλητή z προκύπτει ο ορισμός του μοντέλου μικτών κατανομών.

Τα μοντέλα μικτών κατανομών αποτελούν μια ισχυρή μέθοδο για την εκτίμηση μιας συνάρτησης πυκνότητας πιθανότητας και αυτός είναι ο σκοπός που δημιουργήθηκαν. Τυγχάνουν όμως μεγάλης αποδοχής και εφαρμογής σε προβλήματα ομαδοποίησης παρατηρήσεων. Στην συνέχεια θα δούμε μια μέθοδο ομαδοποίησης που μπορεί μας εξασφαλίσει η εκτίμηση συνάρτησης πυκνότητας πιθανότητας με μοντέλα μικτών κατανομών. Βασική υπόθεση για τον σκοπό αυτό, είναι πως οι παρατηρήσεις που γεννούνται από την ίδια συνιστώσα κατανομή, σχηματίζουν μια ενιαία ομάδα.

Ας υποθέσουμε ότι έχουμε εκτελέσει μια διαδικασία μάθησης με την βοήθεια της οποίας προσδιορίστηκαν οι τιμές των παραμέτρων του μοντέλου μικτών κατανομών. Μπορούμε να υποθέσουμε, πως η εκ των προτέρων πιθανότητα π_k μιας συνιστώσας κατανομής, εκφράζει την αναλογία των παρατηρήσεων που παράγονται από την κατανομή αυτή σε σχέση με το σύνολο των παρατηρήσεων. Επιπλέον, μέσω της θεωρητικής σύνδεσης των συνιστωσών κατανομών και των ομάδων έχουμε βασικά στοιχεία για τον προσδιορισμό της κάθε ομάδας c_k (κέντρο, διακύμανση κ.α.). Τέλος, σημαντική προϋπόθεση για την διαδικασία εκπαίδευσης που θα ακολουθήσουμε αλλά και για την υπό ανάλυση μέθοδο ομαδοποίησης, είναι η δυνατότητα υπολογισμού των εκ των υστέρων πιθανοτήτων για μια τυχαία παρατήρηση να έχει παραχθεί από την k -οστή συνιστώσα κατανομή του μοντέλου. Μέσω εφαρμογής του θεωρήματος Bayes έχουμε:

$$p(k|x, \theta) = \frac{\pi_k f(x; \theta_k)}{\sum_{l=1}^K \pi_l f(x; \theta_l)} \quad (1.17)$$

Οι εκ των υστέρων πιθανότητες θα πρέπει από τον ορισμό τους να ικανοποιούν την σχέση:

$$\sum_{k=1}^K p(k|x, \theta) = 1 \quad (1.18)$$

Συμπεραίνουμε, πως μετά τον υπολογισμό των εκ των υστέρων πιθανοτήτων για όλες τις παρατηρήσεις, η ανάθεση μίας τυχαίας παρατήρησης x^i γίνεται στην ομάδα c_k με την μεγαλύτερη εκ των υστέρων πιθανότητα [1].

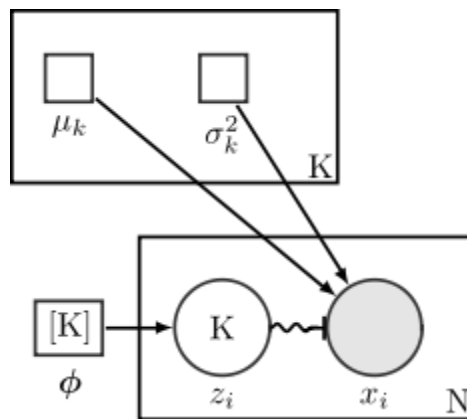
1.5. Μικτές κανονικές κατανομές

Στην περίπτωση που οι συνιστώσες κατανομές του μοντέλου μικτών κατανομών είναι συναρτήσεις πυκνότητας πιθανότητας κανονικών κατανομών, τότε ορίζεται ένα μοντέλο μικτών κανονικών κατανομών και στην μονοδιάστατη περίπτωση ισχύει:

$$p(x) = \sum_{k=1}^K \pi_k N(x|\mu_k, \sigma_k^2) \quad (1.19)$$

Το μοντέλο μικτών κανονικών κατανομών είναι γενικό και μπορεί να προσεγγίσει με ακρίβεια περίπλοκες κατανομές. Μια πολύ σημαντική ιδιότητα του είναι ότι παράγει συνεχείς και ολοκληρώσιμες συναρτήσεις, προσδίδοντας του καλές αναλυτικές ιδιότητες.

Ακόμη, σύμφωνα με το θεώρημα καθολικής προσέγγισης, έχει αποδειχτεί ότι κάτω από ορισμένες συνθήκες, το μοντέλο μπορεί να προσεγγίσει με όσο το δυνατόν μεγάλη ακρίβεια θέλουμε κάθε συνεχή συνάρτηση με πεπερασμένο πλήθος ασυνεχειών.



Σχήμα 1.3: Γραφικό μοντέλο μικτών κανονικών κατανομών, που αναπαριστά την διαδικασία της δημιουργίας των παρατηρήσεων μέσω εξαρτήσεων τυχαίων μεταβλητών. Με τα μικρά τετράγωνα συμβολίζονται οι παράμετροι του μοντέλου, με κύκλους (κόμβους) συμβολίζονται οι τυχαίες μεταβλητές και οι κατανομές τους, ενώ οι ακμές μεταξύ δυο κύκλων (κόμβων) δηλώνουν την αλληλεπίδραση των τυχαίων μεταβλητών. Τέλος, τα γεμισμένα σχήματα υποδηλώνουν τις παρατηρούμενες τιμές, ενώ ο δείκτης $[K]$, υποδηλώνει διάνυσμα μεγέθους K .

1.6. Μέθοδοι εκτίμησης παραμέτρων

1.6.1. Μέγιστη Πιθανοφάνεια

Σε αυτή την ενότητα θα παρουσιάσουμε μια θεμελιώδη μέθοδο εκτίμησης των παραμέτρων για τα στατιστικά μοντέλα που υποθέτουμε, ενώ στη συνέχεια θα γίνει εφαρμογή της μεθόδου στην κανονική κατανομή και στα μοντέλα μικτών κατανομών.

Έχοντας αρχικά αποφασίσει την μορφή της παραμετρικής συνάρτησης που θα χρησιμοποιήσουμε για την εκτίμηση της κατανομής των παρατηρήσεων, υπάρχει σαφής ανάγκη του προσδιορισμού κατάλληλων τιμών για τις παραμέτρους της. Μια αποδοτική και ευρέως διαδεδομένη μέθοδος είναι αυτή της “Μέγιστης Πιθανοφάνειας” (Maximum Likelihood). Στην διαδικασία εφαρμογής της μεθόδου, ορίζεται ως αντικειμενική συνάρτηση η συνάρτηση πιθανοφάνειας των παρατηρήσεων και στην συνέχεια επιστρατεύονται τεχνικές βελτιστοποίησης, για την μεγιστοποίηση της συνάρτησης αυτής.

Έστω ένα σύνολο παρατηρήσεων:

$$X = \{x_j^i\}, \quad i = 1 \dots N, \quad j = 1 \dots D$$

που παράχθηκε ανεξάρτητα και ισόνομα από μια άγνωστη συνάρτηση πυκνότητας πιθανότητας $p_0(x)$. Υποθέτουμε ότι η άγνωστη αυτή συνάρτηση, ανήκει στο παραμετρικό μοντέλο, οπότε ισχύει $p_0 = p(x; \theta_0)$. Το διάνυσμα παραμέτρων θ_0 είναι άγνωστο και αποτελεί την “αληθή” παραμετρική τιμή. Θα επιθυμούσαμε λοιπόν, να βρούμε μια εκτίμηση $\hat{\theta}$ του διανύσματος παραμέτρων, με την ιδιότητα να ελαχιστοποιεί το σφάλμα $\varepsilon = |\theta_0 - \hat{\theta}|$.

Για την εφαρμογή της μεθόδου μέγιστης πιθανοφάνειας, αρχικά υπολογίζουμε την από κοινού συνάρτηση πυκνότητας πιθανότητας των παρατηρήσεων:

$$P(x_1, x_2, \dots, x_n | \theta) = P(x_1 | \theta) P(x_2 | \theta) \dots P(x_n | \theta) \quad (1.20)$$

και ορίζουμε:

$$L(\theta; x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^N p(x_i | \theta) \quad (1.21)$$

Στην παραπάνω συνάρτηση, θεωρούμε ότι οι παρατηρήσεις $x_1, x_2, \dots, x_n$ αποτελούν σταθερές, ενώ αντίθετα, το διάνυσμα παραμέτρων θ αποτελεί τις άγνωστες μεταβλητές της συνάρτησης. Η $L(\theta|x_1, x_2, \dots, x_n)$ ονομάζεται συνάρτηση πιθανοφάνειας των παραμέτρων θ δοθέντος του συνόλου των παρατηρήσεων X .

Εκτιμητής μέγιστης πιθανοφάνειας των παραμέτρων ορίζεται η τιμή εκείνη που μεγιστοποιεί την ποσότητα $L(\theta|X)$ δηλαδή:

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} L(\theta|X) \quad (1.22)$$

Αντιλαμβανόμαστε, ότι η εκτίμηση $\hat{\theta}_{MLE}$ των παραμέτρων, θα οδηγήσει στην καλύτερη δυνατή περιγραφή των παρατηρήσεων. Για αναλυτικούς λόγους, είναι βολικότερο να δουλεύουμε με το λογάριθμο της συνάρτησης πιθανοφάνειας που ορίζεται από την σχέση:

$$LL(\theta|x_1, x_2, \dots, x_n) = \log \prod_{i=1}^N p(x_i|\theta) = \sum_{i=1}^N \log p(x_i|\theta) \quad (1.23)$$

Εφόσον ο λογάριθμος είναι γνησίως αύξουσα συνάρτηση, η τιμή $\hat{\theta}_{MLE}$ που μεγιστοποιεί την συνάρτηση της λογαριθμικής πιθανοφάνειας, μεγιστοποιεί επίσης και την συνάρτηση της πιθανοφάνειας.

Επεκτείνοντας την μέθοδο, μπορούμε να κάνουμε την υπόθεση ότι οι παράμετροι θ του μοντέλου μας ακολουθούν κάποια εκ των προτέρων κατανομή. Στην περίπτωση αυτή, αναζητούμε την μέγιστη εκ των υστέρων (MAP) εκτίμηση των παραμέτρων του μοντέλου, μεγιστοποιώντας την πιθανότητα των παραμέτρων θ δοθέντος του συνόλου των παρατηρήσεων X . Για την εκτίμηση MAP, εφαρμόζουμε το θεώρημα του Bayes και έχουμε:

$$P(\theta; x_1, x_2, \dots, x_n) = \frac{p(x_1, x_2, \dots, x_n|\theta)P(\theta)}{P(x_1, x_2, \dots, x_n)} \quad (1.24)$$

όπου $p(\theta)$ είναι η εκ των προτέρων κατανομή των παραμέτρων του μοντέλου. Ο παρονομαστής $P(x_1, x_2, \dots, x_n)$ είναι ανεξάρτητος από τις παραμέτρους οπότε ένας MAP εκτιμητής προκύπτει από την μεγιστοποίηση του αριθμητή:

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}} p(x_1, x_2, \dots, x_n | \theta) P(\theta) \quad (1.25)$$

Να σημειωθεί, ότι στην περίπτωση που η εκ των προτέρων κατανομή $p(\theta)$ είναι ομοιόμορφη, ο $\hat{\theta}_{MAP}$ ταυτίζεται με τον εκτιμητή μέγιστης πιθανοφάνειας:

$$\hat{\theta}_{MAP} = \hat{\theta}_{MLE} \quad (1.26)$$

Ολοκληρώνοντας την παρουσίαση της μεθόδου, πρέπει να αναλυθεί ο τρόπος βελτιστοποίησης της αντικειμενικής συνάρτησης για την εξαγωγή των παραμέτρων του μοντέλου. Η λογαριθμική πιθανοφάνεια είναι συνήθως παραγωγίσιμη ως προς το διάνυμα παραμέτρων θ και άρα το μέγιστο της μπορεί να υπολογιστεί σε κάποιες περιπτώσεις, με την βοήθεια γνωστών μεθόδων του διαφορικού λογισμού. Συνήθως οδηγούμαστε στην επίλυση ενός συστήματος εξισώσεων που προκύπτει από την σχέση:

$$\nabla LL(\theta|X) = \frac{\partial LL(\theta|X)}{\partial \theta} = 0 \quad (1.27)$$

Στην περίπτωση όπου η επίλυση του συστήματος, δεν γίνεται αναλυτικά και σε κλειστή μορφή, καταφεύγουμε σε αριθμητικές τεχνικές όπως είναι οι επαναληπτικές μέθοδοι βελτιστοποίησης.

1.6.2. Εφαρμογή της μέγιστης πιθανοφάνειας στην κανονική κατανομή

Εάν στην εξίσωση της λογαριθμικής πιθανοφάνειας, αντικαταστήσουμε την συνάρτηση πυκνότητας πιθανότητας της πολυδιάστατης κανονικής κατανομής, προκύπτει η παρακάτω σχέση:

$$LL(\theta|X) = \sum_{i=1}^N \left[-\frac{D}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right] \quad (1.28)$$

Παίρνοντας τις μερικές παραγώγους της σχέσης ως προς τις παραμέτρους μ και Σ και θέτοντας αυτές ίσες με το μηδέν, προκύπτουν οι εξισώσεις:

$$\frac{\partial LL(\theta|X)}{\partial \mu} = \sum_{i=1}^N \Sigma^{-1} (x_i - \mu) = 0 \quad (1.29)$$

$$\frac{\partial LL(\theta|X)}{\partial \Sigma} = \sum_{i=1}^N \left[-\frac{1}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-2} (x_i - \mu)(x_i - \mu)^T \right] = 0 \quad (1.30)$$

Επιλύοντας τις παραπάνω σχέσεις, έχουμε τις εκτιμήσεις για τις παραμέτρους μ και Σ :

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.31)$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T \quad (1.32)$$

Ο εκτιμητής για τη μέση τιμή, είναι ο δειγματικός μέσος όρος των παρατηρήσεων, ενώ αντίστοιχα για τον πίνακα συμμεταβλητότητας, είναι ο δειγματικός πίνακας συμμεταβλητότητας. Οι δυο αυτοί εκτιμητές ορίζουν και τον εκτιμητή μέγιστης πιθανοφάνειας των παραμέτρων του μοντέλου της πολυδιάστατης κανονικής κατανομής.

1.6.3. Εφαρμογή σε μικτές κατανομές

Σε εφαρμογές της στατιστικής μηχανικής μάθησης που υποθέτουν ένα μικτό μοντέλο κατανομών, μας ενδιαφέρει να προσδιορίσουμε την πιθανοτική συμμετοχή μιας παρατήρησης σε μια συνιστώσα κατανομή.

Όπως έχουμε δει, σε ένα μικτό μοντέλο με γνωστές παραμέτρους $\theta = (\pi_k, \theta_k)$, $k = 1 \dots K$, μπορούμε άμεσα να υπολογίσουμε την εκ των υστέρων πιθανότητα της κάθε παρατήρησης, δηλαδή την πιθανότητα να έχει παραχθεί από μια συνιστώσα

κατανομή, δοθέντων των παραμέτρων της. Στην πλειονότητα των περιπτώσεων όμως, οι παράμετροι Θ του μικτού μοντέλου είναι άγνωστοι και θα πρέπει να εκτιμηθούν.

Μια άμεση σκέψη είναι να χρησιμοποιήσουμε και σε αυτήν την περίπτωση την μέθοδο της μέγιστης πιθανοφάνειας, μεγιστοποιώντας την συνάρτηση:

$$LL(\theta|X) = \sum_{i=1}^N \log p(x_i; \theta) = \sum_{i=1}^N \log \sum_{k=1}^K \pi_k p(x_i; \theta_k) \quad (1.33)$$

ως προς τις παραμέτρους:

$$\theta = (\pi_k, \theta_k), \quad k = 1 \dots K.$$

Σε αντίθεση με την εκτίμηση μέγιστης πιθανοφάνειας για μια κατανομή, παρουσιάζονται κάποιες δυσκολίες. Στην παραπάνω συνάρτηση $LL(\theta|X)$, εξαιτίας του αθροίσματος που υπάρχει στον λογάριθμο, παρουσιάζεται υψηλή μη-γραμμικότητα που συνεπάγεται ύπαρξη πολλών τοπικών μεγίστων. Συμπερασματικά, στη μεγιστοποίηση της λογαριθμικής πιθανοφάνειας μικτών κατανομών, υπάρχουν θεωρητικά ζητήματα σχετικά με την μοναδικότητα του εκτιμητή μέγιστης πιθανοφάνειας, άρα και της επιλογής του κατάλληλου διανύσματος παραμέτρων για το μικτό μοντέλο.

Οι παραπάνω δυσκολίες του αλγορίθμου θα αντιμετωπιζόταν, εάν για κάθε παρατήρηση είχαμε την γνώση της συνιστώσας κατανομής που την δημιούργησε, μέσω μεταβλητών $z_i \in \mathbb{R}^K$ που υποδείκνυαν από ποια συνιστώσα κατανομή παράχθηκε η κάθε παρατήρηση x_i . Στην περίπτωση αυτή, η εκτίμηση των παραμέτρων θα ολοκληρωνόταν, εκτελώντας M ξεχωριστές διαδικασίες εκτίμησης μέγιστης πιθανοφάνειας. Τέλος, η ύπαρξη των μεταβλητών z_i , ανάγει την εκτίμηση των παραμέτρων ενός μοντέλου μικτών κατανομών, σε ένα πρόβλημα με κρυμμένες μεταβλητές, για την επίλυση του οποίου μπορεί να εφαρμοστεί αποδοτικά ο αλγόριθμος EM (Expectation Maximization) τον οποίο παρουσιάσουμε στην συνέχεια.

1.7. Βελτιστοποίηση μέσω του αλγορίθμου Expectation-Maximization (EM)

1.7.1. Ορισμός του αλγορίθμου - Εφαρμογή στα μοντέλα μικτών κατανομών

Ο αλγόριθμος Expectation-Maximization (EM), αποτελεί μια επαναληπτική μέθοδο εύρεσης εκτιμητών μέγιστης πιθανοφάνειας για τις παραμέτρους ενός στατιστικού μοντέλου, όταν αυτό εξαρτάται από κρυμμένες μεταβλητές [2]. Ξεκινώντας από μια αρχική εκτίμηση $\theta^{(0)}$ των παραμέτρων του μοντέλου, ο αλγόριθμος εκτελεί δύο βήματα. Αρχικά, στο Expectation step (E-βήμα), υπολογίζεται ένα τοπικό κάτω φράγμα της λογαριθμικής πιθανοφάνειας και μεγιστοποιείται ως προς την κατανομή των κρυμμένων μεταβλητών. Παρακάτω, θα δείξουμε ότι αυτό είναι ισοδύναμο με το να υπολογίσουμε τη εκ των υστέρων κατανομή των κρυμμένων μεταβλητών, δοθέντων των παρατηρήσεων και των τρεχουσών εκτιμήσεων των παραμέτρων. Στο Maximization step (M-βήμα), μεγιστοποιείται το κάτω φράγμα ως προς τις παραμέτρους θ , υποθέτοντας ότι η κατανομή των κρυμμένων μεταβλητών που βρέθηκε στο E-βήμα, είναι σωστή. Τα δύο βήματα επαναλαμβάνονται μέχρι να υπάρξει σύγκλιση σε κάποιο τοπικό μέγιστο της συνάρτησης πιθανοφάνειας.

Ο αλγόριθμος EM είναι ιδιαίτερα ερμηνεύσιμος και αποδοτικός στην εφαρμογή του σε μοντέλα μικτών κατανομών. Έστω ένα μοντέλο μικτών κατανομών με παραμέτρους $\theta = (\pi_k, \theta_k)$, $k = 1 \dots K$ και ένα σύνολο παρατηρήσεων

$$X = \{x_j^i\}, \quad i = 1 \dots N, \quad j = 1 \dots D$$

Το σύνολο X , αποκαλείται ελλιπές σύνολο παρατηρήσεων. Ορίζουμε ως πλήρες σύνολο το:

$$Y = (X, Z), \quad \text{όπου } Z = \{z_1, z_2, \dots, z_N\}, z_i \in \mathbb{R}^K \quad (1.34)$$

Το Z αποτελείται από τις N κρυμμένες (μη-παρατηρούμενες) μεταβλητές του μοντέλου z_i , οι οποίες υποδεικνύουν την συνιστώσα κατανομή από την οποία παράχθηκε η κάθε παρατήρηση x_i . Για ένα διάνυσμα z_i ισχύει $z_{ik} = 1$ όταν το x_i παράχθηκε από την k -συνιστώσα κατανομή ενώ οι υπόλοιπες θέσεις του διανύσματος είναι μηδενικές.

Από τον ορισμό αυτό προκύπτει ότι τα στοιχεία του συνόλου Z είναι σε ένα προς ένα αντιστοιχία με τα στοιχεία του συνόλου X . Μπορούμε λοιπόν να ορίσουμε την από κοινού κατανομή $P(X, Z|\Theta)$, μεταξύ των παρατηρούμενων και των κρυμμένων (μη-παρατηρούμενων) τιμών. Με τη βοήθεια αυτής της κατανομής, η συνάρτηση λογαριθμικής πιθανοφάνειας, για τα πλήρη δεδομένα, ορίζεται ως:

$$LLc(\theta|Y) = LLc(\theta|X, Z) = \log[p(X, Z; \theta)] \quad (1.35)$$

Επειδή το σύνολο Z αποτελείται από κρυμμένες μεταβλητές, η λογαριθμική πιθανοφάνεια LLc είναι ακαθόριστη, οπότε ο αλγόριθμος EM την λαμβάνει ως τυχαία μεταβλητή και υπολογίζει την αναμενόμενη τιμή της, ως προς την κατανομή $q(Z)$ των z_i , $i = 1 \dots N$. Η κατανομή $q(Z)$, όπως θα δείξουμε στη συνέχεια, είναι ίση με $p(Z|X, \Theta^{(t)})$, όπου $\Theta^{(t)}$ είναι η τρέχουσα τιμή των παραμέτρων.

Με δεδομένη τώρα την από κοινού κατανομή $p(X, Z; \theta)$, μπορούμε να επαναδιατυπώσουμε την λογαριθμική πιθανοφάνεια του μοντέλου μικτών κατανομών, με περιθωριοποίηση της κρυμμένης μεταβλητής Z :

$$LL(\theta|X) = \log p(X; \theta) = \log \sum_Z p(X, Z; \theta) \quad (1.36)$$

Η βασική ιδέα του τρόπου με τον οποίο γίνεται η βελτιστοποίηση μέσω του αλγορίθμου EM, είναι η κατασκευή ενός ευέλικτου κάτω φράγματος $G(\theta, q(Z))$ της συνάρτησης $LL(\theta|X)$, για το οποίο εξ'ορισμού θα ισχύει $G(\theta, q(Z)) \leq LL(\theta|X)$. Το κάτω φράγμα, θα έχει ως παραμέτρους τα θ και την κατανομή που ακολουθεί η κρυμμένη μεταβλητή Z . Ιδανικά, επιθυμούμε, ο υπολογισμός του G , σε αντίθεση με την πιθανοφάνεια $LL(\theta|X)$ να περιέχει άθροισμα λογαρίθμων, αντί για λογάριθμο αθροίσματος. Για να κατασκευή μιας τέτοιας ποσότητας, ξαναγράφουμε την λογαριθμική πιθανοφάνεια ως εξής:

$$LL(\theta|X) = \log \sum_Z q(Z) \frac{p(X, Z; \theta)}{q(Z)} \quad (1.37)$$

$$LL(\theta|X) = \log \sum_Z q(Z) \frac{p(X, Z; \theta)}{q(Z)}$$

όπου $q(Z)$ είναι προς το παρόν μια αυθαίρετη κατανομή για την τυχαία μεταβλητή Z , την οποία θα προσδιορίσουμε στην συνέχεια.

Εξαιτίας της κυρτότητας της συνάντησης του λογάριθμου, εφαρμόζουμε στην παραπάνω σχέση την ανισότητα του Jensen για υπολογίσουμε το κάτω φράγμα G και έχουμε:

$$\begin{aligned} LL(\theta|X) &= \log \sum_Z q(Z) \frac{p(X, Z; \theta)}{q(Z)} \geq \sum_Z q(Z) \log \frac{p(X, Z; \theta)}{q(Z)} \\ &= \sum_Z [q(Z) \log p(X, Z; \theta) - q(Z) \log q(Z)] = G(\theta, q(Z)) \end{aligned} \quad (1.38)$$

Η ανισότητα είναι αληθής για κάθε έγκυρη κατανομή $q(Z)$, παρόλα αυτά, εκείνο που ιδανικά επιθυμούμε, δεν είναι να κατασκευάσουμε ένα οποιοδήποτε κάτω φράγμα, αλλά ένα βέλτιστο (ή σφιχτό) κάτω φράγμα της λογαριθμικής πιθανοφάνειας. Πιο συγκεκριμένα, επιδιώκουμε να προσδιορίσουμε εκείνη την κατανομή $q(Z)$, με την ιδιότητα, το κάτω φράγμα $G(\theta, q(Z))$ να εφάπτεται στην συνάρτηση $LL(\theta|X)$ στην τρέχουσα εκτίμηση των παραμέτρων $\theta^{(t)}$ (Σχήμα 1.4), δηλαδή να ισχύει η σχέση:

$$G(\theta^{(t)}, q(Z)) = LL(\theta^{(t)}|X) \quad (1.39)$$

Η παραπάνω ιδιότητα εξασφαλίζει, ότι μετά από κάθε βελτιστοποίηση της συνάρτησης $G(\theta, q(Z))$ ως προς θ , θα βελτιστοποιείται ταυτόχρονα και η συνάρτηση $LL(\theta|X)$. Για να βρούμε λοιπόν το βέλτιστο κάτω φράγμα πρέπει να μεγιστοποιήσουμε την ποσότητα $G(\theta, q(Z))$ ως προς την άγνωστη κατανομή $q(Z)$. Για τον σκοπό αυτό, έχουμε:

$$\begin{aligned} G(\theta, q(Z)) &= \sum_Z q(Z) \log \frac{p(X, Z; \theta)}{q(Z)} = E_Z \left[\log \frac{p(X, Z; \theta)}{q(Z)} \right] = \\ &= E_Z \left[\log \frac{p(Z|X, \theta)p(X; \theta)}{q(Z)} \right] = E_Z \left[\log \frac{p(Z|X, \theta)}{q(Z)} + \log p(X; \theta) \right] = \\ &= E_Z \left[\log \frac{p(Z|X, \theta)}{q(Z)} \right] + E_Z [\log p(X; \theta)] \\ &= -E_Z \left[\log \frac{q(Z)}{p(Z|X, \theta)} \right] + \log p(X; \theta) = \\ &= -D[q(Z)||p(Z|X, \theta)] + LL(\theta|X) \end{aligned} \quad (1.40)$$

Υποθέτοντας τώρα ότι η $q(Z)$ είναι συνάρτηση πυκνότητας πιθανότητας κάποιας άγνωστης κατανομής, η ποσότητα $D[q(Z)||p(Z|X, \theta)]$ αναπαριστά την απόσταση Kullback-Leibler μεταξύ των κατανομών $q(Z)$ και $p(Z|X, \theta)$. Η απόσταση αυτή είναι εξ' ορισμού μη αρνητική και είναι ίση με μηδέν, όταν οι δύο κατανομές είναι ίσες. Από αυτό συνεπάγεται ότι η ποσότητα $G(\theta, q(Z))$ μεγιστοποιείται ως προς $q(Z)$ όταν ισχύει:

$$D[q(Z)||p(Z|X, \theta)] = 0 \Leftrightarrow q(Z) = p(Z|X, \theta) \quad (1.41)$$

Από την τελευταία σχέση, είναι εύκολο να αποδείξουμε ότι όταν ισχύει η ισότητα, η συνάρτηση του βέλτιστου (σφιχτού) κάτω φράγματος $G(\theta, q(Z))$, εφάπτεται με την συνάρτηση της λογαριθμικής πιθανοφάνειας στην τρέχουσα τιμή των παραμέτρων θ .

Η παραπάνω διαδικασία προσδιορισμού της κατανομής $q(Z)$ από την οποία προκύπτει το βέλτιστο κάτω φράγμα με δεδομένη την τρέχουσα τιμή των παραμέτρων $\theta^{(t)}$, αποτελεί το E-βήμα του αλγορίθμου.

Για την βελτιστοποίηση των παραμέτρων, δηλαδή για την εκτέλεση του M-βήματος του αλγορίθμου, πρέπει να μεγιστοποιήσουμε το κάτω φράγμα $G(\theta, q(Z))$ ως προς το διάνυσμα παραμέτρων θ , έχοντας δεδομένη την κατανομή $q(Z)$ που προσδιορίσαμε στο E-βήμα. Η διαδικασία αυτή εξαρτάται από το υπό μελέτη πρόβλημα και η επίλυση γίνεται αρκετές φορές με αναλυτικό τρόπο. Από την σχέση:

$$G(\theta, q(Z)) = \sum_Z [q(Z) \log p(X, Z; \theta) - q(Z) \log q(Z)] \quad (1.42)$$

παρατηρούμε ότι για να πετύχουμε το ζητούμενο, αρκεί να μεγιστοποιήσουμε ως προς θ την ποσότητα:

$$\begin{aligned} \sum_Z q(Z) \log p(X, Z; \theta) &= \sum_Z p(Z|X, \theta) \log p(X, Z; \theta) \\ &= E_{Z|X, \theta}[\log p(X, Z; \theta)] \end{aligned} \quad (1.43)$$

Συμπερασματικά, ας υποθέσουμε ότι οι τρέχουσες καλύτερες εκτιμήσεις των παραμέτρων θ και q είναι οι:

$$\theta^{(t)}, \theta^{(t+1)}, q^{(t)}, q^{(t+1)}$$

Στην περίπτωση αυτή, ο αλγόριθμος EM, διατυπώνεται συνοπτικά ως εξής:

E-βήμα:

$$q^{(t+1)}(Z) = \arg \max_q G(\theta^{(t)}, q(Z)) = p(Z|X, \theta^{(t)})$$

M-βήμα:

$$\theta^{(t+1)} = \arg \max_{\theta} G(\theta, q^{(t+1)}(Z)) = \arg \max_{\theta} E_{Z|X, \theta^t}[\log p(X, Z; \theta)]$$

Μια γραφική αναπαράσταση ενός μοναδικού βήματος του αλγορίθμου EM φαίνεται και στο Σχήμα 1.4.

1.7.2. Σύγκλιση του αλγορίθμου EM

Σε αυτήν την παράγραφο θα μας απασχολήσει η σύγκλιση του αλγορίθμου EM σε μια γενική προσπάθεια απόδειξης. Δεν θα επικεντρωθούμε σε ιδιότητες της σύγκλισης, για τις οποίες μπορεί κάποιος να αντλήσει πληροφορίες από το βιβλίο των McLachlan και Krishnan [3].

Υποθέτουμε, ότι $\theta^{(t)}$ και $\theta^{(t+1)}$ είναι εκτιμήσεις των παραμέτρων του μοντέλου, που προέκυψαν από δύο διαδοχικά βήματα του αλγορίθμου EM. Από την στιγμή που το $\theta^{(t+1)}$ έχει την ιδιότητα να μεγιστοποιεί την συνάρτηση του κάτω φράγματος G , εύκολα προκύπτει η σχέση:

$$G(\theta^{(t+1)}, q(Z)) \geq G(\theta^{(t)}, q(Z)) \quad (1.44)$$

και επειδή το G είναι ένα βέλτιστο (σφιχτό) κάτω φράγμα της λογαριθμικής πιθανοφάνειας ισχύει ότι:

$$LL(\theta^{(t+1)}|X) \geq G(\theta^{(t+1)}, q(Z)) \quad (1.45)$$

Επίσης, εξ' ορισμού ισχύει και η σχέση:

$$G(\Theta^{(t)}, q(Z)) = LL(\Theta^{(t)}|X) \quad (1.46)$$

Οπότε συνδυάζοντας όλες τις παραπάνω σχέσεις, έχουμε τελικά:

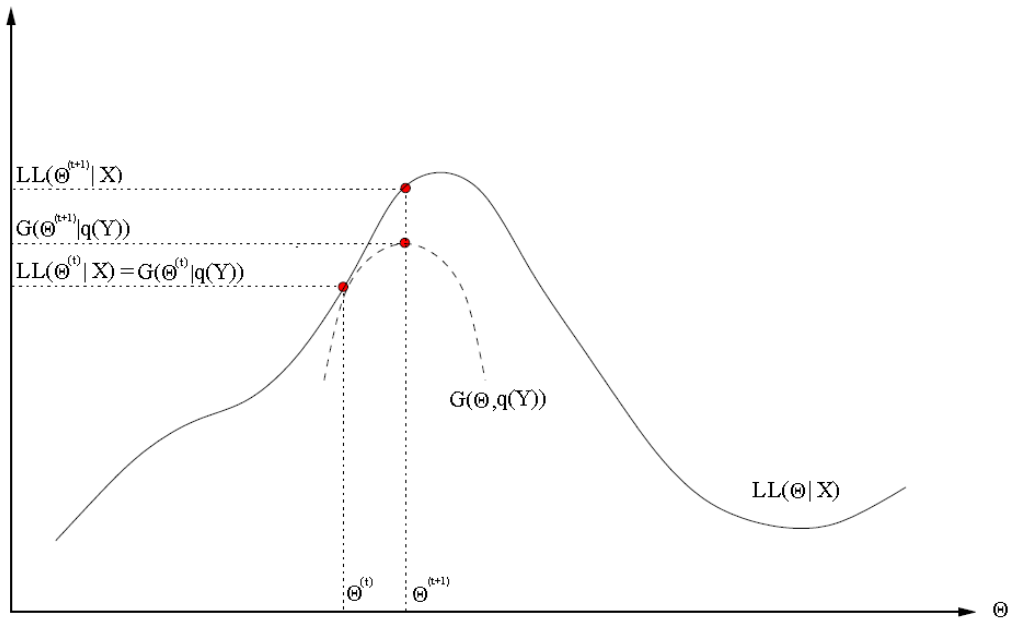
$$LL(\Theta^{(t+1)}|X) \geq LL(\Theta^{(t)}|X) \quad (1.47)$$

Η τελευταία σχέση αποδεικνύει ότι η πιθανοφάνεια αυξάνεται σε κάθε επανάληψη. Πιο γενικά, τα δύο βήματα του αλγορίθμου επαναλαμβάνονται, μέχρι να βρεθεί κάποιο τοπικό μέγιστο και η (αύξουσα) ακολουθία των τιμών της πιθανοφάνειας που παράγονται σε κάθε βήμα, συγκλίνει. Στην περίπτωση αυτή ο αλγόριθμος τερματίζει σε τοπικό μέγιστο της πιθανοφάνειας.

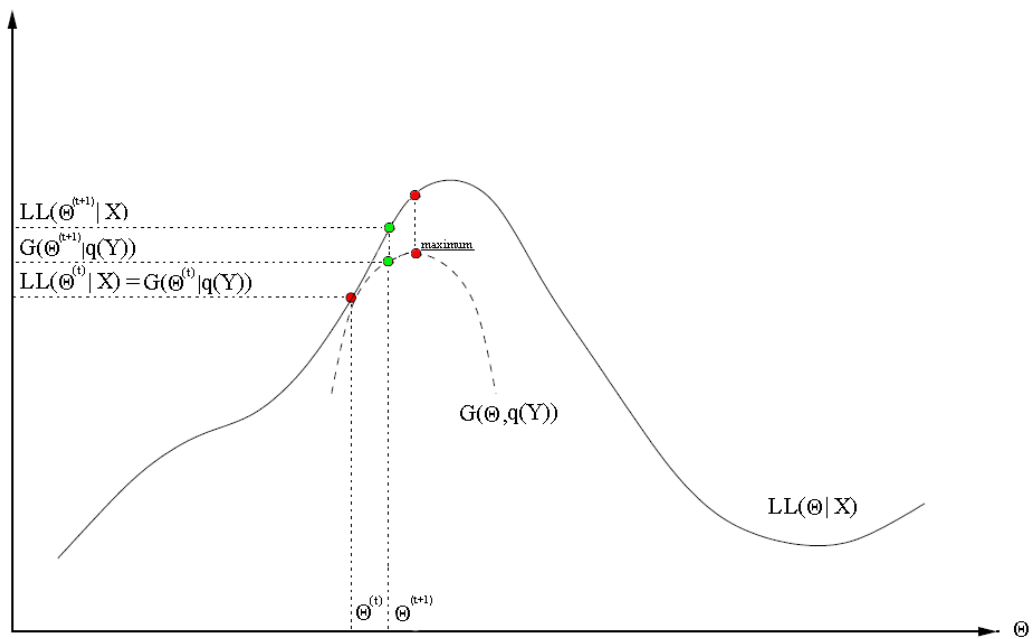
1.7.3. Ο αλγόριθμος *Generalized Expectation Maximization (GEM)*

Όπως είδαμε στην παρουσίαση του αλγορίθμου EM, στο M-βήμα του αλγορίθμου η νέα εκτίμηση των παραμέτρων $\Theta^{(t+1)}$ επιλέγεται έτσι ώστε να μεγιστοποιεί το κάτω φράγμα $G(\Theta|q^{(t+1)}(Z))$. Εφόσον η αύξηση της συνάρτησης του κάτω φράγματος, συνεπάγεται αύξηση και στην συνάρτηση της πιθανοφάνειας, θα μπορούσαμε να χαλαρώσουμε το προηγούμενο κριτήριο και αντί να απαιτούμε να βρεθεί το μέγιστο της συνάρτησης του κάτω φράγματος G ως προς Θ , να επιδιώκουμε απλώς να την αυξήσουμε. Με αυτό τον τρόπο, επιτυγχάνουμε και πάλι το ζητούμενο, δηλαδή αύξηση στην συνάρτηση της πιθανοφάνειας.

Αυτή η προσέγγιση είναι γνωστή ως αλγόριθμος GEM (Generalized Expectation Maximization). Ο GEM είναι πολύ χρήσιμος σε περιπτώσεις όπου η μεγιστοποίηση του κάτω φράγματος G είναι δύσκολη ή δεν υπάρχει λύση διατυπωμένη σε κλειστή μορφή. Σε αυτή την περίπτωση εφαρμόζονται μέθοδοι αριθμητικής βελτιστοποίησης, η χρήση των οποίων δεν οδηγεί πάντα στην εύρεση ολικού μεγίστου, αλλά σε μια καλύτερη εκτίμηση από το αρχικό σημείο εκκίνησης. Η απόδειξη της σύγκλισης του αλγορίθμου αυτού είναι παρόμοια με αυτή του αλγορίθμου EM και γι' αυτό παραλείπεται. Ο τρόπος με τον οποίο λειτουργεί ο αλγόριθμος για μια επανάληψη, παρουσιάζεται στο Σχήμα 1.5.



Σχήμα 1.4 Απεικόνιση ενός βήματος του EM. Η συνάρτηση G , αποτελεί το κάτω φράγμα της συνάρτησης πιθανοφάνειας. Οι δύο συναρτήσεις είναι ίσες στο σημείο $\Theta^{(t)}$. Στο M-βήμα το $\Theta^{(t+1)}$ ορίζεται ως η τιμή του Θ που μεγιστοποιεί το G .



Σχήμα 1.5 Ο τρόπος λειτουργίας του αλγόριθμου GEM για μια επανάληψη. Παρατηρούμε ότι παρόλο που δεν υπολογίζεται το μέγιστο του κάτω φράγματος G , η πιθανοφάνεια αυξάνεται.

1.7.4. Εφαρμογή του αλγορίθμου EM σε μικτές κανονικές κατανομές

Μέχρι τώρα ο αλγόριθμος EM παρουσιάστηκε περισσότερο σε θεωρητικό επίπεδο. Σε αυτή την παράγραφο θα εξετάσουμε τον τρόπο με τον οποίο μπορεί να εφαρμοστεί στην εκπαίδευση ενός μοντέλου μικτών κανονικών κατανομών.

Έστω ότι έχουμε ένα σύνολο παρατηρήσεων που παράχθηκαν ανεξάρτητα από μία κανονική κατανομή $N(x|\mu_k, \Sigma_k)$ σταθμισμένη με βάρη π_k :

$$X = \{x_j^i\}, \quad i = 1 \dots N, \quad j = 1 \dots D$$

Ο ορισμός ενός μικτού μοντέλου K κανονικών κατανομών όπως διατυπώθηκε στην παράγραφο 1.5 είναι:

$$p(x) = \sum_{k=1}^K \pi_k N(x|\mu_k, \Sigma_k) \quad (1.48)$$

$$0 \leq \pi_k \leq 1, \quad \sum_{k=1}^K \pi_k = 1$$

Τόσο οι παράμετροι π_k , όσο και οι παράμετροι $\theta_k = \{\mu_k, \sigma_k^2\}$, $k = 1 \dots K$ για κάθε συνιστώσα κατανομή, πρέπει να εκτιμηθούν. Υπενθυμίζεται ότι για ένα διάνυσμα z_i , ισχύει $z_{ik} = 1$ όταν το x_i παράχθηκε από την k -συνιστώσα κατανομή ενώ οι υπόλοιπες θέσεις του διανύσματος είναι μηδενικές.

Η πλήρης λογαριθμική πιθανοφάνεια του μοντέλου, έχει την μορφή:

$$\log P(X, Z) = \sum_{i=1}^N \sum_{k=1}^K z_{ik} \log[\pi_k N(\mu_k, \Sigma_k)] \quad (1.49)$$

Ε-Βήμα

Σύμφωνα με την σχέση (1.42) στο Ε-βήμα πρέπει να υπολογίσουμε την $p(Z|X, \theta^{(t)})$, δηλαδή την κατανομή των μη παρατηρούμενων μεταβλητών, δοθέντων του συνόλου των παρατηρήσεων και των τρεχουσών εκτιμήσεων των παραμέτρων. Ισοδύναμα, μπορούμε να υπολογίσουμε τις εκ των υστέρων πιθανότητες για κάθε παρατήρηση x_i να έχει παραχθεί από κάθε μια από τις K συνιστώσες κατανομές:

$$p(z_{ik} = 1|x_i) = \frac{\pi_k N(x_i|z_{ik} = 1)}{p(x_i)} = \frac{\pi_k N(x_i|\mu_k, \Sigma_k)}{\sum_{l=1}^K \pi_l N(x_i|\mu_l, \Sigma_l)} \quad (1.50)$$

Μ-βήμα

Σύμφωνα πάλι με την σχέση (1.42), στο Μ-βήμα θα πρέπει να μεγιστοποιήσουμε την ποσότητα:

$$E_{Z|X, \theta^t} [\log p(X, Z; \theta)] \quad (1.51)$$

ως προς τις παραμέτρους:

$$\theta = \{\pi_k, (\mu_k, \Sigma_k)\}, k = 1 \dots K$$

Στην περίπτωση των μικτών μοντέλων κανονικών κατανομών αυτό μπορεί να πραγματοποιηθεί αναλυτικά σε κλειστή μορφή. Η αναμενόμενη πλήρη λογαριθμική πιθανοφάνεια με δεδομένη την εκ των υστέρων κατανομή των παρατηρήσεων είναι:

$$\begin{aligned} E_Z [\log p(X, Z; \theta)] &= \sum_{i=1}^N \sum_{k=1}^K p(z_{ik} = 1|x_i) (\log \pi_k + \log N(\mu_k, \Sigma_k)) = \\ &= \sum_{i=1}^N \sum_{k=1}^K p(z_{ik} = 1|x_i) \log N(\mu_k, \Sigma_k) + \sum_{i=1}^N \sum_{k=1}^K p(z_{ik} = 1|x_i) \log \pi_k \end{aligned} \quad (1.52)$$

Η εκ των υστέρων πιθανότητα $p(z_{ik} = 1|x_i)$ έχουν ήδη υπολογιστεί στο Ε-βήμα, και άρα η παραπάνω αντικειμενική συνάρτηση (1.52) είναι ένα άθροισμα δύο ασυσχέτιστων μεταξύ τους όρων. Ο πρώτος εμπεριέχει την συνάρτηση πυκνότητας πιθανότητας της κάθε συνιστώσας κανονικής κατανομής και ο δεύτερος την αντίστοιχη εκ των προτέρων πιθανότητα. Άρα, σε κάθε επανάληψη του αλγορίθμου, για την εύρεση των παραμέτρων της κάθε συνιστώσας κανονικής κατανομής, θα πρέπει να μεγιστοποιήσουμε τον πρώτο όρο, ενώ για την εκ των προτέρων πιθανότητα της κάθε κατανομής, θα μεγιστοποιήσουμε τον δεύτερο. Εφαρμόζοντας τα παραπάνω προκύπτουν οι παρακάτω τύποι ενημέρωσης των παραμέτρων του μοντέλου:

$$\pi_k^{new} = \frac{1}{N} \sum_{i=1}^N p(z_{ik} = 1|x_i) \quad (1.53)$$

$$\mu_k^{new} = \frac{\sum_{i=1}^N x_i p(z_{ik} = 1|x_i)}{\sum_{i=1}^N p(z_{ik} = 1|x_i)} \quad (1.54)$$

$$\Sigma_k^{new} = \frac{\sum_{i=1}^N p(z_{ik} = 1|x_i)(x_i - \mu_j^{new})(x_i - \mu_j^{new})^T}{\sum_{i=1}^N p(z_{ik} = 1|x_i)} \quad (1.55)$$

Παρατηρούμε ότι οι δύο τελευταίες εξισώσεις είναι παρόμοιες με αυτές που χρησιμοποιούνται στην εκτίμηση των παραμέτρων μιας μοναδικής κανονικής κατανομής. Η μοναδική διαφορά είναι ότι στο μοντέλο μικτών κανονικών κατανομών, η κάθε παρατήρηση x_i , είναι σταθμισμένη με την εκ των υστέρων πιθανότητα $p(z_{ik} = 1|x_i)$.

Μέσω των αναλυτικών μορφών ενημέρωσης των παραμέτρων και έπειτα από κάθε επανάληψη του αλγορίθμου EM, έχει αποδειχτεί η αύξηση της τιμής στην συνάρτηση πιθανοφάνειας του μοντέλου μικτών κατανομών.

1.7.5. Σχολιασμός του αλγορίθμου EM

Τα πλεονεκτήματα του EM είναι η απλότητα του, η καλή απόδοση του σε προβλήματα που περιλαμβάνουν κρυμμένες μεταβλητές καθώς και η σίγουρη σύγκλιση του σε κάποιο (τοπικό) μέγιστο. Επίσης, στα μοντέλα μικτών κατανομών έχουμε άμεση ικανοποίηση των παρακάτω περιορισμών:

$$0 \leq \pi_k \leq 1, \quad \sum_{k=1}^K \pi_k = 1$$

καθώς επίσης και ότι ο πίνακας συμμεταβλητότητας που παράγεται από την σχέση (1.55) στα μοντέλα μικτών κανονικών κατανομών, είναι συμμετρικός και θετικά ημισορισμένος.

Ένα βασικό μειονέκτημα του αλγορίθμου είναι η μεγάλη εξάρτησή του από τις αρχικές τιμές των παραμέτρων. Στην περίπτωση που η αρχικοποίηση είναι τυχαία, ή πολύ διαφορετική από την τελική επιθυμητή λύση, ο αλγόριθμος δεν επιστρέφει ικανοποιητικές τιμές για τις παραμέτρους του μοντέλου. Για την εξασφάλιση

κατάλληλων αρχικών τιμών, συνήθως χρησιμοποιείται ένας ανεξάρτητος αλγόριθμος που επιστρέφει μια αποδεκτή λύση, στην οποία θα αρχικοποιηθεί ο EM, για να παράγει την τελική εκτίμηση των παραμέτρων του μοντέλου. Αναφορικά με την ταχύτητα σύγκλισης θα λέγαμε ότι ο αλγόριθμος σε γενικές γραμμές αυξάνει με γρήγορους ρυθμούς την πιθανοφάνεια στα πρώτα του βήματα, ενώ η σύγκλιση του γίνεται αισθητά πιο αργή καθώς πλησιάζει σε ένα τοπικό μέγιστο. Πιο αναλυτικά [1], σε σύγκριση με τον αλγόριθμο K-μέσων (K-means) ο αλγόριθμος EM χρειάζεται περισσότερες επαναλήψεις μέχρι να συγκλίνει, ενώ η κάθε επανάληψη του αλγορίθμου είναι αρκετά πιο δαπανηρή από την αντίστοιχη του αλγορίθμου K-μέσων. Για την αντιμετώπιση τα παραπάνω ζητημάτων, ο αλγόριθμος K-μέσων συνήθως εφαρμόζεται για την αρχικοποίηση ενός μοντέλου μικτών κανονικών κατανομών.

Τέλος, ένα πολύ σημαντικό ζήτημα που θα μας απασχολήσει στο κλείσιμο της παρουσίασης του EM αλγορίθμου αλλά και της εισαγωγής συνολικά, είναι το πρόβλημα του προσδιορισμού του πλήθους των συνιστωσών κατανομών ενός μοντέλου μικτών κατανομών, που επηρεάζει συνολικά την απόδοση του αλγορίθμου EM.

1.8. Το κριτήριο dip

Η εκτίμηση του πλήθους των συνιστωσών κατανομών για ένα μοντέλο μικτών κατανομών είναι ένα θεμελιώδες πρόβλημα στην στατιστική μηχανική μάθηση ενώ συνδέεται άμεσα και με το πρόβλημα της ομαδοποίησης. Η μέθοδος που θα εφαρμόσουμε για την προσέγγιση του βέλτιστου πλήθους συνιστωσών κρίνεται θεμελιώδης για να πετύχουμε μια αποδοτική και ακριβή εκτίμηση των παραμέτρων του μοντέλου μέσω του EM αλγορίθμου.

Στην παρούσα εργασία, χρησιμοποιούμε το κριτήριο dip του Hartigan's [4], με το οποίο ελέγχεται η ιδιότητα της μονοκόρυφης (unimodal) συνάρτησης. Μέσω της ιδιότητας αυτής, προσδιορίζεται εάν η εμπειρική συνάρτηση κατανομής που προκύπτει από ένα σύνολο παρατηρήσεων X , έχει μία, ή περισσότερες κορυφές και θα χρησιμοποιηθεί για τον προσδιορισμό των συνιστωσών κατανομών ενός μικτού μοντέλου.

Στην συνέχεια παρουσιάζονται κάποιοι ορισμοί για την ανάλυση του κριτηρίου, ενώ στα επόμενα κεφάλαια γίνεται η εφαρμογή του για την εκτίμηση του πλήθους των συνιστωσών κατανομών ενός μικτού μοντέλου.

Έστω ένα σύνολο μονοδιάστατων παρατηρήσεων $X = \{x^i\}$, $i = 1 \dots N$. Ο αλγόριθμος που εκτελεί το κριτήριο dip , εφαρμόζεται στην *εμπειρική αθροιστική συνάρτηση κατανομής* (*empirical cdf*) που ορίζεται από την σχέση:

$$F_n(t) = \frac{1}{n} \sum_n I(x^i \leq t) \quad (1.56)$$

και αποτελεί μια μη-παραμετρική συνάρτηση που περιγράφει τα δεδομένα μιας άγνωστης αθροιστικής συνάρτησης κατανομής (cdf), δίχως να ανατρέξει σε παραμετρικούς συντελεστές εκτίμησης. Υποθέτει ότι η μορφή της κατανομής που περιγράφει το δείγμα των παρατηρήσεων μιας τυχαίας μεταβλητής είναι άγνωστη και ότι οι παρατηρήσεις είναι ανεξάρτητες.

Μια συνάρτηση $F(t)$ είναι *μονοκόρυφη* (*unimodal*) με κορυφή το διάστημα $s_m = \{(t_L, t_U) : t_L \leq t_U\}$ εάν είναι κυρτή στο $s_L = (-\infty, t_L]$, σταθερή στο $[t_L, t_U]$ και κοίλη στο διάστημα $s_U = [t_U, \infty)$. Ο ορισμός αυτός εξασφαλίζει, ότι καθώς απομακρυνόμαστε από την κορυφή, η συνάρτηση πυκνότητας πιθανότητας δεν αυξάνεται.

Έστω δύο φραγμένες συναρτήσεις F, G και έστω:

$$\rho(F, G) = \max_t |F(t) - G(t)| \quad (1.57)$$

$$\rho(F, \mathcal{A}) = \min_{G \in \mathcal{A}} \rho(F, G) \quad (1.58)$$

όπου με $\mathcal{A}$ αναπαριστούμε κάθε σύνολο φραγμένων συναρτήσεων. Έστω τώρα $\mathcal{U}$, το σύνολο όλων των μονοκόρυφων (*unimodal*) συναρτήσεων. Το κριτήριο dip για μια συνάρτηση κατανομής F ορίζεται ως εξής:

$$dip(F) = \min_{G \in \mathcal{U}} \rho(F, G) \quad (1.59)$$

Για το dip , ισχύουν οι παρακάτω ιδιότητες:

$$\text{dip}(F_1) \leq \text{dip}(F_2) + \rho(F_1, F_2) \quad (1.60)$$

$$\text{dip}(F) = 0, \text{ για } F \in \mathcal{U}, \text{ ενώ } \text{dip}(F) > 0, \text{ όταν } F \notin \mathcal{U} \quad (1.61)$$

Από την τελευταία σχέση συμπεραίνουμε ότι το dip αναπαριστά την απόκλιση της F από την ιδιότητα της μονοκόρυφης συνάρτησης. Ακόμα, ορίζουμε

- ως μέγιστο κυρτό minorant (g.c.m.) της συνάρτησης F στο διάστημα $(-\infty, a]$, το $\sup G(x)$ για $x \leq a$, όπου το $\sup$ αντιστοιχεί σε κάθε κυρτή συνάρτηση G στο $(-\infty, a]$ που δεν είναι πουθενά μεγαλύτερη από την F .
- ως ελάχιστο κοίλο minorant (l.c.m.) της συνάρτησης F στο διάστημα $[a, \infty)$, το $\inf L(x)$ για $x \geq a$, όπου το $\inf$ αντιστοιχεί σε κάθε κοίλη συνάρτηση L στο $[a, \infty)$ που δεν είναι πουθενά μικρότερη από την F .

Συμπερασματικά, το κριτήριο dip υπολογίζει αρχικά τις μέγιστες αποκλίσεις μεταξύ μιας συνάρτησης κατανομής F και των συναρτήσεων κατανομών από το σύνολο $\mathcal{U}$, και τελικά παίρνει την τιμή της ελάχιστης απόκλισης.

Μια χρήσιμη ιδιότητα του dip , είναι ότι εάν η F_n αποτελεί μια δειγματική κατανομή n -παρατηρήσεων από την F , τότε ισχύει:

$$\lim_{n \rightarrow \infty} \text{dip}(F_n) = \text{dip}(F) \quad (1.62)$$

Επίσης, σύμφωνα με [4],[7], το σύνολο των ομοιόμορφων κατανομών είναι το καταλληλότερο για την μηδενική υπόθεση του κριτηρίου, καθώς οι τιμές του κριτηρίου dip για το σύνολο ομοιόμορφων κατανομών, είναι στοχαστικά μεγαλύτερες από άλλες unimodal κατανομές.

Για ένα σύνολο μονοδιάστατων παρατηρήσεων $X = \{x^i\}$, $i = 1 \dots N$, ο αλγόριθμος που εκτελεί το κριτήριο dip , συνοψίζεται στα παρακάτω βήματα. Αρχικά λαμβάνει ως είσοδο την αντίστοιχη εμπειρική συνάρτηση κατανομής πιθανότητας. Στη συνέχεια, ελέγχει τα $\frac{n(n-1)}{2}$ πιθανά διαστήματα κορυφών $[t_L, t_U]$ μεταξύ των n ανεξάρτητων και ταξινομημένων παρατηρήσεων. Για όλους τους συνδυασμούς, υπολογίζει σε $O(n)$ χρόνο τις αντίστοιχες μέγιστες κυρτές minorants και ελάχιστες κοίλες majorants

καμπύλες στα διαστήματα $(\min_t F_n, t_L)$ και $(t_U, \max_t F_n)$. Για κάθε εμπειρική συνάρτηση κατανομής F_n η πολυπλοκότητα μια εκτέλεσης του κριτηρίου dip , είναι $O(n)$.

Τέλος, για τον υπολογισμό της p-value του κριτηρίου χρησιμοποιούνται στατιστικοί πίνακες. Η τιμή της p-value, εκφράζει την πιθανότητα το $\text{dip}(F_n)$ να είναι μικρότερο από τις τιμές $\text{dip}(U_n^r)$, όπου U_n^r είναι μια συνάρτηση κατανομής n παρατηρήσεων, που δημιουργήθηκαν από την ομοιόμορφη κατανομή $U[0,1]$:

$$P = \#[\text{dip}(F_n) \leq \text{dip}(U_n^r)]/b, \quad r = 1, \dots, b \quad (1.63)$$

Η μηδενική υπόθεση του τεστ: H_0 : “Η F_n είναι unimodal”, γίνεται αποδεκτή με επίπεδο στατιστικής σημαντικότητας α , όταν ισχύει $p\text{-value} > \alpha$. Στην αντίθετη περίπτωση, η H_0 απορρίπτεται και δεχόμαστε την εναλλακτική υπόθεση H_1 : “Η F_n είναι multimodal”.

Στα επόμενα κεφάλαια θα εφαρμόσουμε αποδοτικά το κριτήριο dip του Hartigan’s για τον προσδιορισμό του πλήθους των συνιστωσών κατανομών στα μικτά μοντέλα συμμετρικών και ασύμμετρων Π-σιγμοειδών κατανομών.

1.9. Διάρθρωση και συνεισφορά της εργασίας

Η παρούσα εργασία αποτελείται από πέντε κεφάλαια. Στο πρώτο κεφάλαιο παρουσιάζεται το πρόβλημα της στατιστικής μοντελοποίησης που συνίσταται στην εκτίμηση της συνάρτησης πυκνότητας πιθανότητας ενός συνόλου παρατηρήσεων. Επίσης παρουσιάζεται η θεωρητική βάση που απαιτήθηκε για την υλοποίηση της διατριβής. Πιο συγκεκριμένα, γίνεται αναφορά σε μεθόδους εκτίμησης παραμέτρων όπως η μέθοδος μέγιστης πιθανοφάνειας (MLE) και ο αλγόριθμος βελτιστοποίησης Expectation-Maximization (EM). Στο τέλος του κεφαλαίου παρουσιάζεται το κριτήριο dip .

Στο δεύτερο κεφάλαιο γίνεται παρουσίαση της Π-σιγμοειδούς κατανομής η οποία ορίζεται ως διαφορά δύο μετατοπισμένων λογιστικών σιγμοειδών συναρτήσεων που έχουν την ίδια κλίση. Στη συνέχεια, αναφέρονται οι στατιστικές ιδιότητες της

κατανομής και περιγράφεται το μικτό μοντέλο Π-σιγμοειδών κατανομών. Στο τέλος του κεφαλαίου, προτείνεται μια μεθοδολογία καθορισμού του πραγματικού αριθμού των συνιστωσών του μικτού μοντέλου Π-σιγμοειδών κατανομών με χρήση του στατιστικού κριτηρίου dip .

Στο τρίτο κεφάλαιο προτείνεται μια επέκταση της Π-σιγμοειδούς κατανομής που ονομάζεται ασύμμετρη Π-σιγμοειδής κατανομή. Μελετώνται οι ιδιότητες της κατανομής, ορίζεται ένα μικτό μοντέλο από ασύμμετρες Π-σιγμοειδείς κατανομές και προτείνεται ένας αλγόριθμος εκπαίδευσης βασισμένος στον αλγόριθμο GEM. Τέλος, προτείνεται μια μεθοδολογία καθορισμού του πραγματικού αριθμού των συνιστωσών του μικτού μοντέλου ασύμμετρων Π-σιγμοειδών κατανομών με χρήση του στατιστικού κριτηρίου dip .

Στο τέταρτο κεφάλαιο γίνεται πειραματική αξιολόγηση της επίδοσης των προτεινόμενων στατιστικών μοντέλων καθώς και των μεθόδων εκπαίδευσης τους, σε τεχνητά και πραγματικά σύνολα παρατηρήσεων.

Στο τελευταίο κεφάλαιο παρουσιάζονται τα συμπεράσματα και γίνονται προτάσεις για μελλοντική έρευνα.

Η πειραματική αξιολόγηση ανέδειξε την βελτιωμένη επίδοση του μικτού μοντέλου ασύμμετρων Π-σιγμοειδών κατανομών στην μοντελοποίηση συνόλων παρατηρήσεων και στην κατάτμηση εικόνων. Επιπλέον, αξιολογήθηκε η δυνατότητα καθορισμού του πραγματικού αριθμού των συνιστωσών κατανομών του μικτού μοντέλου. Συμπερασματικά, στην παρούσα διατριβή προτείνεται ένα αποδοτικό παραμετρικό μοντέλο για την μοντελοποίηση ενός συνόλου παρατηρήσεων.

ΚΕΦΑΛΑΙΟ 2. Η ΚΑΤΑΝΟΜΗ Π-SIGMOID

- 2.1. Γενικά
 - 2.2. Η σιγμοειδής συνάρτηση
 - 2.3. Ορισμός της συνάρτησης πυκνότητας πιθανότητας Π-sigmoid
 - 2.4. Ιδιότητες της Π-sigmoid
 - 2.5. Η πολυδιάστατη Π-sigmoid κατανομή
 - 2.6. Μικτά μοντέλα Π-sigmoid κατανομών (ΠsMM)
 - 2.7. Εκτίμηση του αριθμού των συνιστωσών κατανομών σε ένα ΠsMM
-

2.1. Γενικά

Στην εισαγωγή αναφέραμε ότι στόχος της στατιστικής μοντελοποίησης είναι η περιγραφή του τρόπου γένεσης των παρατηρήσεων. Επίσης, συνδέσαμε την μοντελοποίηση, με την εκτίμηση μιας άγνωστης συνάρτησης πυκνότητας πιθανότητας. Στην διαδικασία εκπαίδευσης των παραμετρικών στατιστικών μοντέλων, παρουσιάσαμε την θεμελιώδη μέθοδο της μέγιστης πιθανοφάνειας, αλλά και τον EM, έναν αποδοτικό αλγόριθμο εκπαίδευσης, με εφαρμογή στα μοντέλα μικτών κατανομών. Τέλος, περιγράψαμε το στατιστικό κριτήριο dip με το οποίο ελέγχεται η ιδιότητα της μονοκόρυφης (unimodal) συνάρτησης.

Στο κεφάλαιο αυτό θα παρουσιάσουμε ένα παραμετρικό μοντέλο που βασίζεται στην σιγμοειδή συνάρτηση και πιο συγκεκριμένα την συνάρτηση πυκνότητας πιθανότητας Π-sigmoid. Επίσης, θα ορίσουμε το μοντέλο μικτών Π-sigmoid κατανομών, καθώς

και της μεθόδου προσδιορισμού του πλήθους των συνιστωσών κατανομών του μοντέλου βασισμένη στο κριτήριο dip .

Κλείνοντας, πρέπει να αναφερθεί πως η κατανομή Π-sigmoid προτάθηκε στα πλαίσια της σχετικής εργασίας [5],[8] και η έρευνα στη παρούσα εργασία οδήγησε στην δημιουργία της AsΠsigmoid, μιας νέας κατανομής, που θα παρουσιαστεί στο επόμενο κεφάλαιο.

Η δημιουργία και η ανάλυση μιας νέας συνάρτησης πυκνότητας πιθανότητας, πέρα από το θεωρητικό ενδιαφέρον που παρουσιάζει, σχετίζεται άμεσα και με τις ερευνητικές ανάγκες που εξυπηρετεί.

Όπως αναφέρθηκε στην εισαγωγή, τα μοντέλα μικτών κατανομών χρησιμοποιούνται αποδοτικά στο πρόβλημα της ομαδοποίησης. Το μοντέλο των μικτών κανονικών κατανομών (GMM) είναι το πλέον διαδεδομένο, καθώς, τόσο οι καλές αναλυτικές του ιδιότητες όσο και το θεώρημα καθολικής προσέγγισης, το καθιστούν μια ισχυρή μέθοδο στο πρόβλημα της ομαδοποίησης.

Ωστόσο, υπάρχουν κάποια μειονεκτήματα που αφορούν την ίδια την φύση της κανονικής κατανομής. Τα αποτελέσματα ομαδοποίησης του μοντέλου μικτών κατανομών, δεν είναι σε μορφή ερμηνεύσιμη από τον άνθρωπο. Συνήθως, αντιλαμβανόμαστε καλύτερα την περιγραφή των παρατηρήσεων υπό μορφή κανόνων, δηλαδή έναν διαχωρισμό του χώρου των παρατηρήσεων σε ορθογώνιες περιοχές. Επίσης, παρά την δεδομένη ευελιξία των μικτών κανονικών κατανομών, υπάρχει αδυναμία στην μοντελοποίηση παρατηρήσεων που δημιουργούνται από ομοιόμορφες κατανομές.

Στο συγκεκριμένο πρόβλημα, η συμβολή της κατανομής Π-sigmoid, είναι ιδιαίτερα σημαντική, καθώς μέσω της αυξημένης ευελιξίας που επιδεικνύει, έχει την δυνατότητα να περιγράφει δεδομένα από διαφορετικές κατανομές (κανονικά, ομοιόμορφα κ.α.) . Επιπλέον, στην διαδικασία της μοντελοποίησης παρατηρήσεων έχει την δυνατότητα παραγωγής των ζητούμενων ερμηνεύσιμων κανόνων, αφού, όπως θα δείξουμε στην συνέχεια, τα σχετικά όρια που παράγει η κατανομή, είναι υπερεπίπεδα.

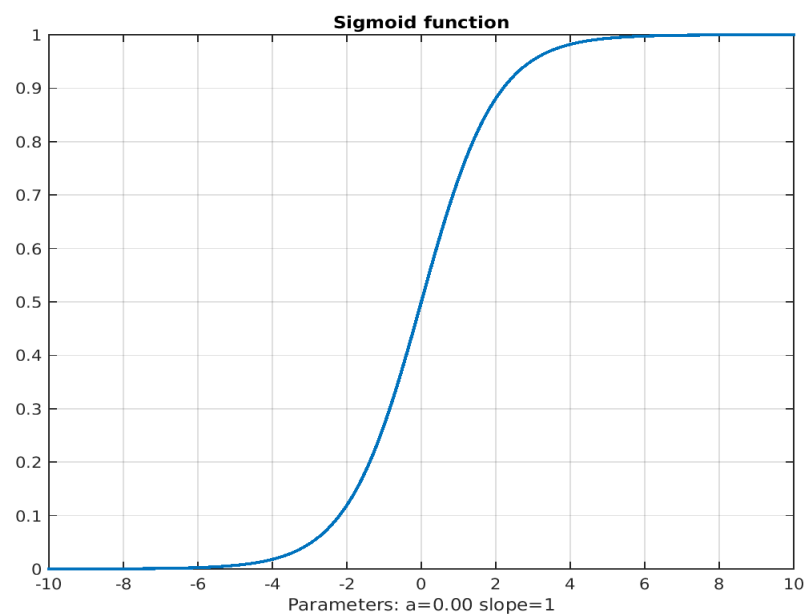
2.2. Η σιγμοειδής συνάρτηση

Η σιγμοειδής συνάρτηση, (γνωστή και ως λογιστική) είναι μια σημαντική συνάρτηση στον χώρο της μηχανικής μάθησης, με πλήθος εφαρμογών, ενώ αποτελεί και την βασική συνιστώσα της υπό μελέτης Π-sigmoid κατανομής. Έχει κυρίαρχο ρόλο στα νευρωνικά δίκτυα ως συνάρτηση ενεργοποίησης, αλλά ταυτόχρονα είναι ιδιαίτερα διαδεδομένη και σε άλλες επιστήμες (οικολογία, ιατρική, χημεία, φυσική κ.α.).

Η σιγμοειδής συνάρτηση ορίζεται από τον τύπο:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

Το πεδίο ορισμού της f είναι το σύνολο των πραγματικών αριθμών ενώ το σύνολο τιμών της είναι το $[0,1]$. Επίσης, είναι μια αύξουσα, φραγμένη και παραγωγίσιμη συνάρτηση. Από την μορφή του συνόλου τιμών, συμπεραίνουμε ότι η συνάρτηση έχει την ιδιότητα να κανονικοποιεί τις παρατηρήσεις του συνόλου X . Το κέντρο της σιγμοειδούς συνάρτησης είναι η αρχή των αξόνων και η γραφική της παράσταση προσεγγίζει το χαρακτηριστικό σχήμα “S” όπως φαίνεται στο Σχήμα 2.1.



Σχήμα 2.1 Η σιγμοειδής συνάρτηση με κέντρο το μηδέν και κλίση μονάδα.

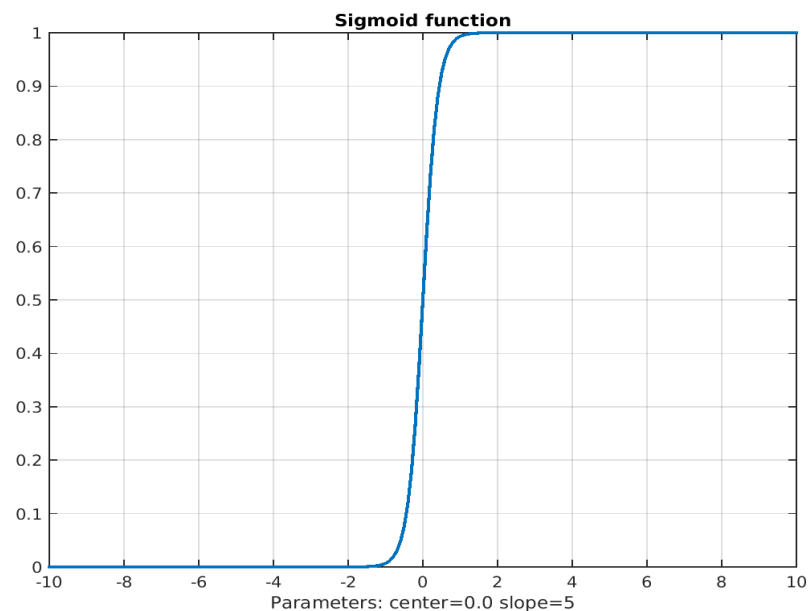
Ιδιαίτερα σημαντική είναι η μετατροπή του ορισμού της σιγμοειδούς συνάρτησης σε μια πιο γενική μορφή, δίνοντας την δυνατότητα μετατόπισης του κέντρου, αλλά και της αλλαγής της κλίσης της συνάρτησης. Για τον σκοπό αυτό, ορίζουμε την συνάρτηση:

$$f(x) = \frac{1}{1 + e^{-\lambda(x-a)}} \quad (2.2)$$

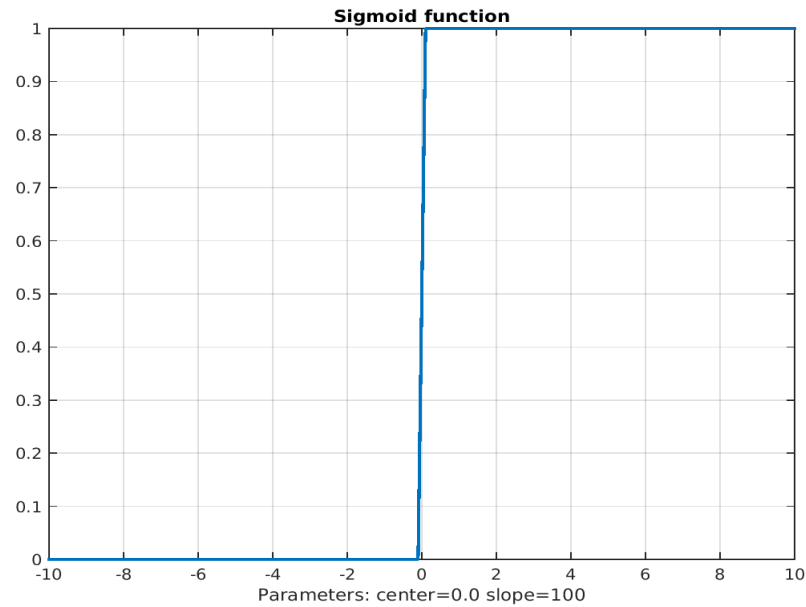
Όπου με a συμβολίζεται το κέντρο και με λ η κλίση της συνάρτησης στο κέντρο a . Αν αντικαταστήσουμε $a = 0$ και $\lambda = 1$, παίρνουμε την σχέση (2.1).

Στην νέα αυτή προσέγγιση, παρατηρούμε ότι μεγάλες τιμές της κλίσης λ , προκαλούν απότομη μετάβαση των τιμών της συνάρτησης από το μηδέν στη μονάδα, ενώ η συμπεριφορά της συνάρτησης σε τιμές του x μακριά από την περιοχή του κέντρου (δηλ. $x \ll a$ και $x \gg a$) τείνει να είναι σταθερή. Τέλος, για $\lambda = 0$, η συνάρτηση γίνεται σταθερή με $f(x) = \frac{1}{2}$.

Στη συνέχεια, παρουσιάζονται γραφικές παραστάσεις της συνάρτησης για διάφορες τιμές των παραμέτρων της.



Σχήμα 2.2 Η γραφική παράσταση της σιγμοειδούς συνάρτησης με κέντρο $a=0$ και κλίση $\lambda=5$.



Σχήμα 2.3 Η γραφική παράσταση της σιγμοειδούς συνάρτησης με κέντρο $a=0$ και κλίση $\lambda=100$. Φαίνεται καθαρά η απότομη μεταβολή της συνάρτησης από το μηδέν στην μονάδα.

Έπειτα από την παρουσίαση του ορισμού της σιγμοειδούς συνάρτησης, θα αναλυθούν κάποιες βασικές της ιδιότητες:

Ιδιότητα συμμετρίας:

$$\sigma(-x) = \frac{1}{1+e^x} = \frac{e^{-x}}{1+e^{-x}} = \frac{1-1+e^{-x}}{1+e^{-x}} = \frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}} = 1 - \sigma(x) \quad (2.3)$$

Αντίστροφη συνάρτηση:

$$\sigma^{-1}(x) = \ln\left(\frac{x}{1-x}\right) \quad (2.4)$$

Ιδιότητα παραγώγου:

$$\begin{aligned} \frac{d\sigma}{dx} &= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{(1+e^{-x})} \frac{e^{-x}}{(1+e^{-x})} = \sigma(x) \frac{e^{-x}}{(1+e^{-x})} \\ &= \sigma(x)\sigma(-x) = \sigma(x)[1-\sigma(x)] \end{aligned} \quad (2.5)$$

Η παραπάνω μορφή της παραγώγου κρίνεται ιδιαίτερα σημαντική και εύχρηστη, καθώς ο υπολογισμός της εξαρτάται αποκλειστικά από τις τιμές της σιγμοειδούς συνάρτησης. Η παράγωγος μηδενίζεται όταν $\sigma(x) = 0$ ή $\sigma(x) = 1$.

Ακόμη, εύκολα προκύπτει η παράγωγος της γενικής μορφής της σιγμοειδούς συνάρτησης. Έστω $z = \lambda(x - a)$. Από τον κανόνα της αλυσίδας ισχύει:

$$\frac{d\sigma}{dx} = \frac{d\sigma}{dz} \frac{dz}{dx} = \sigma(x)[1 - \sigma(x)]\lambda \quad (2.6)$$

2.3. Ορισμός της συνάρτησης πυκνότητας πιθανότητας Π-sigmoid

Ύστερα από την παρουσίαση της σιγμοειδούς συνάρτησης, θα παρουσιαστεί η κατανομή Π-sigmoid [5],[8].

Ορισμός: Η συνάρτηση πυκνότητας πιθανότητας Π-sigmoid ορίζεται ως η διαφορά δύο σιγμοειδών συναρτήσεων με κέντρα a, b όπου $b > a$. Ο τύπος της δίνεται από την σχέση:

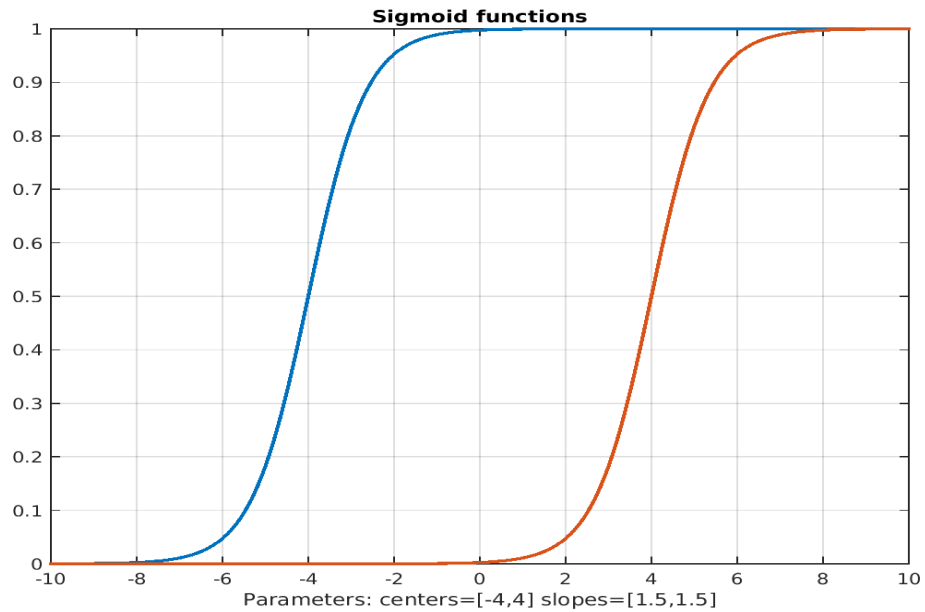
$$f(x) = \left(\frac{1}{b-a} \right) \left[\frac{1}{1 + e^{-\lambda(x-a)}} - \frac{1}{1 + e^{-\lambda(x-b)}} \right] \quad (2.7)$$

Όπου το λ είναι η κοινή κλίση των δύο σιγμοειδών συναρτήσεων στα κέντρα τους a, b αντίστοιχα, η οποία πρέπει να είναι μεγαλύτερη του μηδενός. Για να δηλώσουμε ότι μια τυχαία μεταβλητή X ακολουθεί την Π-sigmoid κατανομή με κλίση λ και κέντρα a, b γράφουμε: $X \sim Ps(a, b, \lambda)$.

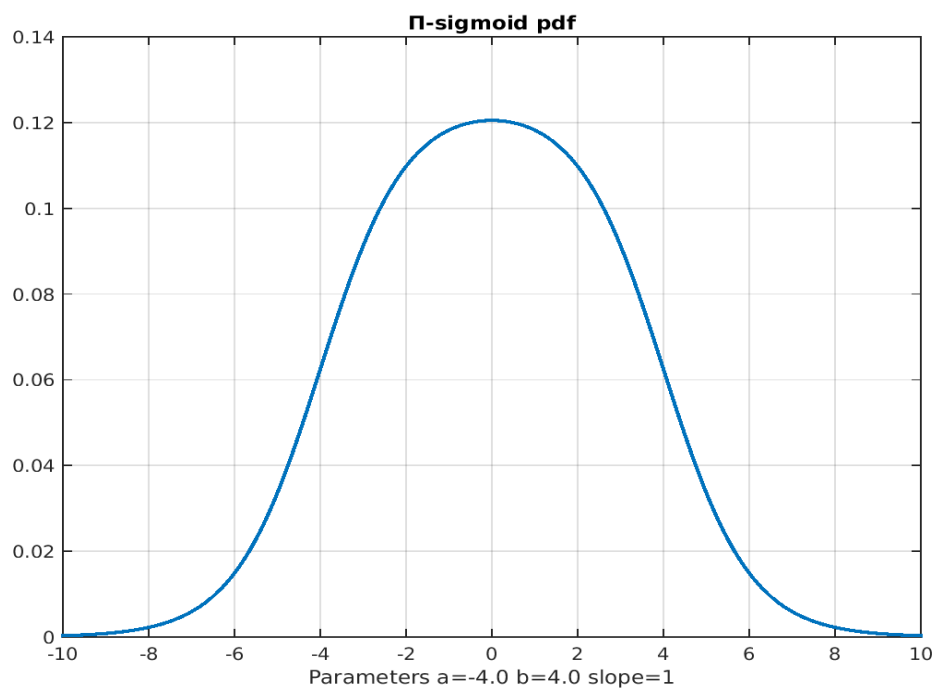
Ο παράγοντας $K_\sigma = \frac{1}{b-a}$ αποτελεί τη σταθερά κανονικοποίησης, εξασφαλίζοντας τη θεμελιώδη σχέση-ορισμό μιας συνάρτησης πυκνότητας πιθανότητας:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

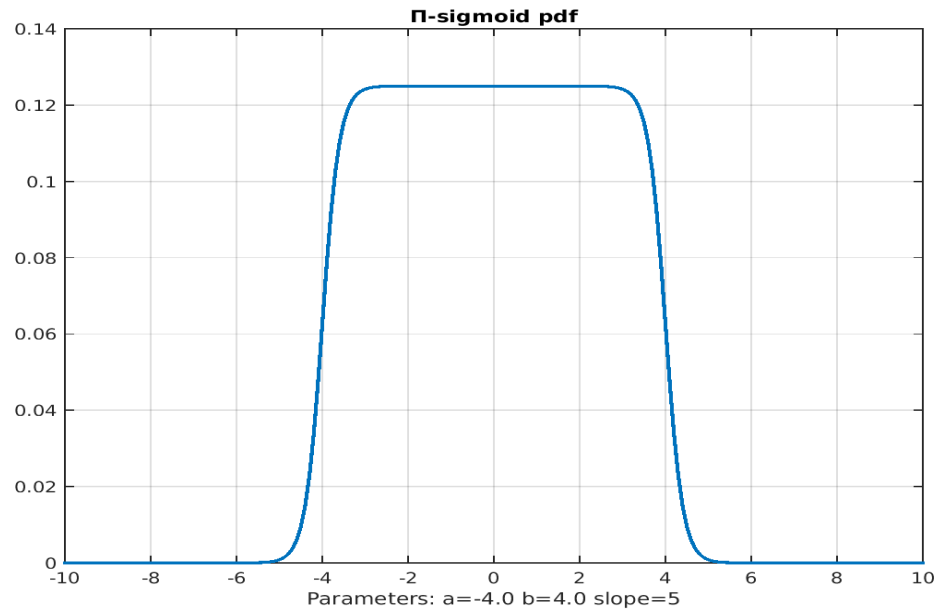
Η σταθερά κανονικοποίησης K_σ είναι ανεξάρτητη της κοινής κλίσης των δύο σιγμοειδών στα αντίστοιχα κέντρα τους. Η παραπάνω διατύπωση μειώνει την πολυπλοκότητα της κατανομής και μεταφέρει καλές αλγεβρικές ιδιότητες στις παραγώγους της. Ακολουθούν γραφικές παραστάσεις της Π-sigmoid για διάφορες τιμές των παραμέτρων της.



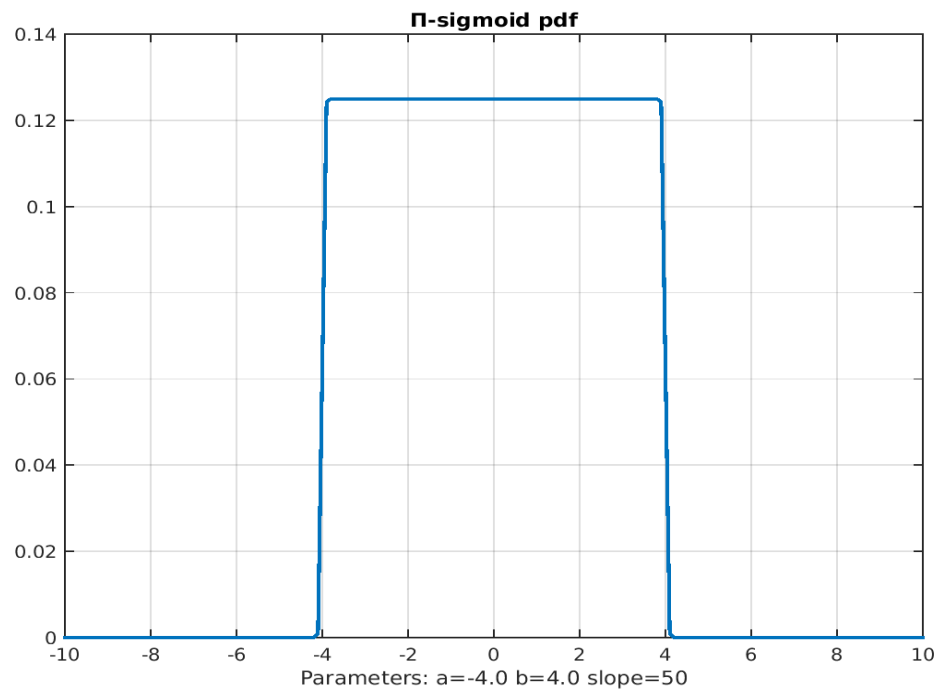
Σχήμα 2.4 Παράθεση δύο διαφορετικών σιγμοειδών συναρτήσεων με κέντρα $a=-4$ και $b=4$ και κλίση $\lambda=1.5$. Το αποτέλεσμα της διαφοράς τους θα είναι μια καμπανοειδής συνάρτηση ενώ για $\lambda \gg 1$ θα τείνει σε σχήμα Π .



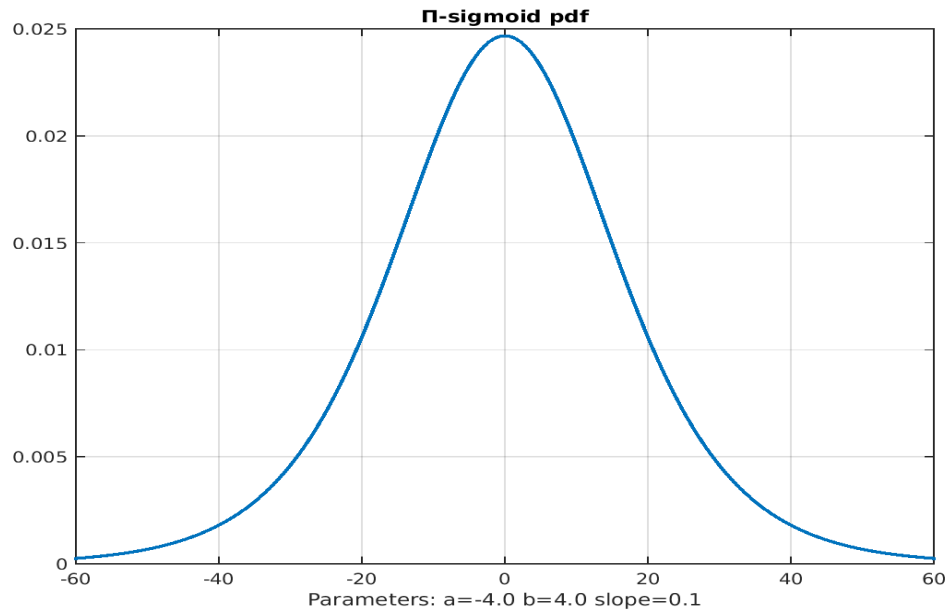
Σχήμα 2.5 Η κατανομή Π -sigmoid με παραμέτρους $a=-4$, $b=4$ και κλίση $\lambda=1$. Να σημειωθεί ότι το σχήμα αυτής της κατανομής προκύπτει από την διαφορά των σιγμοειδών συναρτήσεων αντίστοιχων με το Σχήμα 2.4.



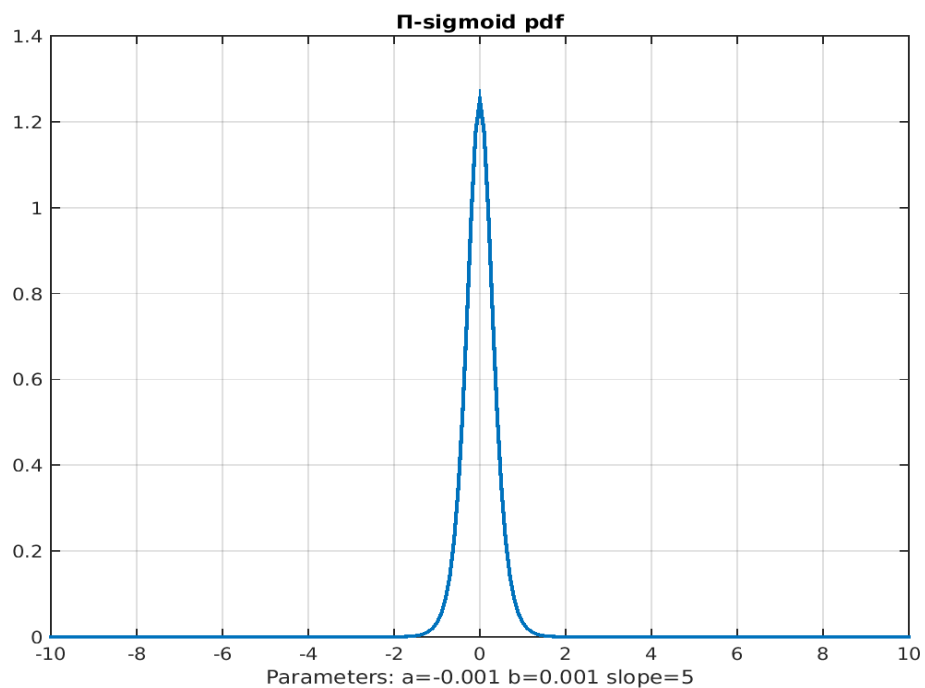
Σχήμα 2.6 Η κατανομή Π-sigmoid με παραμέτρους $a=-4$, $b=4$ και κλίση $\lambda=5$. Μεγαλώνοντας την τιμή του λ , το γεωμετρικό σχήμα της κατανομής αρχίζει να προσεγγίζει το σχήμα Π .



Σχήμα 2.7 Η κατανομή Π-sigmoid με παραμέτρους $a=-4$, $b=4$ και κλίση $\lambda=50$. Μεγαλώνοντας περισσότερο την τιμή του λ , η κατανομή προσεγγίζει με πολύ ικανοποιητικό τρόπο την ομοιόμορφη.



Σχήμα 2.8 Η κατανομή Π-sigmoid με παραμέτρους $a=-4$, $b=4$ και κλίση $\lambda=0.1$. Η συνάρτηση πλατειάζει σημαντικά και για αυτό ανοίγουμε το διάστημα των τιμών από το $[-10\ 10]$ στο $[-60\ 60]$ για να είναι ορατό το σχήμα της.



Σχήμα 2.9 Η κατανομή Π-sigmoid με παραμέτρους $a=-0.001$, $b=0.001$ και κλίση $\lambda=5$. Η συνάρτηση γίνεται απότομη, όταν η σχετική απόσταση του a με το b γίνει μικρή.

2.4. Ιδιότητες της Π-sigmoid

Έπειτα από την μαθηματική διατύπωση καθώς και τη γραφική περιγραφή της συνάρτησης πυκνότητας πιθανότητας, θα παρουσιάσουμε τις παραμετρικές της ιδιότητες και βασικές στατιστικές ιδιότητες.

Η παράμετρος λ

Η κοινή κλίση λ των δυο σιγμοειδών συνιστωσών συναρτήσεων εξασφαλίζει την ιδιότητα της συμμετρίας και επηρεάζει σε μεγάλο βαθμό την συμπεριφορά της κατανομής. Για μεγάλες τιμές της κλίσης, η συνάρτηση προσεγγίζει την ομοιόμορφη κατανομή, ιδιότητα μοναδική, καθώς επιτυγχάνεται διατηρώντας την ιδιότητα της συνέχειας. Για μικρότερες τιμές της κλίσης (δηλ. $\lambda = 1 + \varepsilon$, $\varepsilon \geq 0$), η συνάρτηση παίρνει μια καμπανοειδή μορφή που προσεγγίζει σχηματικά την κανονική κατανομή, ενώ για τιμές της κλίσης $\lambda < 1$, η συνάρτηση αποκτά μεγάλο άνοιγμα (πλατειάζει), αποκτώντας την δυνατότητα περιγραφής δεδομένων με μεγάλη διασπορά.

Οι παράμετροι a και b

Οι δύο αυτές παράμετροι της κατανομής, αποτελούν τα κέντρα των δυο σιγμοειδών συνιστωσών συναρτήσεων και έχουν αντίστοιχες ιδιότητες με την παράμετρο της διασποράς στην κανονική κατανομή. Αύξηση στην απόσταση $|b - a|$, υποδηλώνει μεγάλη διασπορά των παρατηρήσεων. Χαρακτηριστικό παράδειγμα, είναι η συμπεριφορά της κατανομής για πολύ μεγάλες τιμές κλίσης ($\lambda \gg 1$). Σε αυτή την περίπτωση, όπως φαίνεται και στο Σχήμα 2.8, η κατανομή ανέρχεται και κατέρχεται στα σημεία a και b αντίστοιχα, σχεδόν κατακόρυφα, συνδέοντας άμεσα την διασπορά της Π-sigmoid κατανομής, με αυτήν της ομοιόμορφης κατανομής, που υπολογίζεται από την σχέση: $\frac{(b-a)^2}{12}$.

Στην έννοια της διασποράς της Π-sigmoid κατανομής, συμμετέχει και η παράμετρος της κλίσης, έχοντας αντιστρόφως ανάλογη συμπεριφορά.

Ακολουθεί η περιγραφή της αθροιστικής συνάρτησης κατανομής, καθώς επίσης και κάποιων βασικών στατιστικών συναρτήσεων.

Αθροιστική συνάρτηση (CDF)

$$\begin{aligned}
 F(x; a, b, \lambda) &= \frac{1}{b-a} \int_{-\infty}^x \left[\frac{1}{1+e^{-\lambda(u-a)}} - \frac{1}{1+e^{-\lambda(u-b)}} \right] du \\
 &= \frac{\ln(e^{-\lambda(x-a)} + 1) - \ln(e^{-\lambda(x-b)} + 1) + \lambda(b-a)}{\lambda(b-a)} \\
 &= \frac{\ln(e^{-\lambda(x-a)} + 1) - \ln(e^{-\lambda(x-b)} + 1)}{\lambda(b-a)} + 1
 \end{aligned} \tag{2.8}$$

Η αναμενόμενη τιμή ή κέντρο

$$E(X) = \frac{1}{b-a} \int_{-\infty}^{+\infty} \left(\frac{1}{1+e^{-\lambda(x-a)}} - \frac{1}{1+e^{-\lambda(x-b)}} \right) x dx = \frac{a+b}{2} \tag{2.9}$$

Διάμεσος (median)

Η Π-sigmoid είναι συμμετρική συνάρτηση, οπότε η διάμεσος ταυτίζεται με την μέση τιμή (και την κορυφή), άρα ισχύει:

$$Median_X = \frac{a+b}{2} \tag{2.10}$$

2.5. Η πολυδιάστατη Π-sigmoid κατανομή

Στην παράγραφο αυτή θα παρουσιάσουμε την κατανομή Π-sigmoid για πολυδιάστατα δεδομένα, υποθέτοντας ότι η κατανομή είναι ανεξάρτητη στις διαστάσεις των παρατηρήσεων. Έστω ένα σύνολο από N παρατηρήσεις, διάστασης D και $p(x; \theta)$ μια συνάρτηση πυκνότητας πιθανότητας. Κάθε παρατήρηση έχει την μορφή:

$$x^i = (x_1^i, x_2^i, \dots, x_D^i), \quad i = 1 \dots N.$$

Σύμφωνα με την υπόθεση της ανεξαρτησίας της $p(x; \theta)$ σε κάθε διάσταση των παρατηρήσεων, η συνάρτηση πυκνότητας πιθανότητας έχει την μορφή:

$$p(x; \theta) = \prod_{j=1}^D p(x_j^i; \theta_j), \quad i = 1 \dots N \tag{2.11}$$

όπου θ_j είναι το διάνυσμα των παραμέτρων της κατανομής σε κάθε διάσταση.

Αντικαθιστώντας στην τελευταία σχέση τον ορισμό της μονοδιάστατης Π-sigmoid (2.7) έχουμε [5],[8]:

$$p(x; \theta) = \prod_{j=1}^D \frac{1}{b_j - a_j} \left[\frac{1}{1 + e^{-\lambda_j(x_j^i - a_j)}} - \frac{1}{1 + e^{-\lambda_j(x_j^i - b_j)}} \right], \quad (2.12)$$

$i = 1 \dots N$

Η τελευταία σχέση αποτελεί την πολυδιάστατη εκδοχή της κατανομής Π-sigmoid, η οποία προκύπτει ως “τομή” των επιμέρους μονοδιάστατων εκδοχών της, όπως αυτές διατυπώνονται και στο γινόμενο της σχέσης (2.11).

Τέλος, στον τύπο της πολυδιάστατης Π-sigmoid (2.12) παρατηρούμε ότι κάθε διάσταση περιλαμβάνει τρεις ανεξάρτητες παραμέτρους:

$$a_j, b_j, \lambda_j, \quad j = 1 \dots D$$

Γενικότερα, για την πλήρη περιγραφή μιας πολυδιάστατης (D -διάστατης) Π-sigmoid κατανομής, πρέπει να προσδιοριστούν συνολικά $3 * D$ παράμετροι.

2.6. Μικτά μοντέλα Π-sigmoid κατανομών (PsMM)

Στις περιπτώσεις που τα σύνολα παρατηρήσεων είναι ανομοιογενή, περιλαμβάνοντας παρατηρήσεις από διαφορετικές κατανομές, μια κατανομή αδυνατεί να τα περιγράψει επαρκώς. Σε αυτήν την περίπτωση υποθέτουμε ένα μοντέλο μικτών κατανομών, το οποίο παρουσιάζει πολύ ισχυρές δυνατότητες μοντελοποίησης παρατηρήσεων.

Σύμφωνα με την σχέση (1.13) ένα PsMM (Π-sigmoid Mixture Model) αποτελεί έναν γραμμικό συνδυασμό πεπερασμένων Π-sigmoid κατανομών, σταθμισμένων με βάρη. Το μοντέλο για K κατανομές έχει την μορφή:

$$p(x) = \sum_{k=1}^K \pi_k Ps(x; a_k, b_k, \lambda_k) \quad (2.13)$$

Κάθε Π-sigmoid κατανομή $Ps(x; a_k, b_k, \lambda_k)$, $k = 1 \dots K$ είναι συνιστώσα του μικτού μοντέλου και προσδιορίζεται από τις παραμέτρους a_k, b_k, λ_k .

Τα π_k ονομάζονται παράμετροι μίξης και πρέπει να ικανοποιούν τους παρακάτω περιορισμούς:

$$0 \leq \pi_k \leq 1, \quad \forall k = 1 \dots K$$

$$\sum_{k=1}^K \pi_k = 1$$

Συνοψίζοντας, όταν το πλήθος των συνιστωσών K είναι γνωστό, τότε ένα μοντέλο μικτών Π-sigmoid κατανομών περιγράφεται πλήρως από τις παρακάτω παραμέτρους:

$$\pi = \{\pi_1, \dots, \pi_k\}, \forall k = 1 \dots K \quad (2.14)$$

$$\alpha = \{a_1, \dots, a_k\}, \forall k = 1 \dots K \quad (2.15)$$

$$b = \{b_1, \dots, b_k\}, \forall k = 1 \dots K \quad (2.16)$$

$$\lambda = \{\lambda_1, \dots, \lambda_k\}, \forall k = 1 \dots K \quad (2.17)$$

Για τον προσδιορισμό τους, χρησιμοποιείται αποδοτικά ο αλγόριθμος GEM και στα πλαίσια αυτής της παραγράφου παραλείπεται.. Τέλος, πρέπει να τονιστεί ότι το πλήθος των συνιστωσών κατανομών, είναι μια άγνωστη υπερπαραμέτρος ενός μικτού μοντέλου, η οποία πρέπει να προσδιοριστεί για την εκκίνηση της διαδικασίας εκπαίδευσης.

2.7. Εκτίμηση του αριθμού των συνιστωσών κατανομών σε ένα PsMM

Στα προβλήματα μηχανικής μάθησης, ο προσδιορισμός κάποιων πρωταρχικών παραμέτρων, επηρεάζει σε μεγάλο βαθμό την γενικευτική ικανότητα και την πολυπλοκότητα των μεθόδων.

Στην στατιστική μοντελοποίηση, ο προσδιορισμός εκτιμήσεων για τις παραμέτρους του μοντέλου που υποθέτουμε, γίνεται με την διαδικασία της εκπαίδευσης και περιλαμβάνει την μεγιστοποίηση μιας στατιστικής συνάρτησης (συνάρτηση πιθανοφάνειας) μέσω κάποιας τεχνικής βελτιστοποίησης (αλγόριθμος EM). Στο τέλος

της διαδικασίας της εκπαίδευσης, εξάγουμε ένα στατιστικό μοντέλο με τις βέλτιστες εκτιμήσεις των παραμέτρων.

Στην πραγματικότητα, τα σύνολα των παρατηρήσεων είναι σύνθετα. Για την μοντελοποίηση τους, τα στατιστικά μοντέλα που υποθέτουμε είναι ισχυρά και η διατύπωση τους περιλαμβάνει επιπλέον υπερπαραμέτρους που πρέπει να προσδιοριστούν. Η διαδικασία της εκπαίδευσης των μοντέλων αυτών είναι ιδιαίτερα απαιτητική, καθώς μια τυπική διαδικασία εκπαίδευσης, αδυνατεί να εκτιμήσει τις τιμές των υπερπαραμέτρων με επιτυχία. Για τον σκοπό αυτό, γίνεται κατάλληλη εφαρμογή ενός στατιστικού κριτηρίου (ή και περισσοτέρων) ικανού να ενισχύσει την διαδικασία.

Ο συνδυασμός της πάγιας διαδικασίας εκπαίδευσης με την εφαρμογή του κριτηρίου, απαιτεί μελέτη του σύνθετου προβλήματος και συνήθως, επιμερισμό του σε απλούστερα προβλήματα.

Το πλήθος των συνιστωσών κατανομών για τα μοντέλα μικτών κατανομών, αποτελεί ένα πρωτεύον πρόβλημα. Η υπερπαραμέτρος αυτή, είναι θεμελιώδης και δεν μπορεί να εκτιμηθεί μέσω της τυπικής διαδικασίας της εκπαίδευσης (EM). Παρόλα αυτά, η τιμή της, επιδρά άμεσα στη απόδοση και στην πολυπλοκότητα, καθώς συνδέεται άρρηκτα με την επιτυχή μοντελοποίηση του συνόλου των παρατηρήσεων.

Στην προσπάθεια προσδιορισμού της υπερπαραμέτρου, θα χρησιμοποιήσουμε το στατιστικό κριτήριο dip και θα δημιουργήσουμε έναν νέο αλγόριθμο εκπαίδευσης.

Το θεωρητικό σκέλος του κριτηρίου dip παρουσιάστηκε στο πρώτο κεφάλαιο. Η εφαρμογή του γίνεται στην εμπειρική συνάρτηση κατανομής που αντιστοιχεί σε ένα σύνολο μονοδιάστατων παρατηρήσεων και εξάγεται μια μη αρνητική τιμή. Η τιμή του κριτηρίου dip , ορίζει εάν η εμπειρική συνάρτηση κατανομής έχει μια ή περισσότερες κορυφές. Μέσω της τιμής αυτής, ορίζεται και η p -value για τον στατιστικό έλεγχο που θα πραγματοποιηθεί.

Μια μικρή τιμή του dip υποδηλώνει μονοκόρυφη εμπειρική κατανομή. Αντίθετα, μια μικρή τιμή της p -value ($p - \text{value} < \alpha$), οδηγεί στην απόρριψη της υπόθεσης H_0 : “Η F_n είναι μονοκόρυφη”, με επίπεδο στατιστικής σημαντικότητας α .

Για την νέα διαδικασία εκπαίδευσης ενός PsMM μέσω του κριτηρίου dip , θέσαμε επίπεδο στατιστικής σημαντικότητας $\alpha = 0$.

Ακόμη, υποθέσαμε ότι μια μονοκόρυφη εμπειρική κατανομή περιγράφει ένα σύνολο παρατηρήσεων που δημιουργήθηκε από μια Π-sigmoid κατανομή, ενώ αντίθετα, μια εμπειρική κατανομή με πολλαπλές κορυφές περιγράφει ένα σύνθετο σύνολο παρατηρήσεων που δημιουργείται από περισσότερες Π-sigmoid κατανομές.

Οι ισοδύναμες υποθέσεις για την επίλυση του προβλήματος ομαδοποίησης μέσω του μοντέλου μικτών κατανομών, ορίζουν ότι μια μονοκόρυφη εμπειρική κατανομή περιγράφει μια ομοιογενή ομάδα παρατηρήσεων, ενώ αντίθετα μια κατανομή με πολλαπλές κορυφές περιγράφει ένα σύνολο παρατηρήσεων το οποίο περιλαμβάνει υποομάδες.

Στην συνέχεια θα παρουσιαστεί η διαδικασία εκπαίδευσης ενός μοντέλου μικτών Π-sigmoid κατανομών, με έμφαση στον υπολογισμό του πλήθους των συνιστωσών του κατανομών.

Έστω ένα σύνολο από N παρατηρήσεις, διάστασης D . Κάθε παρατήρηση έχει την μορφή:

$$x^i = \{x_1^i, x_2^i \dots x_D^i\}, \quad i = 1 \dots N$$

Συμβολίζουμε με Π^j το σύνολο μονοδιάστατων παρατηρήσεων, που προκύπτει από την προβολή ενός υπό μελέτη συνόλου, στον j -άξονα των παρατηρήσεων $j = 1 \dots D$.

Σύμφωνα με τα παραπάνω, όταν μια τυχαία παρατήρηση $x^i = \{x_1^i, x_2^i \dots x_D^i\}$, $i = 1 \dots N$, προβάλλεται στον j -άξονα, $j = 1 \dots D$, ισχύει $x_j^i \in \Pi^j$.

Ακόμα, συμβολίζουμε με dip_pvalue_j την $p-value$ που προκύπτει από την εφαρμογή του κριτηρίου dip στα στοιχεία του συνόλου Π^j .

1. Αρχικά υποθέτουμε ότι το σύνολο των παρατηρήσεων X , παράγεται εξ ολοκλήρου από μια κατανομή οπότε αρχικοποιούμε το πλήθος συνιστωσών K στη τιμή ένα ($K=1$).
2. Στην συνέχεια, και ανεξάρτητα για την κάθε διάσταση των παρατηρήσεων του συνόλου X , υπολογίζουμε την τιμή dip_pvalue_j των συνόλων:

$$\Pi^j = X^j = \{x_j^1, x_j^2 \dots x_j^N\}, j = 1 \dots D$$

(ο υπολογισμός των dip_pvalue_j γίνεται συνολικά D φορές).

Για κάθε σύνολο Π^j συγκρίνουμε την αντίστοιχη τιμή της dip_pvalue_j με το μηδέν.

- Εάν $\exists j \in 1 \dots D : dip_pvalue_j = 0$, τότε η προβολή του X στην j διάσταση, περιγράφεται από μια εμπειρική συνάρτηση κατανομής με πολλαπλές κορυφές, οπότε συμπεραίνουμε ότι το υπό μελέτη σύνολο παρατηρήσεων X , παράγεται από περισσότερες (τουλάχιστον δύο) συνιστώσες κατανομές. Για τον σκοπό αυτό, το πλήθος των συνιστωσών κατανομών αυξάνεται κατά ένα και ο αλγόριθμος της εκπαίδευσης αρχικοποιείται στην τιμή $K=K+1$.
- Εάν $dip_pvalue_j > 0 \forall j \in 1 \dots D$, τότε η προβολή του X σε κάθε διάσταση περιγράφεται από μια μονοκόρυφη εμπειρική κατανομή, οπότε συμπεραίνουμε ότι το σύνολο των παρατηρήσεων X παράγεται εξ ολοκλήρου από μία κατανομή. Το τελικό συνολικό πλήθος συνιστωσών κατανομών του μοντέλου είναι ένα ($K=1$), δηλαδή το μικτό μοντέλο εκφυλίζεται στην Π-σιγμοειδή κατανομή.

Εάν $K \neq 1$, η διαδικασία εκπαίδευσης συνεχίζεται με τα παρακάτω βήματα:

3. Εκπαιδεύουμε ένα μικτό μοντέλο με K Π-sigmoid κατανομές μέσω του EM αλγορίθμου και προσδιορίζουμε τις εκτιμήσεις:

$$\pi_k, g_k, a_k, r_k, k = 1 \dots K.$$

4. Κάνουμε ομαδοποίηση των πολυδιάστατων παρατηρήσεων, μέσω των εκ των υστέρων πιθανοτήτων που υπολογίστηκαν στο τελευταίο Ε-βήμα του αλγορίθμου. Υπολογίζουμε K -ομάδες $C_k, k = 1 \dots K$
5. Για κάθε μια από τις ομάδες $C_k, k = 1 \dots K$ πολυδιάστατων παρατηρήσεων που προσδιορίστηκαν στο προηγούμενο βήμα, υπολογίζουμε ανεξάρτητα σε κάθε διάσταση $j \in 1 \dots D$, την τιμή dip_pvalue_j των συνόλων:

$$\Pi^j = C_k^j$$

(ο υπολογισμός των dip_pvalue_j γίνεται συνολικά $D * K$ φορές)

Για κάθε σύνολο Π^j συγκρίνουμε την τιμή της dip_pvalue_j με το μηδέν:

- Εάν $\exists j \in 1 \dots D : dip_pvalue_j = 0$, τότε η προβολή της ομάδας C_k στην j διάσταση, περιγράφεται από μια εμπειρική συνάρτηση κατανομής με πολλαπλές κορυφές, οπότε συμπεραίνουμε ότι η υπό μελέτη ομάδα παρατηρήσεων C_k , παράγεται από περισσότερες (τουλάχιστον δύο) συνιστώσες κατανομές. Για τον σκοπό αυτό το πλήθος των συνολικών συνιστωσών κατανομών του μοντέλου, αυξάνεται κατά ένα και ο αλγόριθμος της εκπαίδευσης επανεκκινείται από το βήμα τρία με νέο πλήθος συνιστωσών: $K = K + 1$.
- Εάν $dip_pvalue_j > 0 \forall j \in 1 \dots D, \forall k = 1 \dots K$, τότε η προβολή κάθε ομάδας C_k σε κάθε διάσταση περιγράφεται από μια μονοκόρυφη εμπειρική κατανομή, οπότε συμπεραίνουμε ότι κάθε ομάδα παρατηρήσεων C_k παράγεται εξ ολοκλήρου από μία κατανομή. Το συνολικό πλήθος συνιστωσών κατανομών του μοντέλου, καθώς και οι τελικές εκτιμήσεις των παραμέτρων είναι οι τρέχουσες.

Τα παραπάνω βήματα, οδηγούν στην ολοκλήρωση της νέας διαδικασίας εκπαίδευσης ενός πολυδιάστατου μικτού μοντέλου Π-sigmoid κατανομών με ταυτόχρονη εκτίμηση του συνόλου των παραμέτρων και του πλήθους των συνιστωσών Π-sigmoid κατανομών.

ΚΕΦΑΛΑΙΟ 3. Η ΑΣΥΜΜΕΤΡΗ Π-SIGMOID ΚΑΤΑΝΟΜΗ

- 3.1. Γενικά
 - 3.2. Ορισμός της συνάρτησης πυκνότητας πιθανότητας asymmetric Π-sigmoid
 - 3.3. Ιδιότητες της asymmetric Π-sigmoid
 - 3.4. Η πολυδιάστατη asymmetric Π-sigmoid κατανομή
 - 3.5. Εκπαίδευση παραμετρικών μοντέλων asymmetric Π-sigmoid κατανομών
 - 3.6. Μικτά μοντέλα asymmetric Π-sigmoid κατανομών
 - 3.7. Εκπαίδευση ενός AsΠsMM μέσω του GEM αλγορίθμου
 - 3.7.1. Αρχικοποίηση του GEM αλγορίθμου
 - 3.7.2. Ε-βήμα (Expectation step)
 - 3.7.3. Μ-βήμα (Maximization step): Περιγραφή της διαδικασίας βελτιστοποίησης
 - 3.8. Αντιμετώπιση θορύβου
 - 3.9. Εκτίμηση του πλήθους των συνιστωσών κατανομών σε ένα μοντέλο μικτών asymmetric Π-sigmoid κατανομών.
-

3.1. Γενικά

Σε αυτό το κεφάλαιο θα παρουσιαστεί αναλυτικά η νέα συνάρτηση πυκνότητας πιθανότητας που ονομάζεται ασύμμετρη Π-σιγμοειδής (asymmetric Π-sigmoid). Αρχικά θα περιγραφούν οι αιτίες της δημιουργίας της και στη συνέχεια θα παρουσιαστεί η συμπεριφορά και οι ιδιότητες της νέας συνάρτησης για διάφορες τιμές των παραμέτρων της. Ακόμη, θα παρουσιαστεί ένα μοντέλο μικτών ασύμμετρων Π-sigmoid κατανομών και θα αναλυθεί ο τρόπος εκπαίδευσης, μέσω του

αλγορίθμου EM. Τέλος, θα παρουσιαστεί μια αποδοτική μέθοδος εκτίμησης του πλήθους των συνιστωσών κατανομών ενός μικτού μοντέλου ασύμμετρων Π-sigmoid κατανομών. Η σχετική ανάλυση θα γίνει στην μονοδιάστατη αλλά και στην πολυδιάστατη περίπτωση.

Στο προηγούμενο κεφάλαιο παρουσιάστηκε αναλυτικά η κατανομή Π-sigmoid, η οποία προκύπτει από δύο σιγμοειδείς συναρτήσεις και παρουσιάζει ευελιξία στην περιγραφή παρατηρήσεων, ενώ παράγει αποδοτικές λύσεις στο πρόβλημα της ομαδοποίησης. Μια Π-sigmoid κατανομή, μέσω της κοινής κλίσης των σιγμοειδών συναρτήσεων από τις οποίες συνίσταται, προσαρμόζει το σχήμα της και επιτυγχάνει μια αποδοτική περιγραφή παρατηρήσεων που παράγονται τόσο από την ομοιόμορφη όσο και την κανονική κατανομή. Και στις δύο αυτές περιπτώσεις η συμμετρική ιδιότητα της κατανομής, ενισχύει την διαδικασία της περιγραφής δεδομένων.

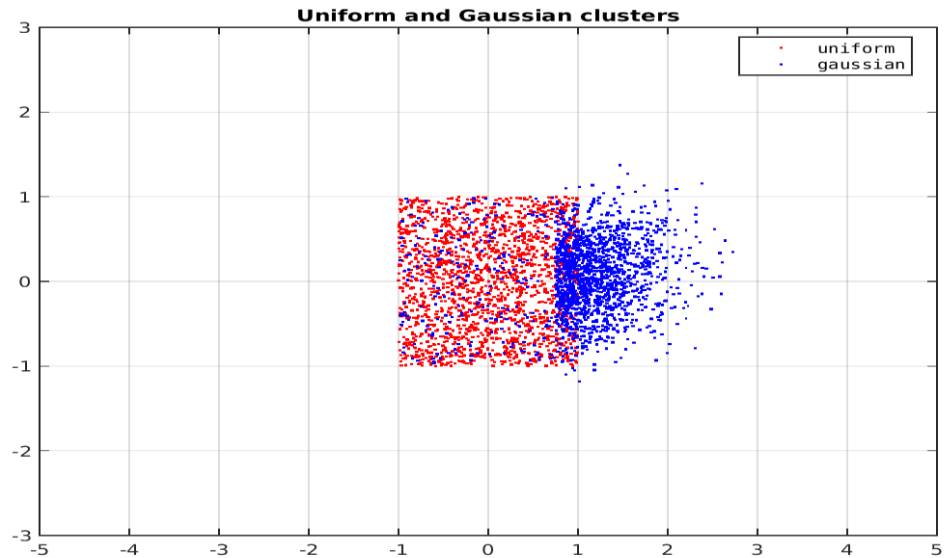
Τι γίνεται όμως στην περίπτωση που οι παρατηρήσεις προκύπτουν από μη συμμετρικές κατανομές;

Μια ιδιαίτερη περίπτωση, αποτελούν οι παρατηρήσεις που παράγονται από τμήματα της κανονικής κατανομής. Η truncated κανονική κατανομή, είναι η συνάρτηση πυκνότητας πιθανότητας μιας τυχαίας μεταβλητής που ακολουθεί την κανονική κατανομή αλλά οι τιμές τις είναι φραγμένες (άνω, κάτω ή και τα δύο).

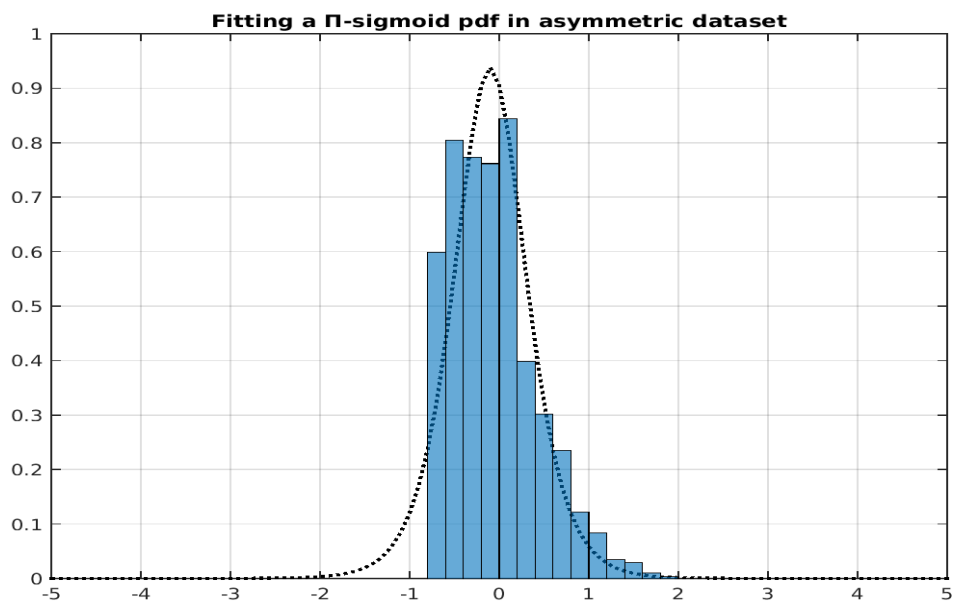
Η κατανομή αυτή, έχει σημαντικές εφαρμογές στη στατιστική αλλά και στην οικονομετρία. Είναι μια χαρακτηριστική ασύμμετρη κατανομή και η μοντελοποίηση των παρατηρήσεων που παράγει, θα μας απασχολήσει στη συνέχεια του κεφαλαίου.

Αντίστοιχα, στο πρόβλημα της ομαδοποίησης, πολλές φορές δημιουργούνται ανομοιογενείς αλλά συνεκτικές ομάδες παρατηρήσεων μέσω επικαλυπτόμενων ομάδων που δεν μπορούν να διακριθούν. Χαρακτηριστικό παράδειγμα αποτελούν οι επικαλυπτόμενες ομάδες κανονικών και ομοιόμορφων παρατηρήσεων τις οποίες μπορούμε να τις θεωρήσουμε ως ενιαίες ομάδες (Σχήμα 3.1). Ανάγοντας λοιπόν το πρόβλημα της ομαδοποίησης σε ένα πρόβλημα εκτίμησης της κατανομής που δημιούργησε τις αντίστοιχες παρατηρήσεις, προκύπτει η ανάγκη για την υπόθεση ενός κατάλληλου παραμετρικού μοντέλου με την μέγιστη δυνατή ακρίβεια στην περιγραφή αντίστοιχων συνόλων παρατηρήσεων. Παρόλα αυτά, η συμμετρική φύση

της κατανομής Π-sigmoid αλλά και της κανονικής κατανομής, δεν επιτρέπει την αποδοτική περιγραφή παρατηρήσεων από ασύμμετρες κατανομές, καθώς η περιγραφή τους από τις συμμετρικές ουρές κρίνεται ανεπιτυχής (Σχήμα 3.2).



Σχήμα 3.1 Επικαλυπτόμενες ομάδες κανονικών και ομοιόμορφων παρατηρήσεων. Το αποτέλεσμα μπορεί να θεωρηθεί ως μια ενιαία ομάδα.



Σχήμα 3.2 Μη αποδοτική περιγραφή ενός μη-συμμετρικού συνόλου παρατηρήσεων από την συμμετρική κατανομή Π-sigmoid.

3.2. Ορισμός της συνάρτησης πυκνότητας πιθανότητας asymmetric Π-sigmoid

Για την αποδοτική μοντελοποίηση παρατηρήσεων που παράγονται από μη συμμετρικές κατανομές, παρατηρούμε ότι απαιτείται μια νέα κατανομή με αυξημένη ευελιξία, με διαφορετική κλίση στις ουρές της, αλλά ταυτόχρονα με την ικανότητα να επιτυγχάνει συμμετρικότητα, ανάλογα με την κατανομή των παρατηρήσεων που καλείται να περιγράψει.

Η διαπίστωση αυτή αλλά και η μελέτη των αδυναμιών των συμμετρικών κατανομών, μας ώθησε στην δημιουργία μιας νέας κατανομής η οποία διατηρεί τις καλές στατιστικές ιδιότητες της Π-sigmoid κατανομής, αλλά ταυτόχρονα είναι μη-συμμετρική.

Ορισμός: Η συνάρτηση πυκνότητας πιθανότητας asymmetric Π-sigmoid ορίζεται ως ένα μικτό μοντέλο δύο Π-sigmoid κατανομών με συντελεστές μίξης $\pi_1 = \pi_2 = \frac{1}{2}$.

Η πρώτη συνιστώσα κατανομή, έχει κέντρα τα σημεία a και b , με $b > a$ και κλίση λ_1 , ενώ η δεύτερη συνιστώσα κατανομή, έχει κέντρα τα σημεία b και c , με $c > b$ και κλίση λ_2 . Ο τύπος της δίνεται από την σχέση:

$$f(x) = \frac{1}{2} [Ps(a, b, \lambda_1) + Ps(b, c, \lambda_2)] \quad (3.1)$$

Εάν γίνει αντικατάσταση των Π-sigmoid συνιστωσών κατανομών προκύπτει η αναλυτική σχέση:

$$f(x) = \frac{1}{2} \left[\frac{1}{b-a} \left[\frac{1}{1 + e^{-\lambda_1(x-a)}} - \frac{1}{1 + e^{-\lambda_1(x-b)}} \right] + \frac{1}{c-b} \left[\frac{1}{1 + e^{-\lambda_2(x-b)}} - \frac{1}{1 + e^{-\lambda_2(x-c)}} \right] \right] \quad (3.2)$$

Οι δύο Π-sigmoid κατανομές έχουν κοινό το κέντρο b και δημιουργούν μια νέα συνεχή συνάρτηση με κέντρα τα σημεία a, b, c με $a < b < c$ και θετικές άνισες κλίσεις λ_1, λ_2 .

Για να δηλώσουμε ότι μια τυχαία μεταβλητή X ακολουθεί την ασύμμετρη Π-sigmoid κατανομή γράφουμε: $X \sim AsPs(a, b, c, \lambda_1, \lambda_2)$

Ως ειδική περίπτωση του παραπάνω ορισμού μπορούμε να θεωρήσουμε την συνάρτηση (3.1) στην οποία το κέντρο b είναι ο μέσος όρος των σημείων a, c δηλαδή ισχύει η σχέση $b = \frac{a+c}{2}$. Η συνάρτηση πυκνότητας πιθανότητας που προκύπτει με τον παραπάνω περιορισμό έχει μειωμένο πλήθος παραμέτρων και παρόμοιες ιδιότητες με την γενική εκδοχή της asymmetric Π-sigmoid που θα παρουσιάσουμε.

Στη συνέχεια θα αποδειχθεί ότι ο παράγοντας $K_\sigma = \frac{1}{2}$ αποτελεί σταθερά κανονικοποίησης στον τύπο της κατανομής εξασφαλίζοντας τη θεμελιώδη σχέση (ορισμό) μιας συνάρτησης πυκνότητας πιθανότητας:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Γνωρίζουμε από την σχέση 2.12 ότι για μια Π-sigmoid κατανομή με κέντρα a, b και κλίση λ_1 ισχύει:

$$\int_{-\infty}^{+\infty} \frac{1}{b-a} \left[\frac{1}{1+e^{-\lambda_1(x-a)}} - \frac{1}{1+e^{-\lambda_1(x-b)}} \right] dx = 1 \quad (3.3)$$

Όμοια, για μια Π-sigmoid κατανομή με κέντρα b, c και κλίση λ_2 , ισχύει:

$$\int_{-\infty}^{+\infty} \frac{1}{c-b} \left[\frac{1}{1+e^{-\lambda_2(x-b)}} - \frac{1}{1+e^{-\lambda_2(x-c)}} \right] dx = 1 \quad (3.4)$$

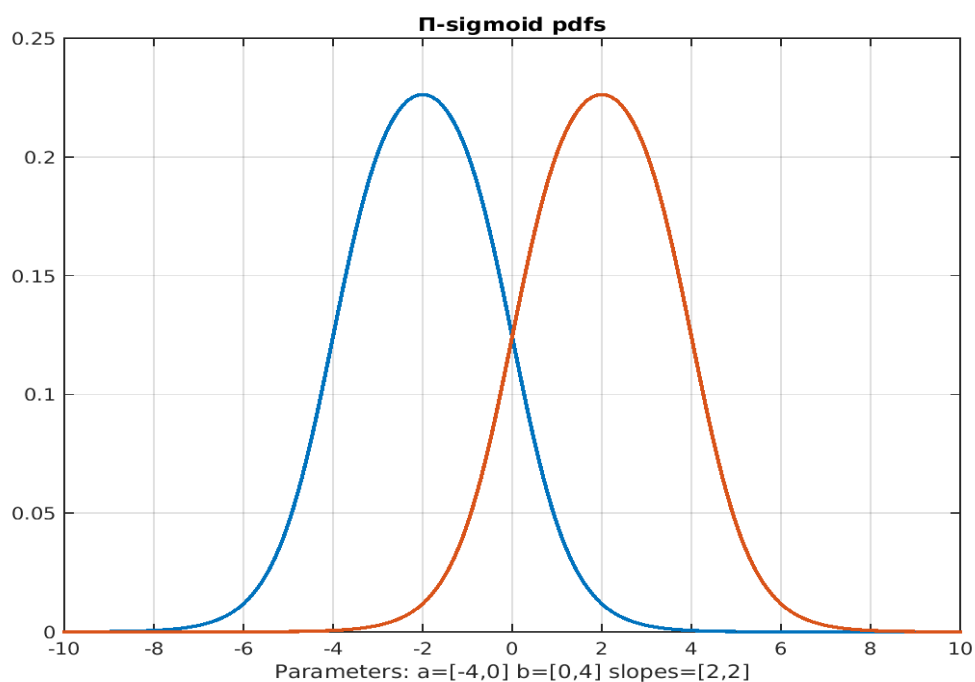
Εύκολα λοιπόν προκύπτει η σχέση:

$$\begin{aligned} K_\sigma^{-1} = \int_{-\infty}^{+\infty} \frac{1}{b-a} \left[\frac{1}{1+e^{-\lambda_1(x-a)}} - \frac{1}{1+e^{-\lambda_1(x-b)}} \right] \\ + \frac{1}{c-b} \left[\frac{1}{1+e^{-\lambda_2(x-b)}} - \frac{1}{1+e^{-\lambda_2(x-c)}} \right] dx = 2 \end{aligned} \quad (3.5)$$

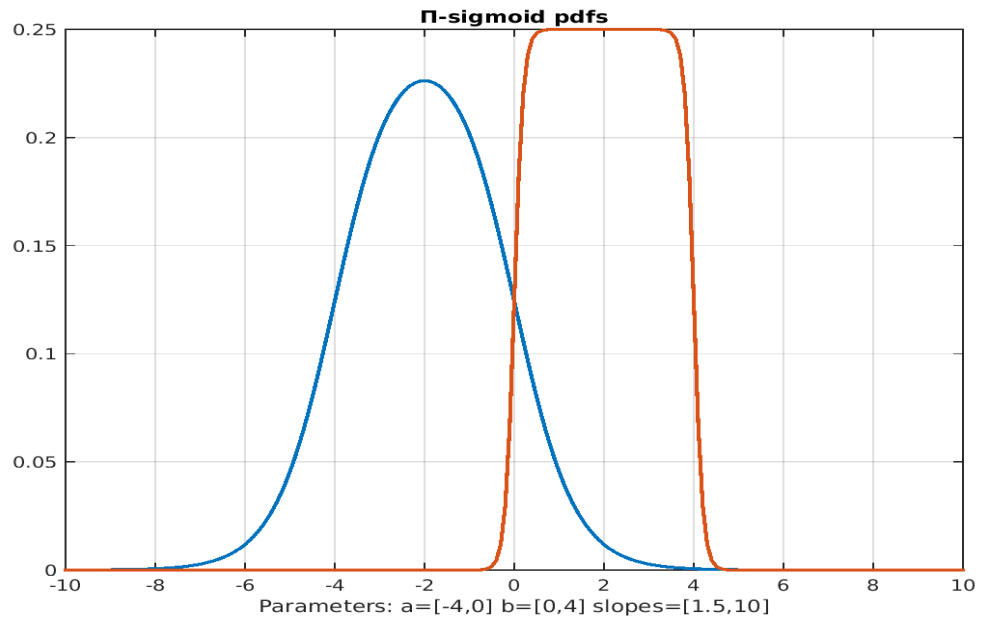
Αποδείχτηκε λοιπόν, πως η τιμή της σταθεράς κανονικοποίησης είναι $K_\sigma = \frac{1}{2}$, καθώς ικανοποιεί την σχέση:

$$\int_{-\infty}^{+\infty} K_\sigma \left[\frac{1}{b-a} \left[\frac{1}{1+e^{-\lambda_1(x-a)}} - \frac{1}{1+e^{-\lambda_1(x-b)}} \right] + \frac{1}{c-b} \left[\frac{1}{1+e^{-\lambda_2(x-b)}} - \frac{1}{1+e^{-\lambda_2(x-c)}} \right] \right] dx = 1 \quad (3.6)$$

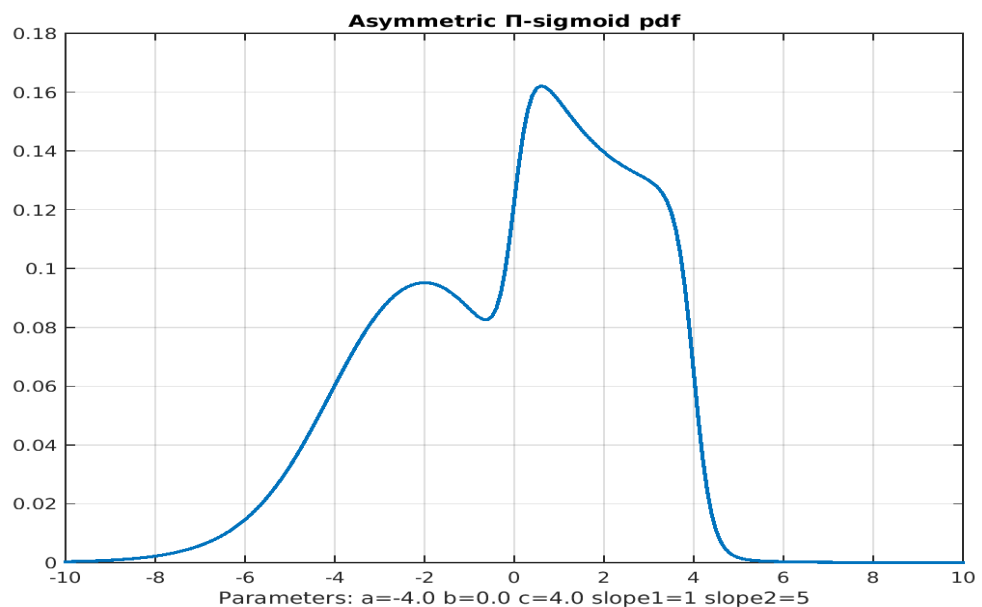
Ακολουθούν γραφικές παραστάσεις της asymmetric Π -sigmoid για διάφορες τιμές των παραμέτρων της.



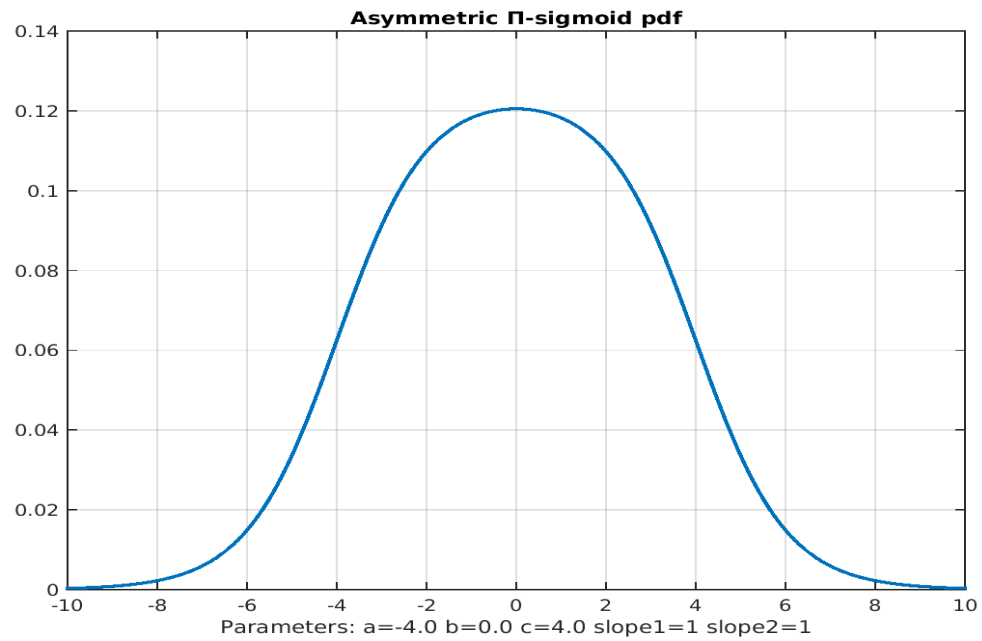
Σχήμα 3.3 Δύο Π -sigmoid κατανομές με κοινό κέντρο $b=0$ και ίσες τιμές κλίσεων. Το αποτέλεσμα που θα προκύψει από το ημίθροισμά τους είναι μια συνεχής και συμμετρική συνάρτηση As Π -sigmoid



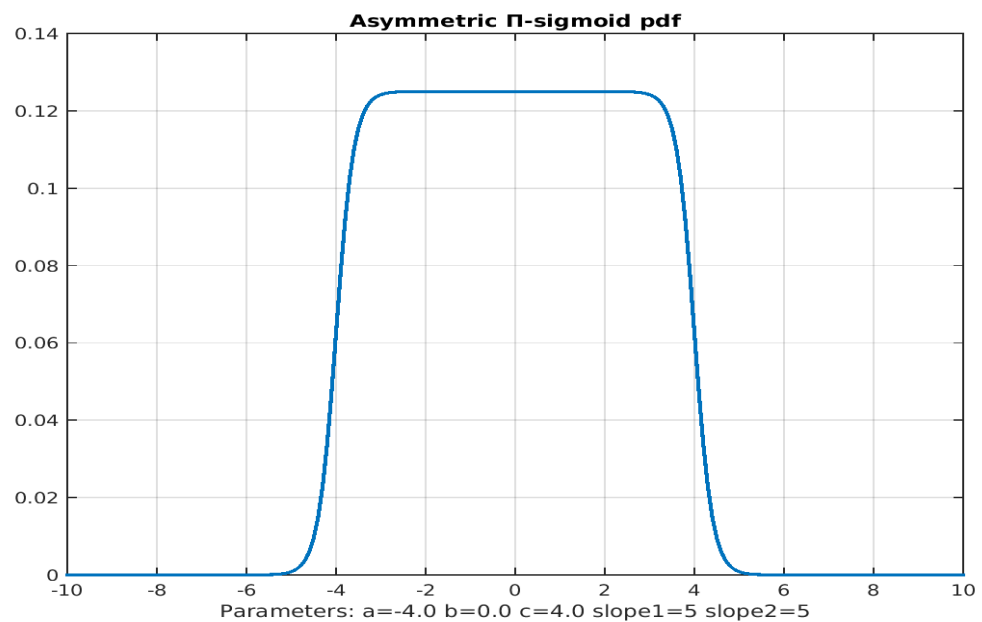
Σχήμα 3.4 Δύο Π-sigmoid κατανομές με κοινό κέντρο $b=0$ και διαφορετικές τιμές κλίσεων. Το αποτέλεσμα που θα προκύψει από το ημιάθροισμά τους είναι μια μη-συμμετρική συνάρτηση AsΠ-sigmoid.



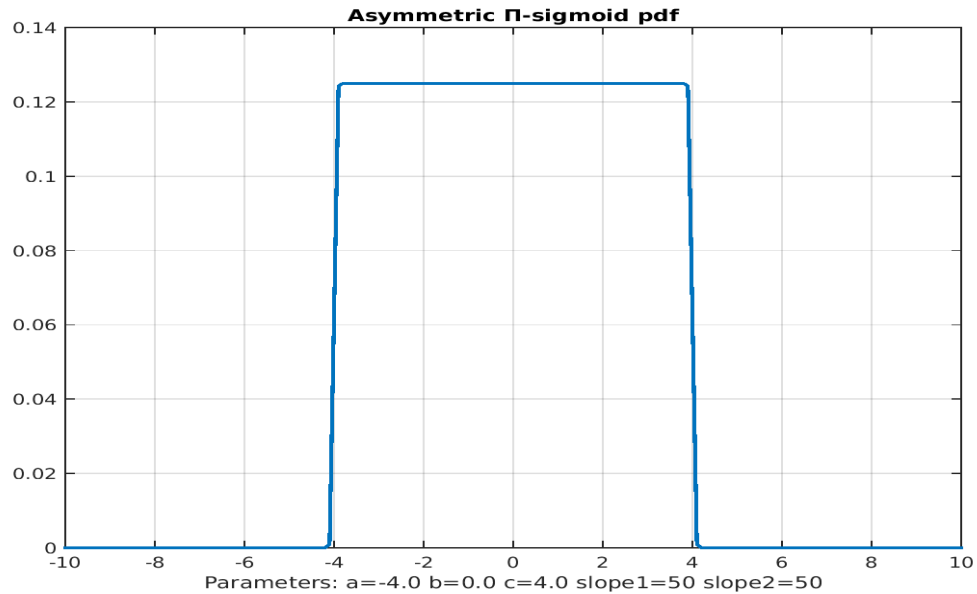
Σχήμα 3.5 Η συνάρτηση πυκνότητας πιθανότητας AsΠ-sigmoid. Παρατηρούμε τη γενική μη-συμμετρική της εκδοχή. Η «ανώμαλη» συμπεριφορά της συνάρτησης στο κέντρο της, είναι αποτέλεσμα του αθροίσματος των δύο Π-sigmoid κατανομών με διαφορετικές κλίσεις.



Σχήμα 3.6 Η συνάρτηση πυκνότητας πιθανότητας AsΠ-sigmoid. Παρατηρούμε την δυνατότητα της κατανομής να πάρει συμμετρικό σχήμα που προσεγγίζει το σχήμα της κανονικής κατανομής, όταν οι δύο Π -sigmoid κατανομές έχουν ίσες κλίσεις με τιμές κοντά στη μονάδα.



Σχήμα 3.7 Η συνάρτηση πυκνότητας πιθανότητας AsΠ-sigmoid. Όσο οι κοινές τιμές στις δύο κλίσεις αυξάνονται, τόσο η μετάβολή των τιμών της συνάρτησης στα κέντρα a, c γίνεται πιο απότομη.



Σχήμα 3.8 Η συνάρτηση πυκνότητας πιθανότητας AsΠ-sigmoid. Με μια μεγάλη αύξηση στις κοινές τιμές των δύο κλίσεων η AsΠ-sigmoid κατανομή προσεγγίζει πλέον το σχήμα της ομοιόμορφης κατανομής.

3.3. Ιδιότητες της asymmetric Π-sigmoid

Έπειτα από την μαθηματική διατύπωση καθώς και τη γραφική περιγραφή της συνάρτησης πυκνότητας πιθανότητας, θα αναλύσουμε τις παραμετρικές της ιδιότητες και θα αποδείξουμε βασικές στατιστικές ιδιότητες.

Οι παράμετροι λ_1, λ_2

Οι δύο άνισες κλίσεις λ_1, λ_2 των δυο Π-sigmoid συνιστωσών κατανομών εξασφαλίζουν την μη-συμμετρική συμπεριφορά της συνάρτησης και δίνουν την απαιτούμενη ευελιξία για την περιγραφή παρατηρήσεων που παράγονται από τυχαίες (μη-συμμετρικές) κατανομές. Επίσης, στην περίπτωση που οι δύο κλίσεις είναι ίσες, $\lambda_1 = \lambda_2$, η κατανομή γίνεται συμμετρική και περιγράφει με επιτυχία δεδομένα και από συμμετρικές κατανομές (κανονικές, ομοιόμορφες κ.α.).

Πιο αναλυτικά, για μεγάλες τιμές της κλίσης λ_1 , η πρώτη συνιστώσα κατανομή προσεγγίζει την ομοιόμορφη κατανομή ενώ για μικρότερες τιμές της κλίσης λ_1 , (δηλ. $\lambda_1 = 1 + \varepsilon$, $\varepsilon \geq 0$), παίρνει μια καμπανοειδή μορφή που προσεγγίζει σχηματικά την κανονική κατανομή. Τέλος, για τιμές κλίσης $\lambda_1 < 1$, η κατανομή αποκτά μεγάλο

άνοιγμα (πλατειάζει), αποκτώντας την δυνατότητα περιγραφής δεδομένων με μεγάλη διασπορά. Ομοίως ισχύει για την συμπεριφορά της δεύτερης συνιστώσας κατανομής μέσω της διαφοροποίησης των τιμών της κλίσης λ_2 .

Συμπεραίνουμε λοιπόν, ότι οι άνισες τιμές στις κλίσεις λ_1, λ_2 των δυο Π-sigmoid συνιστωσών κατανομών, προσδίδουν στο μοντέλο μια ιδιαίτερα ευέλικτη συμπεριφορά, δίνοντας του το πλεονέκτημα της διαφοροποίησης της συμπεριφοράς για κάθε ένα από δύο του τμήματα.

Η σημασία των παραμέτρων a, b, c

Οι τρεις αυτοί παράμετροι της κατανομής, ορίζουν τα κέντρα των δυο συνιστωσών Π-sigmoid κατανομών. Αρχικά, πρέπει να τονιστεί ότι το κοινό κέντρο b που έχουν οι δύο συνιστώσες κατανομές είναι πρωταρχικής σημασίας. Το κέντρο αυτό, εξασφαλίζει την ιδιότητα της συνέχειας της συνάρτησης αφού ορίζει την επικάλυψη (άθροισμα) δύο σιγμοειδών συναρτήσεων με κοινό κέντρο αλλά διαφορετικών κλίσεων.

Επίσης, η παράμετρος b έχει την αντίστοιχη ιδιότητα της μέσης τιμής στην κανονική κατανομή, ενώ οι παράμετροι a, c ορίζουν αντίστοιχες ιδιότητες με την παράμετρο της διασποράς στην κανονική κατανομή. Αύξηση στην απόσταση $|c - a|$, υποδηλώνει μεγάλη διασπορά των παρατηρήσεων.

Τέλος, πρέπει να περιγραφεί η συμπεριφορά της κατανομής για πολύ μεγάλες τιμές κλίσης ($\lambda_1, \lambda_2 \gg 1$). Σε αυτή την περίπτωση, όπως φαίνεται και στο (Σχήμα 3.8), η κατανομή ανέρχεται και κατέρχεται στα σημεία a και c αντίστοιχα, σχεδόν κατακόρυφα, συνδέοντας άμεσα την διασπορά της asymmetric Π-sigmoid κατανομής, με αυτήν της ομοιόμορφης κατανομής, που υπολογίζεται από την σχέση: $\frac{(b-a)^2}{12}$.

Συμπερασματικά λοιπόν, στην έννοια της διασποράς της Π-sigmoid κατανομής, συμμετέχει και η παράμετρος της κλίσης, έχοντας όμως αντιστρόφως ανάλογη συμπεριφορά.

Ακολουθεί η περιγραφή της αθροιστικής συνάρτησης κατανομής, καθώς επίσης και κάποιων βασικών στατιστικών συναρτήσεων.

- **Αθροιστική συνάρτηση (CDF)**

$$\begin{aligned}
 F(x; a, b, c, \lambda_1, \lambda_2) &= \int_{-\infty}^x \frac{1}{2} \left[\frac{1}{b-a} \left[\frac{1}{1+e^{-\lambda_1(u-a)}} - \frac{1}{1+e^{-\lambda_1(u-b)}} \right] \right. \\
 &\quad \left. + \frac{1}{c-b} \left[\frac{1}{1+e^{-\lambda_2(u-b)}} - \frac{1}{1+e^{-\lambda_2(u-c)}} \right] \right] du = \\
 &= \frac{1}{2(b-a)} \int_{-\infty}^x \left[\frac{1}{1+e^{-\lambda_1(u-a)}} - \frac{1}{1+e^{-\lambda_1(u-b)}} \right] du \\
 &\quad + \frac{1}{2(c-b)} \int_{-\infty}^x \left[\frac{1}{1+e^{-\lambda_2(u-b)}} - \frac{1}{1+e^{-\lambda_2(u-c)}} \right] du \\
 &= \frac{\ln(e^{-\lambda_1(x-a)} + 1) - \ln(e^{-\lambda_1(x-b)} + 1) + \lambda_1(b-a)}{2\lambda_1(b-a)} \\
 &\quad + \frac{\ln(e^{-\lambda_2(x-b)} + 1) - \ln(e^{-\lambda_2(x-c)} + 1) + \lambda_2(c-b)}{2\lambda_2(c-b)} \\
 &= \frac{\ln(e^{-\lambda_1(x-a)} + 1) - \ln(e^{-\lambda_1(x-b)} + 1)}{2\lambda_1(b-a)} + \frac{1}{2} \\
 &\quad + \frac{\ln(e^{-\lambda_2(x-b)} + 1) - \ln(e^{-\lambda_2(x-c)} + 1)}{2\lambda_2(c-b)} + \frac{1}{2} \\
 &= \frac{\ln(e^{-\lambda_1(x-a)} + 1) - \ln(e^{-\lambda_1(x-b)} + 1)}{2\lambda_1(b-a)} \\
 &\quad + \frac{\ln(e^{-\lambda_2(x-b)} + 1) - \ln(e^{-\lambda_2(x-c)} + 1)}{2\lambda_2(c-b)} + 1
 \end{aligned} \tag{3.7}$$

- **Η αναμενόμενη τιμή ή κέντρο**

$$\begin{aligned}
 E(X) &= \int_{-\infty}^{+\infty} \frac{1}{2} \left[\frac{1}{b-a} \left[\frac{1}{1+e^{-\lambda_1(x-a)}} - \frac{1}{1+e^{-\lambda_1(x-b)}} \right] \right. \\
 &\quad \left. + \frac{1}{c-b} \left[\frac{1}{1+e^{-\lambda_2(x-b)}} - \frac{1}{1+e^{-\lambda_2(x-c)}} \right] \right] x dx \\
 &= \frac{1}{2(b-a)} \int_{-\infty}^{+\infty} \left(\frac{1}{1+e^{-\lambda_1(x-a)}} - \frac{1}{1+e^{-\lambda_1(x-b)}} \right) x dx \tag{3.8} \\
 &\quad + \frac{1}{2(c-b)} \int_{-\infty}^{+\infty} \left(\frac{1}{1+e^{-\lambda_2(x-b)}} - \frac{1}{1+e^{-\lambda_2(x-c)}} \right) x dx \\
 &= \frac{a+b}{4} + \frac{b+c}{4} = \frac{a+2b+c}{4}
 \end{aligned}$$

Εάν στον τύπο της asymmetric Π-sigmoid λάβουμε τον περιορισμό $b = \frac{a+c}{2}$, τότε η σχέση για την αναμενόμενη τιμή είναι:

$$E(X) = \frac{a + 2b + c}{4} = b \quad (3.9)$$

3.4. Η πολυδιάστατη asymmetric Π-sigmoid κατανομή

Στην παράγραφο αυτή θα παρουσιάσουμε την κατανομή asymmetric Π-sigmoid για πολυδιάστατα δεδομένα, υποθέτοντας ότι η κατανομή είναι ανεξάρτητη στις διαστάσεις των παρατηρήσεων.

Έστω ένα σύνολο από N παρατηρήσεις, διάστασης D και $p(x; \theta)$ μια συνάρτηση πυκνότητας πιθανότητας. Κάθε παρατήρηση έχει την μορφή:

$$x^i = (x_1^i, x_2^i, \dots, x_D^i), \quad i = 1 \dots N$$

Σύμφωνα με την υπόθεση της ανεξαρτησίας της $p(x; \theta)$ σε κάθε διάσταση των παρατηρήσεων, και σύμφωνα με την σχέση 2.16 η συνάρτηση πυκνότητας πιθανότητας έχει την μορφή:

$$p(x; \theta) = \prod_{j=1}^D p(x_j^i; \theta_j), \quad i = 1 \dots N$$

όπου θ_j είναι το διάνυσμα των παραμέτρων της κατανομής σε κάθε διάσταση.

Αντικαθιστώντας στην παραπάνω σχέση τον ορισμό της μονοδιάστατης asymmetric Π-sigmoid έχουμε:

$$\begin{aligned} & AsPs(a, b, c, \lambda_1, \lambda_2) \\ &= \prod_{j=1}^D \left[\frac{1}{2(b_j - a_j)} \left[\frac{1}{1 + e^{-\lambda_{1j}(x_j^i - a_j)}} - \frac{1}{1 + e^{-\lambda_{1j}(x_j^i - b_j)}} \right] \right. \\ & \quad \left. + \frac{1}{2(c_j - b_j)} \left[\frac{1}{1 + e^{-\lambda_{2j}(x_j^i - b_j)}} - \frac{1}{1 + e^{-\lambda_{2j}(x_j^i - c_j)}} \right] \right], \end{aligned} \quad (3.10)$$

$i = 1 \dots N$

Η τελευταία σχέση αποτελεί την πολυδιάστατη εκδοχή της κατανομής asymmetric Π-sigmoid, η οποία προκύπτει ως “τομή” των επιμέρους μονοδιάστατων εκδοχών της. Τέλος, στον τύπο της πολυδιάστατης asymmetric Π-sigmoid παρατηρούμε ότι κάθε διάσταση περιλαμβάνει πέντε ανεξάρτητες παραμέτρους:

$$\alpha_j, b_j, c_j, \lambda_{1j}, \lambda_{2j} \quad j = 1 \dots D$$

Γενικότερα, για την πλήρη περιγραφή μιας πολυδιάστατης asymmetric Π-sigmoid κατανομής, πρέπει να προσδιοριστούν συνολικά $5 * D$ παράμετροι όπου D η διάσταση του χώρου των παρατηρήσεων.

3.5. Εκπαίδευση της asymmetric Π-sigmoid κατανομής

Στην παράγραφο αυτή, θα παρουσιάσουμε την εκτίμηση των παραμέτρων μιας asymmetric Π-sigmoid κατανομής, μέσω της μεγιστοποίησης της συνάρτησης πιθανοφάνειας, όπως ορίστηκε στην σχέση (1.21). Μετά την ολοκλήρωση της διαδικασίας αυτής, οι εκτιμητές μέγιστης πιθανοφάνειας ορίζουν την βέλτιστη περιγραφή του συνόλου των παρατηρήσεων από την asymmetric Π-sigmoid κατανομή.

Έστω λοιπόν ένα σύνολο παρατηρήσεων που υποθέτουμε ότι παράχθηκε ανεξάρτητα από μια πολυδιάστατη asymmetric Π-sigmoid κατανομή:

$$X = \{x_j^i\}, \quad i = 1 \dots N, \quad j = 1 \dots D$$

Τότε, σύμφωνα με τη σχέση (1.21) η συνάρτηση πιθανοφάνειας, έχει την μορφή:

$$\begin{aligned} L(X|\theta) &= \prod_{i=1}^N As\Pi_s(x^i|\theta) \\ &= \prod_{i=1}^N \prod_{j=1}^D \frac{1}{2} \left[\frac{1}{(b_j - a_j)} \left[\frac{1}{1 + e^{-\lambda_{1j}(x_j^i - a_j)}} - \frac{1}{1 + e^{-\lambda_{1j}(x_j^i - b_j)}} \right] \right. \\ &\quad \left. + \frac{1}{(c_j - b_j)} \left[\frac{1}{1 + e^{-\lambda_{2j}(x_j^i - b_j)}} - \frac{1}{1 + e^{-\lambda_{2j}(x_j^i - c_j)}} \right] \right] \end{aligned} \quad (3.11)$$

όπου $\alpha_j, b_j, c_j, \lambda_{1j}, \lambda_{2j} \in \mathbb{R}^D$

Καταφεύγουμε στη συνάρτηση λογαριθμικής πιθανοφάνειας, η οποία σύμφωνα με τη σχέση (1.23) έχει την μορφή:

$$\begin{aligned}
 LL(X|\theta) &= \sum_{i=1}^N \log (AsPs(x^i|\theta)) \\
 &= \sum_{i=1}^N \sum_{j=1}^D \log \left[\frac{1}{2} \left[\frac{1}{(b_j - a_j)} \left[\frac{1}{1 + e^{-\lambda_{1j}(x_j^i - a_j)}} - \frac{1}{1 + e^{-\lambda_{1j}(x_j^i - b_j)}} \right] \right. \right. \\
 &\quad \left. \left. + \frac{1}{(c_j - b_j)} \left[\frac{1}{1 + e^{-\lambda_{2j}(x_j^i - b_j)}} - \frac{1}{1 + e^{-\lambda_{2j}(x_j^i - c_j)}} \right] \right] \right] \quad (3.12)
 \end{aligned}$$

Για την βελτιστοποίηση της παραπάνω σχέσης, πρέπει να υπολογιστούν οι μερικές παράγωγοι για κάθε μια από τις παραμέτρους $\alpha_j, b_j, c_j, \lambda_{1j}, \lambda_{2j}$ $j = 1 \dots D$. Πριν προχωρήσουμε όμως στον υπολογισμό τους, πρέπει να τονιστεί ότι για την ορθή υλοποίηση της διαδικασίας της βελτιστοποίησης, πρέπει να μετασχηματιστεί ο ορισμός της asymmetric Π-sigmoid κατανομής.

Οι περιορισμοί που προκύπτουν από τον ορισμό της κατανομής, πρέπει να ικανοποιούνται σε κάθε λύση που προκύπτει από τον αλγόριθμο EM.

Πιο συγκεκριμένα, πρέπει να ισχύουν οι σχέσεις:

$$\alpha_j < b_j, \quad j = 1 \dots D \quad (3.13)$$

$$b_j < c_j, \quad j = 1 \dots D \quad (3.14)$$

$$\lambda_{1j} \geq 0, \quad j = 1 \dots D \quad (3.15)$$

$$\lambda_{2j} \geq 0, \quad j = 1 \dots D \quad (3.16)$$

Για τον σκοπό αυτό οι τροποποιήσεις που θα πραγματοποιηθούν συνοψίζονται στις παρακάτω αντικαταστάσεις:

$$b_j = \alpha_j + r_{1j}^2, \quad j = 1 \dots D \quad (3.17)$$

$$c_j = b_j + r_{2j}^2, \quad j = 1 \dots D \quad (3.18)$$

$$\lambda_{1j} = g_{1j}^2, \quad j = 1 \dots D \quad (3.19)$$

$$\lambda_{2j} = g_{2j}^2, \quad j = 1 \dots D \quad (3.20)$$

Με τον τρόπο αυτό εξασφαλίζουμε ότι τα κέντρα α_j, b_j, c_j ικανοποιούν πάντοτε τις σχέσεις (3.13-3.14), καθώς ισχύει η ανισότητα $\alpha_j < b_j < c_j$ και επίσης οι κλίσεις $\lambda_{1j}, \lambda_{2j}$ ικανοποιούν τις σχέσεις (3.15-3.16), καθώς αντικαταστάθηκαν από μη-αρνητικές ποσότητες.

Οι παραπάνω αντικαταστάσεις ορίζουν την νέα μορφή της asymmetric Π-sigmoid κατανομής η οποία στην μονοδιάστατη περίπτωση, περιγράφεται από τον τύπο:

$$\begin{aligned} AsPs(x; g_1, g_2, a, r_1, r_2) \\ = \frac{1}{2} \left[\frac{1}{r_1^2} \left[\frac{1}{1 + e^{-g_1^2(x-a)}} - \frac{1}{1 + e^{-g_1^2(x-a-r_1^2)}} \right] \right. \\ \left. + \frac{1}{r_2^2} \left[\frac{1}{1 + e^{-g_2^2(x-a-r_1^2)}} - \frac{1}{1 + e^{-g_2^2(x-a-r_1^2-r_2^2)}} \right] \right] \end{aligned} \quad (3.21)$$

Ενώ στη πολυδιάστατη περίπτωση, περιγράφεται από τον τύπο:

$$\begin{aligned} AsPs(x^i; g_1, g_2, a, r_1, r_2) \\ = \prod_{j=1}^D \frac{1}{2} \left[\frac{1}{r_{1j}^2} \left[\frac{1}{1 + e^{-g_{1j}^2(x_j^i - a_j)}} - \frac{1}{1 + e^{-g_{1j}^2(x_j^i - a_j - r_{1j}^2)}} \right] \right. \\ \left. + \frac{1}{r_{2j}^2} \left[\frac{1}{1 + e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2)}} - \frac{1}{1 + e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2 - r_{2j}^2)}} \right] \right], \\ i = 1 \dots N \end{aligned} \quad (3.22)$$

Οι δύο επιμέρους συνιστώσες Π-sigmoid κατανομές της ανωτέρω asymmetric Π-sigmoid περιγράφονται από τις σχέσεις:

$$Ps(x^i; g_1, a, r_1) = \frac{1}{r_{1j}^2} \left[\frac{1}{1 + e^{-g_{1j}^2(x_j^i - a_j)}} - \frac{1}{1 + e^{-g_{1j}^2(x_j^i - a_j - r_{1j}^2)}} \right] \quad (3.23)$$

$$\begin{aligned} \Pi s(x^i; g_2, a, r_1, r_2) \\ = \frac{1}{r_{2j}^2} \left[\frac{1}{1 + e^{-g_{2j}^2(x_j^i - \alpha_j - r_{1j}^2)}} - \frac{1}{1 + e^{-g_{2j}^2(x_j^i - \alpha_j - r_{1j}^2 - r_{2j}^2)}} \right] \end{aligned} \quad (3.24)$$

Τέλος, η νέα μορφή της λογαριθμικής πιθανοφάνειας θα είναι η παρακάτω:

$$\begin{aligned} LL(X|\theta) &= \sum_{i=1}^N \log \left(As \Pi s(x^i|\theta) \right) \\ &= \sum_{i=1}^N \sum_{j=1}^D \log \left[\frac{1}{2} \left[\frac{1}{r_{1j}^2} \left[\frac{1}{1 + e^{-g_{1j}^2(x_j^i - \alpha_j)}} - \frac{1}{1 + e^{-g_{1j}^2(x_j^i - \alpha_j - r_{1j}^2)}} \right] \right. \right. \\ &\quad \left. \left. + \frac{1}{r_{2j}^2} \left[\frac{1}{1 + e^{-g_{2j}^2(x_j^i - \alpha_j - r_{1j}^2)}} - \frac{1}{1 + e^{-g_{2j}^2(x_j^i - \alpha_j - r_{1j}^2 - r_{2j}^2)}} \right] \right] \right] \end{aligned} \quad (3.25)$$

Για την βελτιστοποίηση της συνάρτησης λογαριθμικής πιθανοφάνειας που περιγράφεται από την παραπάνω σχέση, θα υπολογίσουμε αρχικά τις μερικές παραγώγους ως προς τις παραμέτρους g_1, g_2, a, r_1, r_2 .

Επειδή κάθε μια από αυτές είναι διάνυσμα διάστασης D , κάθε συνιστώσα του διανύσματος αποτελεί και μια ανεξάρτητη παράμετρο ορίζοντας το συνολικό αριθμό παραμέτρων σε $5 * D$.

Οι σχέσεις των μερικών παραγώγων παρουσιάζονται στην συνέχεια:

Μερική παράγωγος ως προς την παράμετρο g_{1j}

$$\begin{aligned} &\frac{\partial LL(X|g_1, g_2, a, r_1, r_2)}{\partial g_{1j}} \\ &= \frac{\partial}{\partial g_{1j}} \left\{ \sum_{i=1}^N \log As \Pi s(x^i; g_1, g_2, a, r_1, r_2) \right\} \\ &= \sum_{i=1}^N \frac{1}{As \Pi s(x^i; g_1, g_2, a, r_1, r_2)} \frac{\partial}{\partial g_{1j}} \left\{ \frac{1}{2} \Pi s(x^i|g_1, a, r_1) \right\} \\ &= \sum_{i=1}^N \frac{\frac{1}{2r_{1j}^2} \left[\frac{e^{-g_{1j}^2(x_j^i - \alpha_j)} 2g_{1j}(x_j^i - \alpha_j)}{(1 + e^{-g_{1j}^2(x_j^i - \alpha_j)})^2} - \frac{e^{-g_{1j}^2(x_j^i - \alpha_j - r_{1j}^2)} 2g_{1j}(x_j^i - \alpha_j - r_{1j}^2)}{(1 + e^{-g_{1j}^2(x_j^i - \alpha_j - r_{1j}^2)})^2} \right]}{As \Pi s(x^i; g_1, g_2, a, r_1, r_2)} \end{aligned} \quad (3.26)$$

Μερική παράγωγος ως προς την παράμετρο g_{2j}

$$\begin{aligned}
 & \frac{\partial LL(X|g_1, g_2, a, r_1, r_2)}{\partial g_{2j}} \\
 &= \frac{\partial}{\partial g_{2j}} \left\{ \sum_{i=1}^N \log As\Pi s(x^i; g_1, g_2, a, r_1, r_2) \right\} \\
 &= \sum_{i=1}^N \frac{1}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \frac{\partial}{\partial g_{2j}} \left\{ \frac{1}{2} \Pi s(x^i|g_2, a, r_1, r_2) \right\} \\
 &= \sum_{i=1}^N \frac{\frac{1}{2r_{2j}^2} \left[\frac{e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2)} 2g_{2j}(x_j^i - \alpha_j - r_{1j}^2)}{(1 + e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2)})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \\
 &\quad - \sum_{i=1}^N \frac{\frac{1}{2r_{2j}^2} \left[\frac{e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2 - r_{2j}^2)} 2g_{2j}(x_j^i - \alpha_j - r_{1j}^2 - r_{2j}^2)}{(1 + e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2 - r_{2j}^2)})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)}
 \end{aligned} \tag{3.27}$$

Μερική παράγωγος ως προς την παράμετρο a_j

$$\begin{aligned}
 & \frac{\partial LL(X|g_1, g_2, a, r_1, r_2)}{\partial a_j} \\
 &= \frac{\partial}{\partial a_j} \left\{ \sum_{i=1}^N \log As\Pi s(x^i; g_1, g_2, a, r_1, r_2) \right\} \\
 &= \sum_{i=1}^N \frac{1}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \frac{\partial}{\partial a_j} \{ As\Pi s(x^i; g_1, g_2, a, r_1, r_2) \} \\
 &= \sum_{i=1}^N \frac{\frac{1}{2r_{1j}^2} \left[\frac{-e^{-g_{1j}^2(x_j^i - a_j)} g_{1j}^2}{(1 + e^{-g_{1j}^2(x_j^i - a_j)})^2} + \frac{e^{-g_{1j}^2(x_j^i - a_j - r_{1j}^2)} g_{1j}^2}{(1 + e^{-g_{1j}^2(x_j^i - a_j - r_{1j}^2)})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \\
 &\quad + \sum_{i=1}^N \frac{\frac{1}{2r_{2j}^2} \left[\frac{e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2)} g_{2j}^2}{(1 + e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2)})^2} + \frac{e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2 - r_{2j}^2)} g_{2j}^2}{(1 + e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2 - r_{2j}^2)})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)}
 \end{aligned} \tag{3.28}$$

Μερική παράγωγος ως προς την παράμετρο r_{1j}

$$\begin{aligned}
& \frac{\partial LL(X|g_1, g_2, a, r_1, r_2)}{\partial r_{1j}} \\
&= \frac{\partial}{\partial r_{1j}} \left\{ \sum_{i=1}^N \log As\Pi s(x^i; g_1, g_2, a, r_1, r_2) \right\} \\
&= \sum_{i=1}^N \frac{1}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \frac{\partial}{\partial r_{1j}} \{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)\} \\
&= \sum_{i=1}^N \frac{-\frac{1}{r_{1j}} \Pi s(x^i; g_1, a, r_1) + \frac{1}{2r_{1j}^2} \left[\frac{e^{-g_{1j}^2(x_j^i - a_j - r_{1j}^2)} 2r_{1j} g_{1j}^2}{(1 + e^{-g_{1j}^2(x_j^i - a_j - r_{1j}^2)})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \\
&+ \sum_{i=1}^N \frac{\frac{1}{2r_{2j}^2} \left[\frac{-e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2)} 2r_{1j} g_{2j}^2}{(1 + e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2)})^2} + \frac{e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2 - r_{2j}^2)} 2r_{1j} g_{2j}^2}{(1 + e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2 - r_{2j}^2)})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)}
\end{aligned} \tag{3.29}$$

Μερική παράγωγος ως προς την παράμετρο r_{2j}

$$\begin{aligned}
& \frac{\partial LL(X|g_1, g_2, a, r_1, r_2)}{\partial r_{2j}} \\
&= \frac{\partial}{\partial r_{2j}} \left\{ \sum_{i=1}^N \log As\Pi s(x^i; g_1, g_2, a, r_1, r_2) \right\} \\
&= \sum_{i=1}^N \frac{1}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \frac{\partial}{\partial r_{2j}} \left\{ \frac{1}{2} \Pi s(x^i|g_2, a, r_1, r_2) \right\} \\
&= \sum_{i=1}^N \frac{-\frac{1}{r_{2j}} \Pi s(x^i; g_2, a, r_1, r_2) + \frac{1}{2r_{2j}^2} \left[\frac{e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2 - r_{2j}^2)} 2r_{2j} g_{2j}^2}{(1 + e^{-g_{2j}^2(x_j^i - a_j - r_{1j}^2 - r_{2j}^2)})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)}
\end{aligned} \tag{3.30}$$

Συμπερασματικά λοιπόν, το αποτέλεσμα των μερικών παραγώγων ως προς κάθε παράμετρο, είναι ένα διάνυσμα διάστασης D , όπου κάθε συνιστώσα του διανύσματος αποτελεί την μερική παράγωγο της αντικειμενικής συνάρτησης, ως προς μια εκ των παραμέτρων g_1, g_2, a, r_1, r_2 σε μια από τις διαστάσεις $j = 1 \dots D$.

Όπως παρατηρούμε από τις σχέσεις των μερικών παραγώγων, δεν είναι εύκολο να πάρουμε μια αναλυτική λύση εξισώνοντας με το μηδέν. Για την βελτιστοποίηση της συνάρτησης της λογαριθμικής πιθανοφάνειας θα χρησιμοποιήσουμε τεχνικές αριθμητικής βελτιστοποίησης. Πιο συγκεκριμένα, θα εφαρμόσουμε την μέθοδο BFGS, που ανήκει στην οικογένεια των Quasi-Newton μεθόδων. Λεπτομέρειες της μεθόδου μπορεί κανείς να βρει στο βιβλίο των Nocedal & Wright [6].

Ο αλγόριθμος αριθμητικής βελτιστοποίησης BFGS, πέρα από την αντικειμενική συνάρτηση, λαμβάνει ως είσοδο το διάνυσμα των μερικών παραγώγων καθώς και μια αρχική τιμή για το διάνυσμα παραμέτρων της αντικειμενικής συνάρτησης.

Πιο αναλυτικά, ως αντικειμενική συνάρτηση f στην περίπτωση μας ορίζεται η συνάρτηση λογαριθμικής πιθανοφάνειας που περιγράφεται στην σχέση (3.25) και εξαρτάται από το διάνυσμα παραμέτρων $\theta = [g_1, g_2, a, r_1, r_2]$. Επειδή υποθέτουμε ανεξαρτησία της asymmetric Π-sigmoid κατανομής ανά διάσταση, θα εκτελέσουμε συνολικά D -διαδικασίες βελτιστοποίησης και θα προκύψουν D -διανύσματα εκτιμήσεων των παραμέτρων, ένα για κάθε διάσταση.

Το διάνυσμα μερικών παραγώγων ανά διάσταση ισούται με:

$$\nabla f = \left[\frac{\partial LL(X|\theta)}{\partial g_{1j}}, \frac{\partial LL(X|\theta)}{\partial g_{2j}}, \frac{\partial LL(X|\theta)}{\partial a_j}, \frac{\partial LL(X|\theta)}{\partial r_{1j}}, \frac{\partial LL(X|\theta)}{\partial r_{2j}} \right], \quad j = 1 \dots D \quad (3.31)$$

ενώ κάθε συνιστώσα του παραπάνω διανύσματος, προκύπτει αναλυτικά από τις σχέσεις (3.26)-(3.30).

Επίσης, μέσω των αντικαταστάσεων (3.17)-(3.20) η διαδικασία της βελτιστοποίησης μπορεί να γίνει πλέον αποδοτικά, καθώς οι τροποποιημένοι ορισμοί των συναρτήσεων περιλαμβάνουν όλους τους περιορισμούς που θέτει η asymmetric Π-sigmoid κατανομή. Σε αντίθετη περίπτωση, θα έπρεπε να εκτελεστεί βελτιστοποίηση με περιορισμούς αυξάνοντας σημαντικά την απόδοση και την πολυπλοκότητα του αλγορίθμου BFGS.

Αυτό που εκκρεμεί για την εκκίνηση του αλγορίθμου, είναι η αρχικοποίηση του διανύσματος παραμέτρων $\theta = [g_1, g_2, a, r_1, r_2]$ της λογαριθμικής πιθανοφάνειας.

Μια αρχικοποίηση των παραμέτρων κοντά στη βέλτιστη λύση εξασφαλίζει την άμεση και αποδοτική εξέλιξη της διαδικασίας της βελτιστοποίησης. Για τον σκοπό αυτό, αρχικά θα εκτελέσουμε την διαδικασία της εκτίμησης μέγιστης πιθανοφάνειας για την κανονική κατανομή και θα προσδιορίσουμε εκτιμητές μέγιστης πιθανοφάνειας (μέση τιμή μ , πίνακας συμμεταβλητότητας Σ) για το σύνολο των παρατηρήσεων που εξετάζουμε. Έπειτα, έχοντας δεδομένες τις παραμέτρους μ , Σ θα προσδιορίσουμε συσχετισμούς με τις παραμέτρους της asymmetric Π-sigmoid κατανομής έτσι να προσδιορίσουμε μια αρχικοποίηση, όσο το δυνατόν πιο κοντά στην βέλτιστη λύση.

Οι συσχετισμοί των παραμέτρων μεταξύ των δύο κατανομών προκύπτουν από τις γεωμετρικές ιδιότητες της μέσης τιμής και του πίνακα συμμεταβλητότητας σε σχέση με τα τρία κέντρα a, b, c της asymmetric Π-sigmoid κατανομής (οι παράμετροι της κλίσης, είναι δύσκολο να συσχετιστούν, οπότε αρχικοποιούνται αυθαίρετα).

Όπως έχει ήδη αναλυθεί, η asymmetric Π-sigmoid κατανομή με κατάλληλες (ίσες) τιμές στις παραμέτρους λ_1, λ_2 έχει την δυνατότητα να σχηματίσει μια καμπανοειδή μορφή η οποία προσεγγίζει με ακρίβεια την γεωμετρική μορφή της κανονικής κατανομής.

Επιπρόσθετα, μέσω της συγκεκριμένης αρχικοποίησης, επιδιώκουμε η asymmetric Π-sigmoid κατανομή να προσεγγίζει (αρχικά) τις γεωμετρικές ιδιότητες της κανονικής κατανομής, έτσι ώστε στα πρώτα βήματα να έχει μια αντίστοιχη συμπεριφορά, ικανή να περιγράψει διάφορα σύνολα παρατηρήσεων.

Σε μονοδιάστατα σύνολα παρατηρήσεων, μπορούμε να αντιστοιχίσουμε το κέντρο b της asymmetric Π-sigmoid κατανομής με την μέση τιμή μ της κανονικής κατανομής, καθώς και τα κέντρα a, c με τις ποσότητες $\mu - k_1\sigma$ και $\mu + k_2\sigma$ αντίστοιχα, όπου $k_1, k_2 \in \mathbb{R}$ κατάλληλες σταθερές.. Για την αρχικοποίηση μιας μονοδιάστατης asymmetric Π-sigmoid κατανομής, επιλέχθηκαν οι τιμές $k_1 = k_2 = \frac{3}{2}$ οπότε έχουμε τις σχέσεις:

$$\begin{aligned} a &= \mu - \frac{3}{2}\sigma \\ b &= \mu \\ c &= \mu + \frac{3}{2}\sigma \end{aligned} \tag{3.32}$$

Επίσης, με έναν σχεδόν αυθαίρετο τρόπο αρχικοποιούμε τις άνισες κλίσεις λ_1, λ_2 με σχετικά μικρές τιμές, έτσι ώστε στα πρώτα βήματα της διαδικασίας της βελτιστοποίησης η κατανομή να παρουσιάζει σχετικά μεγάλη διασπορά και στην συνέχεια να υπάρχει ευχέρεια στην προσαρμογή του σχήματος της.

Συμπερασματικά, προκύπτουν οι παρακάτω αρχικές τιμές των παραμέτρων:

$$\begin{aligned}
 g_1 &= 1.95 \\
 g_2 &= 2.05 \\
 \alpha &= \mu - \frac{3}{2}\sigma \\
 r_1 &= \sqrt{\frac{3}{2}\sigma} \\
 r_2 &= \sqrt{\frac{3}{2}\sigma}
 \end{aligned} \tag{3.33}$$

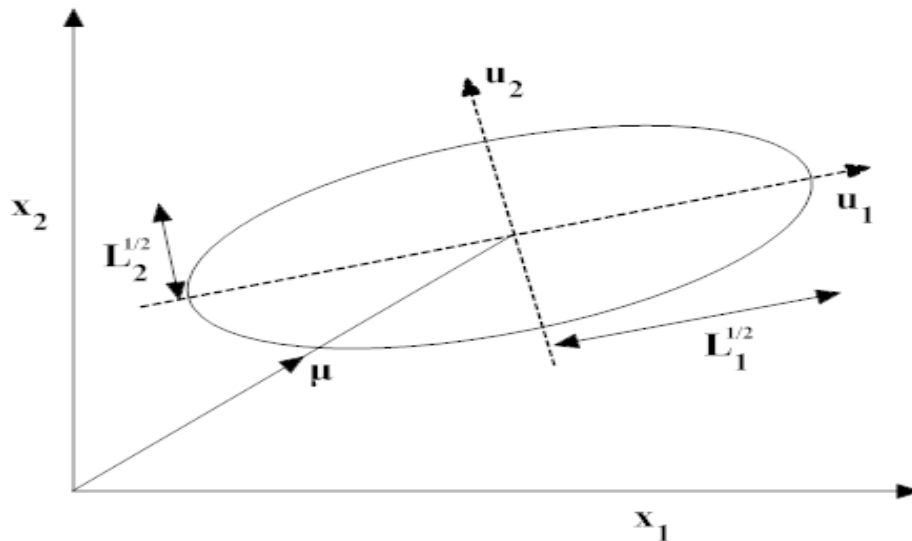
Γενικότερα, στην πολυδιάστατη περίπτωση, θα θέλαμε τα υπερεπίπεδα $x - a_d$ και $x - a_d - r_{1d}^2 - r_{2d}^2$ να είναι παράλληλα με το ιδιοδιάνυσμα u_d του πίνακα συμμεταβλητότητας Σ . Ταυτόχρονα, πρέπει να είναι τοποθετημένα εκατέρωθεν της μέσης τιμής μ και η απόσταση τους από αυτή να είναι ίση με $0.5L_d^{1/2}$ (όπου L_d είναι ιδιοτιμή του πίνακα Σ).

Συνοψίζοντας την παραπάνω διαδικασία αρχικοποίησης στη πολυδιάστατη περίπτωση, προκύπτουν οι σχέσεις:

$$\begin{aligned}
 g_{1j} &= 1.95 \\
 g_{2j} &= 2.05 \\
 a_j &= \mu - \sqrt{L_j} \\
 r_{1j} &= \sqrt{2\sqrt{L_j}} \\
 r_{2j} &= \sqrt{2\sqrt{L_j}}
 \end{aligned} \tag{3.34}$$

Οι τρεις τελευταίες εκτιμήσεις προκύπτουν από τον τύπο υπολογισμού απόστασης μεταξύ δύο υπερεπιπέδων. Πιο αναλυτικά, για τον προσδιορισμό μιας αρχικοποίησης των παραμέτρων της j -οστής διάστασης, αρχικά υπολογίζεται ένα υπερεπίπεδο e_j , παράλληλο με το j -οστό ιδιοδιάνυσμα u_j του πίνακα συμμεταβλητότητας Σ που διέρχεται από το σημείο μ .

Για τον υπολογισμό της αρχικοποίησης των παραμέτρων a_j, c_j της πολυδιάστατης asymmetric Π-sigmoid κατανομής υπολογίζονται τα υπερεπίπεδα που είναι παράλληλα με το ε_j και βρίσκονται εκατέρωθεν αυτού, σε απόσταση $\sqrt{L_j}$ το καθένα.

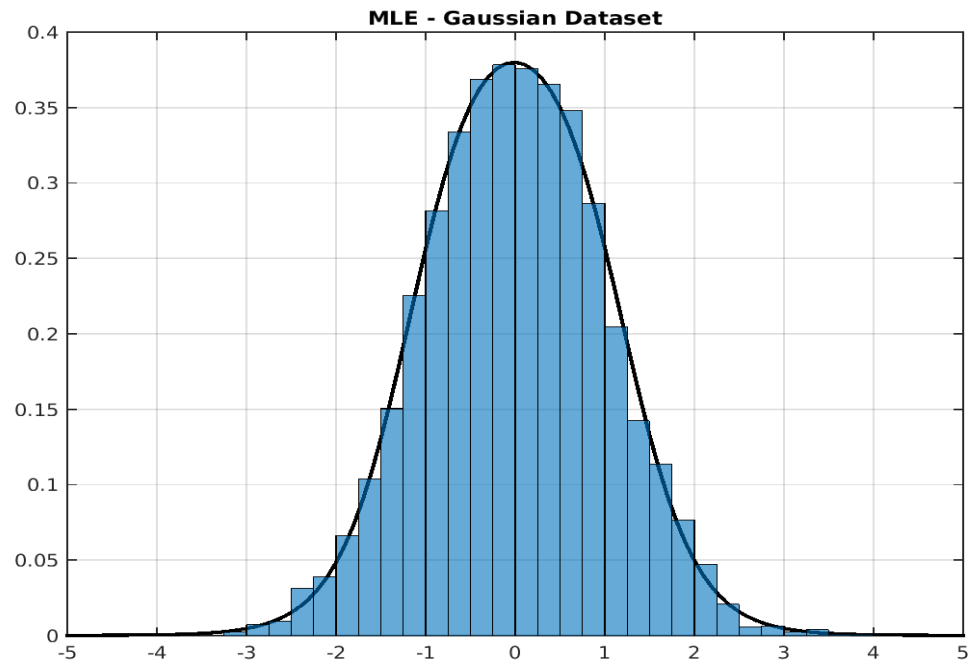


Σχήμα 3.9 Ένα περίγραμμα της κανονικής κατανομής στις δύο διαστάσεις η οποία χαρακτηρίζεται από το κέντρο μ και πίνακα συμμεταβλητότητας του οποίου τα ιδιοδιανύσματα είναι u_1 και u_2 , με αντίστοιχες ιδιοτιμές L_1 και L_2 .

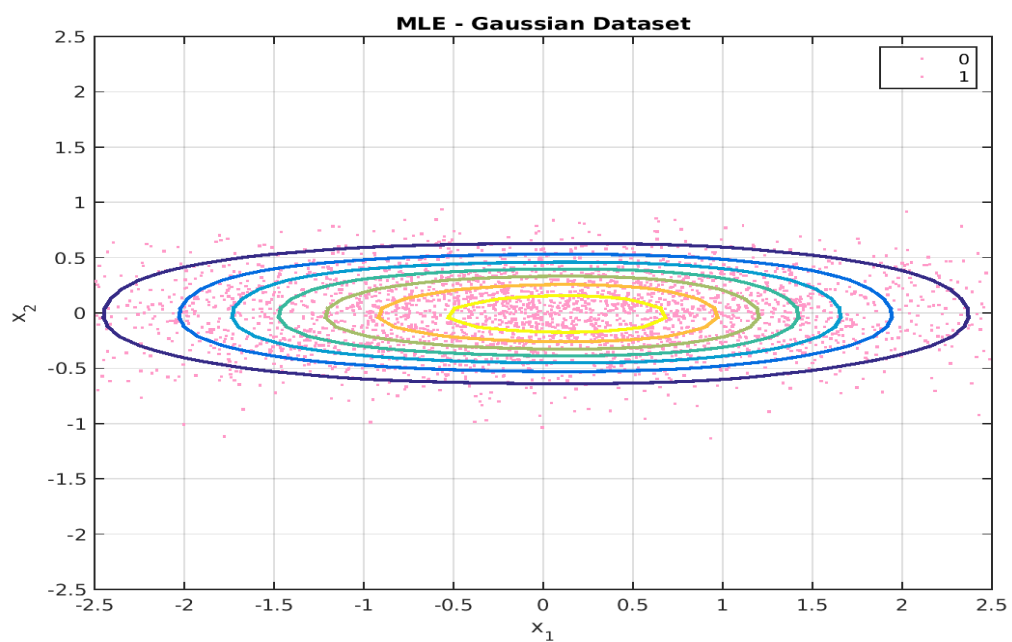
Έχοντας ολοκληρώσει την διαδικασία υπολογισμού των μερικών παραγώγων της λογαριθμικής συνάρτησης πιθανοφάνειας, καθώς και της αρχικοποίησης του διανύσματος παραμέτρων της, θα εφαρμοστεί η διαδικασία της βελτιστοποίησης μέσω της μεθόδου BFGS.

Στην διαδικασία της βελτιστοποίησης στοχεύουμε στην μεγιστοποίηση της λογαριθμικής πιθανοφάνειας (ή ισοδύναμα της πιθανοφάνειας). Για τεχνικούς όμως λόγους στην θεωρία της βελτιστοποίησης επιλέγουμε την ελαχιστοποίηση μιας αντικειμενικής συνάρτησης. Για τον σκοπό αυτό, θεωρούμε ως αντικειμενική συνάρτηση την αρνητική λογαριθμική πιθανοφάνεια, καθώς και το αντίθετο διάνυσμα των μερικών παραγώγων που υπολογίστηκε στη σχέση (3.31).

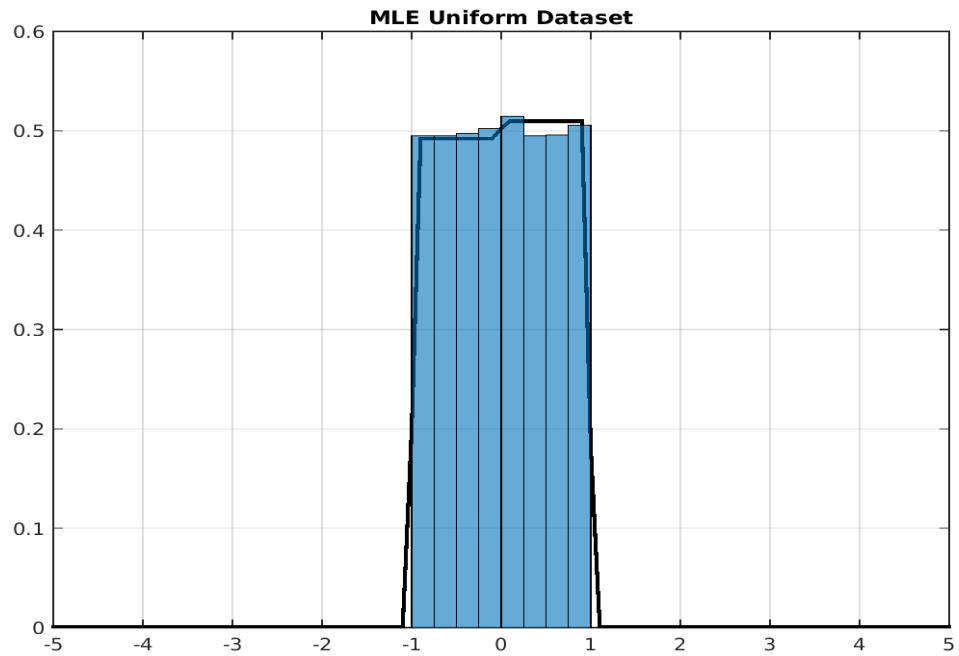
Στα Σχήματα 3.10 ως 3.15 παρουσιάζονται τα αποτελέσματα της εφαρμογής της μεθόδου MLE για την κατανομή asymmetric Π-sigmoid. Η μέθοδος εφαρμόζεται σε τρία ανεξάρτητα σύνολα παρατηρήσεων, με κανονικά, ομοιόμορφα και truncated κανονικά δεδομένα αντίστοιχα. Πειραματικά αποτελέσματα παρουσιάζονται στο επόμενο κεφάλαιο.



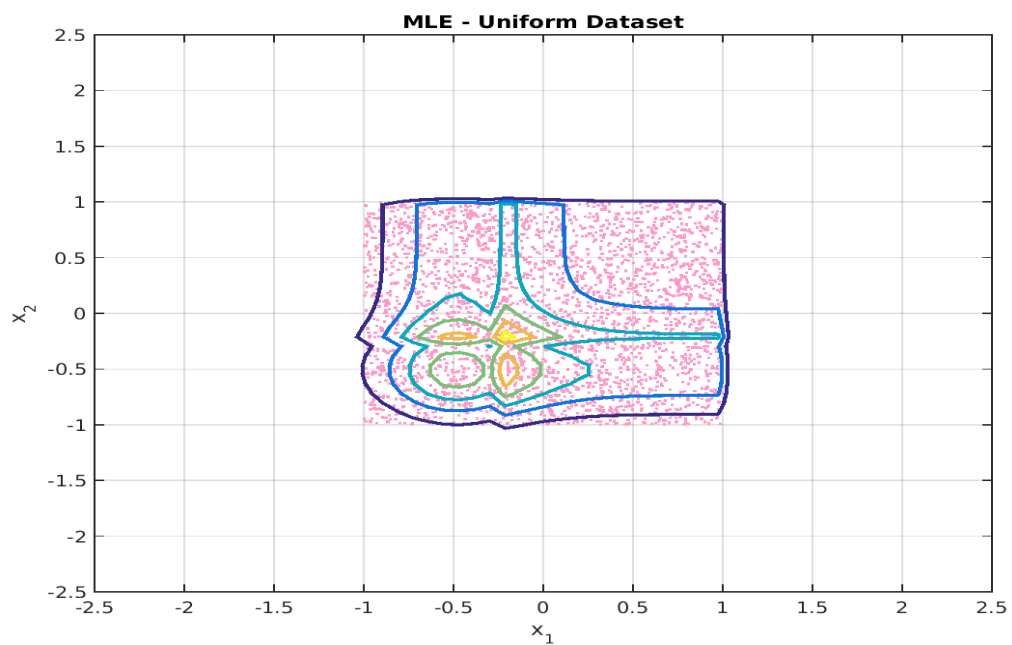
Σχήμα 3.10 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε μονοδιάστατο σύνολο κανονικών παρατηρήσεων.



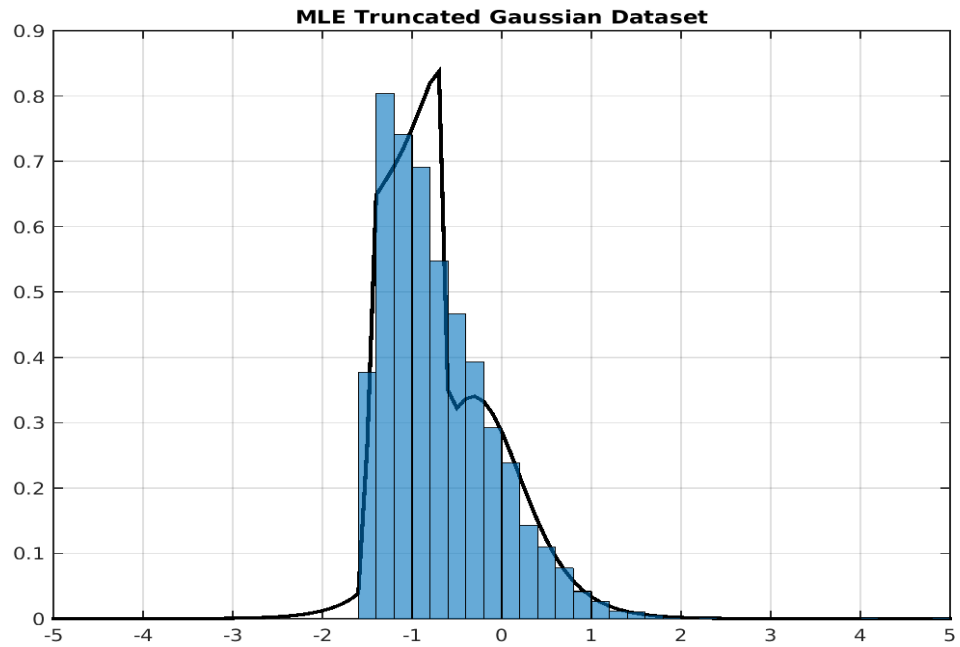
Σχήμα 3.11 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε δισδιάστατο σύνολο κανονικών παρατηρήσεων.



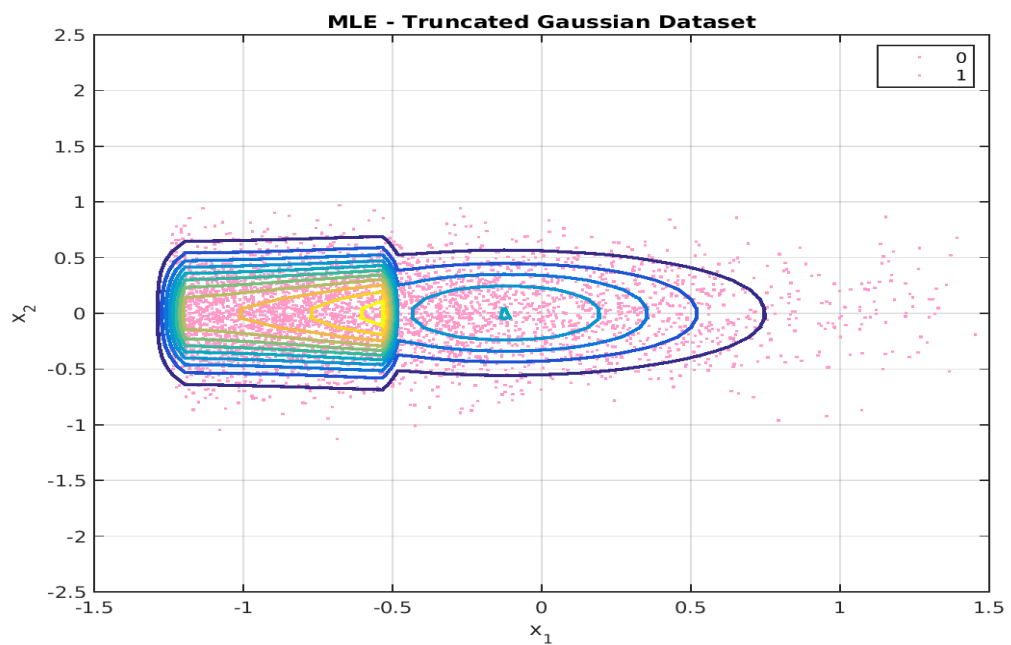
Σχήμα 3.12 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε μονοδιάστατο σύνολο ομοιόμορφων παρατηρήσεων.



Σχήμα 3.13 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε δισδιάστατο σύνολο ομοιόμορφων παρατηρήσεων.



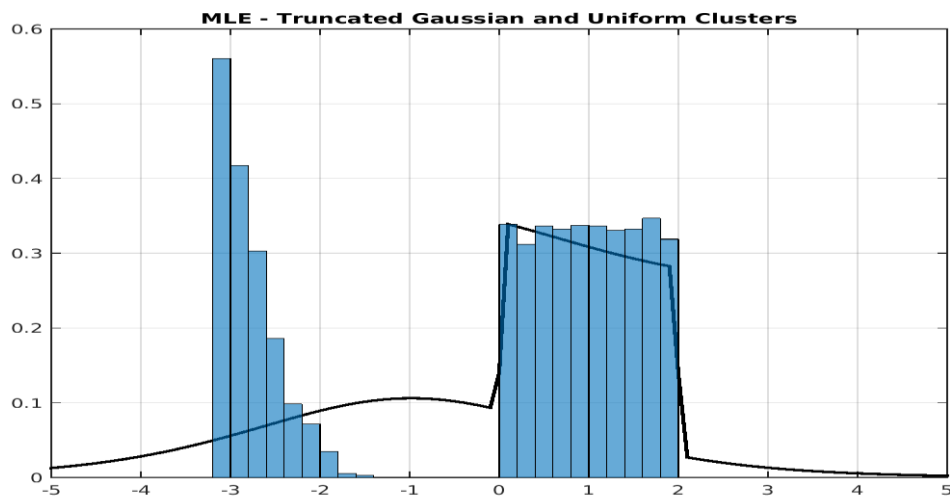
Σχήμα 3.14 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε μονοδιάστατο σύνολο truncated κανονικών παρατηρήσεων.



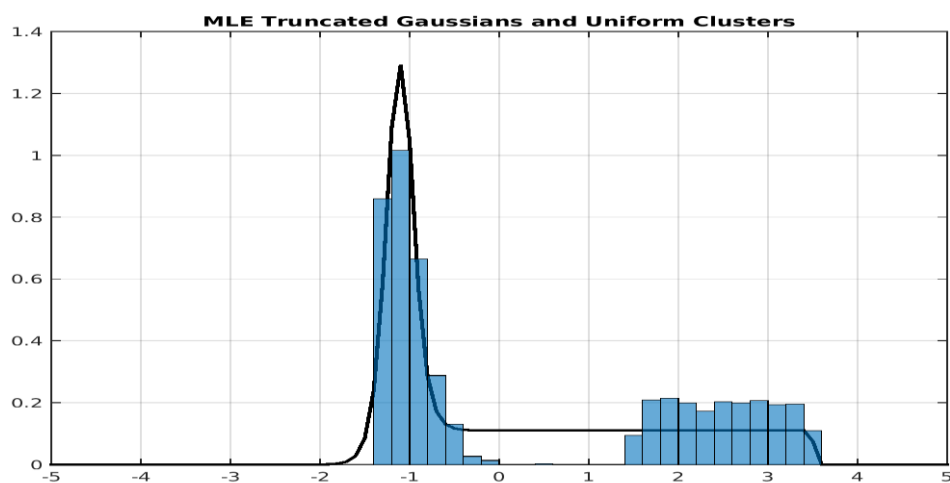
Σχήμα 3.15 Εφαρμογή της μεθόδου εκτίμησης μέγιστης πιθανοφάνειας σε δισδιάστατο σύνολο truncated κανονικών παρατηρήσεων.

3.6. Μικτά μοντέλα asymmetric Π-sigmoid κατανομών

Παρόλο που μια asymmetric Π-sigmoid κατανομή έχει μεγάλη ευελιξία στην περιγραφή παρατηρήσεων που παράγονται από διαφορετικές κατανομές, συναντά σημαντικά προβλήματα στην περιγραφή πραγματικών συνόλων παρατηρήσεων. Όπως φαίνεται στα Σχήματα (3.16)-(3.17), η περιγραφή ενός συνόλου παρατηρήσεων που σχηματίζει πολλαπλές ομάδες μέσω μιας asymmetric Π-sigmoid κρίνεται ανεπαρκής έχοντας μειωμένη ακρίβεια.



Σχήμα 3.16 Ανεπιτυχής περιγραφή μέσω μιας AsΠ-sigmoid δύο ομάδων παρατηρήσεων



Σχήμα 3.17 Ανεπιτυχής περιγραφή μέσω μιας AsΠ-sigmoid δύο ομάδων παρατηρήσεων.

Για την μοντελοποίηση ενός σύνθετου συνόλου παρατηρήσεων καταφεύγουμε σε ένα μοντέλο μικτών κατανομών, το οποίο παρουσιάζει πολύ ισχυρές δυνατότητες μοντελοποίησης πραγματικών συνόλων παρατηρήσεων. Σύμφωνα με την σχέση (1.13) ένα μοντέλο μικτών asymmetric Π-sigmoid κατανομών (AsPIsMM) αποτελεί έναν γραμμικό συνδυασμό πεπερασμένων asymmetric Π-sigmoid κατανομών, σταθμισμένων με βάρη. Το μοντέλο για K κατανομές έχει την μορφή:

$$p(x) = \sum_{k=1}^K \pi_k AsPIs(x; g_{1k}, g_{2k}, a_k, r_{1k}, r_{2k}) \quad (3.35)$$

Κάθε asymmetric Π-sigmoid κατανομή $AsPIs(x; g_1, g_2, a, r_1, r_2), k = 1 \dots K$ είναι συνιστώσα του μικτού μοντέλου και προσδιορίζεται από τις παραμέτρους $g_{1k}, g_{2k}, a_k, r_{1k}, r_{2k}$.

Τα π_k ονομάζονται παράμετροι μίξης και πρέπει να ικανοποιούν τους παρακάτω περιορισμούς:

$$0 \leq \pi_k \leq 1, \quad \forall k = 1 \dots K$$

$$\sum_{k=1}^K \pi_k = 1$$

Συνοψίζοντας, όταν το πλήθος των συνιστωσών K είναι γνωστό, τότε ένα μοντέλο μικτών asymmetric Π-sigmoid κατανομών περιγράφεται πλήρως από τις παρακάτω παραμέτρους:

$$\pi = \{\pi_1, \dots, \pi_K\}, \forall k = 1 \dots K$$

$$g_1 = \{g_{11}, \dots, g_{1K}\}, \forall k = 1 \dots K$$

$$g_2 = \{g_{21}, \dots, g_{2K}\}, \forall k = 1 \dots K$$

$$a = \{a_1, \dots, a_K\}, \forall k = 1 \dots K$$

$$r_1 = \{r_{11}, \dots, r_{1K}\}, \forall k = 1 \dots K$$

$$r_2 = \{r_{21}, \dots, r_{2K}\}, \forall k = 1 \dots K$$

Για τον προσδιορισμό τους, χρησιμοποιείται αποδοτικά ο αλγόριθμος GEM και θα παρουσιαστεί αναλυτικά στην επόμενη παράγραφο.

3.7. Εκπαίδευση ενός AsPisMM μέσω του GEM αλγορίθμου

Σε αυτή την ενότητα θα παρουσιάσουμε μια αποδοτική μέθοδο με την οποία θα προσδιοριστούν κάποιες ικανοποιητικές (ιδανικά οι βέλτιστες) εκτιμήσεις για τις παραμέτρους ενός μικτού μοντέλου asymmetric Π-sigmoid κατανομών όπως περιγράφεται από την σχέση (3.35).

Όπως έχει ήδη αναφερθεί, για την επιτυχή εκτίμηση των παραμέτρων θα γίνει βελτιστοποίηση της συνάρτησης της λογαριθμικής πιθανοφάνειας. Η διαδικασία της βελτιστοποίησης θα γίνει μέσω του αλγορίθμου βελτιστοποίησης Generalized Expectation Maximization (GEM) που παρουσιάστηκε στο πρώτο κεφάλαιο καθώς το μικτό μοντέλο, εξαρτάται από κρυμμένες μεταβλητές.

Η επιλογή του συγκεκριμένου αλγορίθμου γίνεται διότι η μεγιστοποίηση του κάτω φράγματος G της λογαριθμικής πιθανοφάνειας που υπολογίζεται στο E-βήμα (Expectation step) δεν μπορεί να γίνει αναλυτικά. Στην περίπτωση αυτή θα εφαρμοστούν μέθοδοι αριθμητικής βελτιστοποίησης, η χρήση των οποίων δεν οδηγεί πάντα στην εύρεση ολικού μεγίστου, αλλά σε μια καλύτερη εκτίμηση παραμέτρων από τις αρχικοποιημένες τιμές με τις οποίες γίνεται εκκίνηση του αλγορίθμου.

Αρχικά, θεωρούμε ότι ο αριθμός K των συνιστωσών κατανομών είναι γνωστός και θα παρουσιάσουμε αναλυτικά την διαδικασία εκπαίδευσης μέσω του αλγορίθμου GEM με την οποία υπολογίζονται οι βέλτιστες εκτιμήσεις των παραμέτρων. Στην επόμενη παράγραφο, θα εντάξουμε στην διαδικασία της εκπαίδευσης το κριτήριο dip και θα παρουσιάσουμε έναν τελικό αλγόριθμο εκπαίδευσης, που θα υπολογίζει το πλήθος των συνιστωσών κατανομών του μικτού μοντέλου καθώς και τις βέλτιστες εκτιμήσεις για κάθε συνιστώσα asymmetric Π-sigmoid κατανομή.

3.7.1. Αρχικοποίηση του GEM αλγορίθμου

Όπως έχει ήδη τονιστεί, η απόδοση του αλγορίθμου GEM εξαρτάται σε πολύ μεγάλο βαθμό από τις αρχικές τιμές από τις οποίες θα ξεκινήσει η διαδικασία της βελτιστοποίησης. Η τεχνική που θα χρησιμοποιήσουμε αποτελεί γενίκευση της μεθόδου αρχικοποίησης στην εκτίμηση μέγιστης πιθανοφάνειας (MLE) για μια asymmetric Π-sigmoid κατανομή.

Για να επιτύχουμε μια αποδοτική αρχικοποίηση του AsPsMM, θα εκπαιδευτεί αρχικά ένα GMM με το ίδιο πλήθος συνιστωσών που θα έχει το αντίστοιχο AsPsMM και στην συνέχεια θα εντοπιστούν κατάλληλοι συσχετισμοί μεταξύ των παραμέτρων της κάθε συνιστώσας κατανομής του GMM με αυτές του AsPsMM.

Στην μονοδιάστατη περίπτωση, έστω μ_k, σ_k^2 οι παράμετροι της k -οστής συνιστώσας του GMM και g, π_k η αντίστοιχη παράμετρος μίξης (prior πιθανότητα). Μπορούμε να αναθέσουμε στα κέντρα b_k της asymmetric Π-sigmoid κατανομής την μέση τιμή μ_k της κάθε κανονικής κατανομής $k = 1 \dots K$, καθώς και στα κέντρα a_k, c_k τις ποσότητες $\mu_k - k_1 \sigma_k$ και $\mu_k + k_2 \sigma_k$ αντίστοιχα, όπου $k_1, k_2 \in \mathbb{R}$ κατάλληλες σταθερές. Στην αρχικοποίηση ενός μικτού μοντέλου μονοδιάστατων asymmetric Π-sigmoid κατανομών, επιλέχθηκαν οι τιμές $k_1 = k_2 = \frac{3}{2}$ οπότε παίρνουμε τις σχέσεις:

$$\begin{aligned} a_k &= \mu_k - \frac{3}{2} \sigma_k, \forall k = 1 \dots K \\ b_k &= \mu_k, \forall k = 1 \dots K \\ c_k &= \mu_k + \frac{3}{2} \sigma_k, \forall k = 1 \dots K \end{aligned} \tag{3.36}$$

Επίσης, όπως αναφέρθηκε, με έναν σχεδόν αυθαίρετο τρόπο αρχικοποιούμε τις άνισες κλίσεις λ_1, λ_2 με σχετικά μικρές τιμές. Τέλος, εξισώνουμε τις παραμέτρους μίξης του AsPsMM με αυτές του GMM.

Συμπερασματικά, προκύπτουν οι παρακάτω αρχικές τιμές των παραμέτρων:

$$\begin{aligned}
\pi_k &= g_{\pi_k}, \forall k = 1 \dots K \\
g_{1k} &= 1.95, \forall k = 1 \dots K \\
g_{2k} &= 2.05, \forall k = 1 \dots K \\
a_k &= \mu - \frac{3}{2}\sigma, \forall k = 1 \dots K \\
r_{1k} &= \sqrt{\frac{3}{2}\sigma_k}, \forall k = 1 \dots K \\
r_{k2} &= \sqrt{\frac{3}{2}\sigma_k}, \forall k = 1 \dots K
\end{aligned} \tag{3.37}$$

Γενικότερα, στην πολυδιάστατη περίπτωση, υποθέτουμε ανεξαρτησία των παρατηρήσεων ανά διάσταση, οπότε κάθε διάσταση περιλαμβάνει ένα ανεξάρτητο σύνολο παραμέτρων το οποίο πρέπει να αρχικοποιηθεί κατάλληλα.

Για τον σκοπό αυτό, έστω μ_k , Σ_k οι παράμετροι της k -οστής συνιστώσας του GMM και g_{π_k} η αντίστοιχη παράμετρος μίξης (prior πιθανότητα). Ακόμη, έστω u_{kj} τα ιδιοδιανύσματα και L_{kj} οι αντίστοιχες ιδιοτιμές του πίνακα Σ_k ($j \in \mathbb{R}^D$).

Ο συσχετισμός των παραμέτρων μιας asymmetric Π-sigmoid κατανομής με τις παραμέτρους μιας κανονικής κατανομής έχει ήδη περιγραφεί, οπότε συνοψίζουμε την διαδικασία αρχικοποίησης με στις παρακάτω τιμές των παραμέτρων:

$$\begin{aligned}
\pi_k &= g_{\pi_k}, \forall k = 1 \dots K, j = 1 \dots D \\
g_{1kj} &= 1.95, \forall k = 1 \dots K, j = 1 \dots D \\
g_{2kj} &= 2.05, \forall k = 1 \dots K, j = 1 \dots D \\
a_{kj} &= \mu_k - \sqrt{L_{kj}}, \forall k = 1 \dots K, j = 1 \dots D \\
r_{1kj} &= \sqrt{2\sqrt{L_{kj}}}, \forall k = 1 \dots K, j = 1 \dots D \\
r_{2kj} &= \sqrt{2\sqrt{L_{kj}}}, \forall k = 1 \dots K, j = 1 \dots D
\end{aligned} \tag{3.38}$$

3.7.2. E-βήμα (Expectation step)

Σε αυτό το βήμα του αλγορίθμου, (όπως και στην περίπτωση του μικτού μοντέλου κανονικών κατανομών) γίνεται ο υπολογισμός των εκ των υστέρων (posteriors) πιθανοτήτων σύμφωνα με τον παρακάτω τύπο:

$$p(k|x^i) = \frac{\pi_k As\Pi s(x^i; g_{1k}, g_{2k}, a, r_{1k}, r_{2k})}{\sum_{l=1}^K \pi_l As\Pi s(x^i; g_{1l}, g_{2l}, a, r_{1l}, r_{2l})}, k = 1 \dots K \quad (3.39)$$

3.7.3. M-βήμα (Maximization step)

Αντικειμενικός στόχος αυτού του βήματος του αλγορίθμου είναι η μεγιστοποίηση της αναμενόμενης πλήρης λογαριθμικής πιθανοφάνειας, ως προς τις παραμέτρους του μικτού μοντέλου: $\theta = \{\pi_k, g_{1k}, g_{2k}, a_k, r_{1k}, r_{2k}\}, k = 1 \dots K$

Η αναμενόμενη πλήρης λογαριθμική πιθανοφάνεια με δεδομένη την εκ των υστέρων κατανομή των παρατηρήσεων περιγράφεται από την σχέση:

$$\begin{aligned} E_Z[\log p(X, Z; \theta)] &= LLc = \sum_{i=1}^N \sum_{k=1}^K p(k|x^i) \log(\pi_k As\Pi s(x^i|k)) = \\ &= \sum_{i=1}^N \sum_{k=1}^K p(k|x^i) \log As\Pi s(x^i; g_1, g_2, a, r_1, r_2) \\ &\quad + \sum_{i=1}^N \sum_{k=1}^K p(k|x^i) \log \pi_k \end{aligned} \quad (3.40)$$

Οι εκ των υστέρων πιθανότητες $p(k|x^i)$ έχουν ήδη υπολογιστεί στο E-βήμα, και άρα η παραπάνω αντικειμενική συνάρτηση είναι ένα άθροισμα δύο ασυσχέτιστων μεταξύ τους όρων. Ο πρώτος εμπεριέχει την συνάρτηση πυκνότητας πιθανότητας της κάθε συνιστώσας asymmetric Π-sigmoid κατανομής και ο δεύτερος την αντίστοιχη εκ των προτέρων πιθανότητα. Άρα, σε κάθε επανάληψη του αλγορίθμου, για την εύρεση των παραμέτρων της κάθε συνιστώσας κατανομής, θα πρέπει να μεγιστοποιείται ο πρώτος όρος, ενώ για την εκ των προτέρων πιθανότητα της κάθε κατανομής, θα μεγιστοποιείται ο δεύτερος.

Ειδικότερα, για την εύρεση των εκ των προτέρων πιθανοτήτων $\pi = \{\pi_1, \dots, \pi_K\}, \forall k = 1 \dots K$, η βελτιστοποίηση πραγματοποιείται αναλυτικά μέσω της κλειστής μορφής:

$$\pi_k^{new} = \frac{1}{N} \sum_{i=1}^N p(k|x^i), \quad k = 1 \dots K \quad (3.41)$$

Όπου $p(k|x^i)$ είναι η αντίστοιχη εκ των υστέρων (posterior) πιθανότητα που υπολογίστηκε στο E-βήμα.

Αναφορικά με την εύρεση των παραμέτρων της κάθε συνιστώσας κατανομής, θα πρέπει να βελτιστοποιηθεί ο πρώτος όρος της σχέσης (3.40):

$$LLc1 = \sum_{i=1}^N \sum_{k=1}^K p(k|x^i) \log AsPs(x^i; g_1, g_2, a, r_1, r_2) \quad (3.42)$$

Παρόλα αυτά, σε αντίθεση με το GMM, η βελτιστοποίηση δεν μπορεί να γίνει αναλυτικά, οπότε καταφεύγουμε στην αριθμητική βελτιστοποίηση και πιο συγκεκριμένα στην μέθοδο BFGS που εφαρμόζεται και στο PsMM. Όπως και στην περίπτωση της μιας κατανομής πρέπει να προσδιορίσουμε το διάνυσμα μερικών παραγώγων της αντικειμενικής συνάρτησης καθώς και μια αρχικοποίηση των παραμέτρων. Η λογική είναι αντίστοιχη. Οι αρχικές τιμές με τις οποίες εκκινείτε ο αλγόριθμος της μεθόδου BFGS είναι οι προηγούμενες εκτιμήσεις των παραμέτρων g_1, g_2, a, r_1, r_2 που παράχθηκαν στην προηγούμενη επανάληψη του αλγορίθμου GEM, δηλαδή ισχύει:

$$\begin{aligned} BFGS_{initialvalues}: \{g_1^{initial}, g_2^{initial}, a^{initial}, r_1^{initial}, r_2^{initial}\} \\ = \{g_1^{old}, g_2^{old}, a^{old}, r_1^{old}, r_2^{old}\} \end{aligned} \quad (3.43)$$

Να σημειωθεί ότι οι παλιές εκτιμήσεις των παραμέτρων $\{g_1^{old}, g_2^{old}, a^{old}, r_1^{old}, r_2^{old}\}$ δεν αποτελούν πλέον την προσέγγιση του μεγίστου για την πλήρη πιθανοφάνεια, καθώς έχει παρέλθει το E-βήμα, στο οποίο έχουν αλλάξει οι εκ των υστέρων πιθανότητες $p(k|x^i)$. Άρα, αποτελούν μια καλή αρχική εκτίμηση η οποία θα βελτιώνεται με την αύξηση του αριθμού των επαναλήψεων του GEM αλγορίθμου, μέχρι να συγκλίνουν σε κάποιο τοπικό μέγιστο της αντικειμενικής συνάρτησης.

Στον υπολογισμό των μερικών παραγώγων πρέπει να τονιστεί ότι κάθε μια από τις παραμέτρους g_1, g_2, a, r_1, r_2 είναι πίνακας διάστασης KD , οπότε κάθε στοιχείο του πίνακα αποτελεί και μια ανεξάρτητη παράμετρο, ορίζοντας το συνολικό αριθμό παραμέτρων σε $5 * K * D$. Ο υπολογισμός αυτός, θα στηριχθεί στις σχέσεις των μερικών παραγώγων, που υπολογίστηκαν για τη μια κατανομή.

Οι σχέσεις των μερικών παραγώγων παρουσιάζονται στην συνέχεια:

Μερική παράγωγος ως προς την παράμετρο g_{1kj}

$$\begin{aligned}
 & \frac{\partial LLc1(X|g_1, g_2, a, r_1, r_2)}{\partial g_{1kj}} \\
 &= \frac{\partial}{\partial g_{1kj}} \left\{ \sum_{i=1}^N \sum_{k=1}^K p(k|x^i) \log As\Pi s(x^i; g_1, g_2, a, r_1, r_2) \right\} \\
 &= p(k|x^i) \sum_{i=1}^N \frac{1}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \frac{\partial}{\partial g_{1kj}} \left\{ \frac{1}{2} \Pi s(x^i|g_1, a, r_1) \right\} \\
 &= p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{1kj}^2} \left[\frac{e^{-g_{1kj}^2(x_j^i - a_{kj})} 2g_{1kj}(x_j^i - \alpha_{kj})}{(1 + e^{-g_{1kj}^2(x_j^i - a_{kj})})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \\
 &- p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{1kj}^2} \left[\frac{e^{-g_{1kj}^2(x_j^i - a_{kj} - r_{1kj}^2)} 2g_{1kj}(x_j^i - \alpha_{kj} - r_{1kj}^2)}{(1 + e^{-g_{1kj}^2(x_j^i - a_{kj} - r_{1kj}^2)})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)}
 \end{aligned} \tag{3.44}$$

Μερική παράγωγος ως προς την παράμετρο g_{2kj}

$$\begin{aligned}
 & \frac{\partial LL(X|g_1, g_2, a, r_1, r_2)}{\partial g_{2kj}} \\
 &= \frac{\partial}{\partial g_{2kj}} \left\{ \sum_{i=1}^N \sum_{k=1}^K p(k|x^i) \log As\Pi s(x^i; g_1, g_2, a, r_1, r_2) \right\} \\
 &= p(k|x^i) \sum_{i=1}^N \frac{1}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \frac{\partial}{\partial g_{2kj}} \left\{ \frac{1}{2} \Pi s(x^i|g_2, a, r_1, r_2) \right\} \\
 &= p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{2kj}^2} \left[\frac{e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2)} 2g_{2kj}(x_j^i - \alpha_{kj} - r_{1kj}^2)}{(1 + e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2)})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \\
 &- p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{2kj}^2} \left[\frac{e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2 - r_{2kj}^2)} 2g_{2kj}(x_j^i - \alpha_{kj} - r_{1kj}^2 - r_{2kj}^2)}{(1 + e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2 - r_{2kj}^2)})^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)}
 \end{aligned} \tag{3.45}$$

Μερική παράγωγος ως προς την παράμετρο a_{kj}

$$\begin{aligned}
& \frac{\partial LL(X|g_1, g_2, a, r_1, r_2)}{\partial a_j} \\
&= \frac{\partial}{\partial a_{kj}} \left\{ \sum_{i=1}^N \sum_{k=1}^K p(k|x^i) \log As\mathcal{P}s(x^i; g_1, g_2, a, r_1, r_2) \right\} \\
&= p(k|x^i) \sum_{i=1}^N \frac{1}{As\mathcal{P}s(x^i; g_1, g_2, a, r_1, r_2)} \frac{\partial}{\partial a_{kj}} \{As\mathcal{P}s(x^i; g_1, g_2, a, r_1, r_2)\} \\
&= p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{1kj}^2} \left[\frac{-e^{-g_{1kj}^2(x_j^i - a_{kj})} g_{1kj}^2}{(1 + e^{-g_{1kj}^2(x_j^i - a_{kj})})^2} \right]}{As\mathcal{P}s(x^i; g_1, g_2, a, r_1, r_2)} \\
&\quad + p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{1kj}^2} \left[\frac{e^{-g_{1kj}^2(x_j^i - a_{kj} - r_{1kj}^2)} g_{1kj}^2}{(1 + e^{-g_{1kj}^2(x_j^i - a_{kj} - r_{1kj}^2)})^2} \right]}{As\mathcal{P}s(x^i; g_1, g_2, a, r_1, r_2)} \\
&\quad + p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{2kj}^2} \left[\frac{e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2)} g_{2kj}^2}{(1 + e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2)})^2} \right]}{As\mathcal{P}s(x^i; g_1, g_2, a, r_1, r_2)} \\
&\quad + p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{2kj}^2} \left[\frac{e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2 - r_{2kj}^2)} g_{2kj}^2}{(1 + e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2 - r_{2kj}^2)})^2} \right]}{As\mathcal{P}s(x^i; g_1, g_2, a, r_1, r_2)}
\end{aligned} \tag{3.46}$$

Μερική παράγωγος ως προς την παράμετρο r_{1kj}

$$\begin{aligned}
& \frac{\partial LL(X|g_1, g_2, a, r_1, r_2)}{\partial r_{1j}} \\
&= \frac{\partial}{\partial r_{1kj}} \left\{ \sum_{i=1}^N \sum_{k=1}^K p(k|x^i) \log As\Pi s(x^i; g_1, g_2, a, r_1, r_2) \right\} \\
&= p(k|x^i) \sum_{i=1}^N \frac{1}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \frac{\partial}{\partial r_{1kj}} \{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)\} \\
&= p(k|x^i) \sum_{i=1}^N \frac{-\frac{1}{r_{1kj}} \Pi s(x^i; g_1, a, r_1)}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \\
&\quad + p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{1kj}^2} \left[\frac{e^{-g_{1kj}^2(x_j^i - a_{kj} - r_{1kj}^2)} 2r_{1kj} g_{1kj}^2}{\left(1 + e^{-g_{1kj}^2(x_j^i - a_{kj} - r_{1kj}^2)}\right)^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \\
&\quad + p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{2kj}^2} \left[\frac{-e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2)} 2r_{1kj} g_{2kj}^2}{\left(1 + e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2)}\right)^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \\
&\quad + p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{2kj}^2} \left[\frac{e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2 - r_{2kj}^2)} 2r_{1kj} g_{2kj}^2}{\left(1 + e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2 - r_{2kj}^2)}\right)^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)}
\end{aligned} \tag{3.47}$$

Μερική παράγωγος ως προς την παράμετρο r_{2kj}

$$\begin{aligned}
& \frac{\partial LL(X|g_1, g_2, a, r_1, r_2)}{\partial r_{2kj}} \\
&= \frac{\partial}{\partial r_{2kj}} \left\{ \sum_{i=1}^N \sum_{k=1}^K p(k|x^i) \log As\Pi s(x^i; g_1, g_2, a, r_1, r_2) \right\} \\
&= p(k|x^i) \sum_{i=1}^N \frac{1}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \frac{\partial}{\partial r_{2kj}} \left\{ \frac{1}{2} \Pi s(x^i|g_2, a, r_1, r_2) \right\} \\
&= p(k|x^i) \sum_{i=1}^N \frac{-\frac{1}{r_{2kj}} \Pi s(x^i; g_2, a, r_1, r_2)}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)} \\
&\quad + p(k|x^i) \sum_{i=1}^N \frac{\frac{1}{2r_{2kj}^2} \left[\frac{e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2 - r_{2kj}^2)} 2r_{2kj} g_{2kj}^2}{\left(1 + e^{-g_{2kj}^2(x_j^i - a_{kj} - r_{1kj}^2 - r_{2kj}^2)}\right)^2} \right]}{As\Pi s(x^i; g_1, g_2, a, r_1, r_2)}
\end{aligned} \tag{3.48}$$

Κάθε μια από τις παραπάνω σχέσεις ορίζει έναν πίνακα διάστασης KD . Οι συνιστώσες του πίνακα ορίζονται από τις μερικές παραγώγους της λογαριθμικής πιθανοφάνειας κάθε συνιστώσας κατανομής για τις παραμέτρους g_1, g_2, a, r_1, r_2 , σε κάθε διάσταση $j = 1 \dots D$, πολλαπλασιασμένες κάθε φορά με τις αντίστοιχες εκ των υστέρων πιθανότητες $p(k|x^i)$.

Αναφορικά με το διάνυσμα των μερικών παραγώγων της αντικειμενικής συνάρτησης ανά διάσταση, η κάθε συνιστώσα υπολογίζεται ως άθροισμα των γραμμών του πίνακα διάστασης KD

Συμπερασματικά, ως αντικειμενική συνάρτηση f στην περίπτωση μας ορίζεται η συνάρτηση (3.40) και εξαρτάται από το διάνυσμα παραμέτρων $\theta = [g_1, g_2, a, r_1, r_2]$. Επειδή υποθέτουμε ανεξαρτησία της asymmetric Π-sigmoid κατανομής ανά διάσταση, θα εκτελέσουμε συνολικά D -διαδικασίες βελτιστοποίησης και θα προκύψουν D -πίνακες εκτιμήσεων των παραμέτρων ανά διάσταση για τις συνιστώσες κατανομής του μικτού μοντέλου.

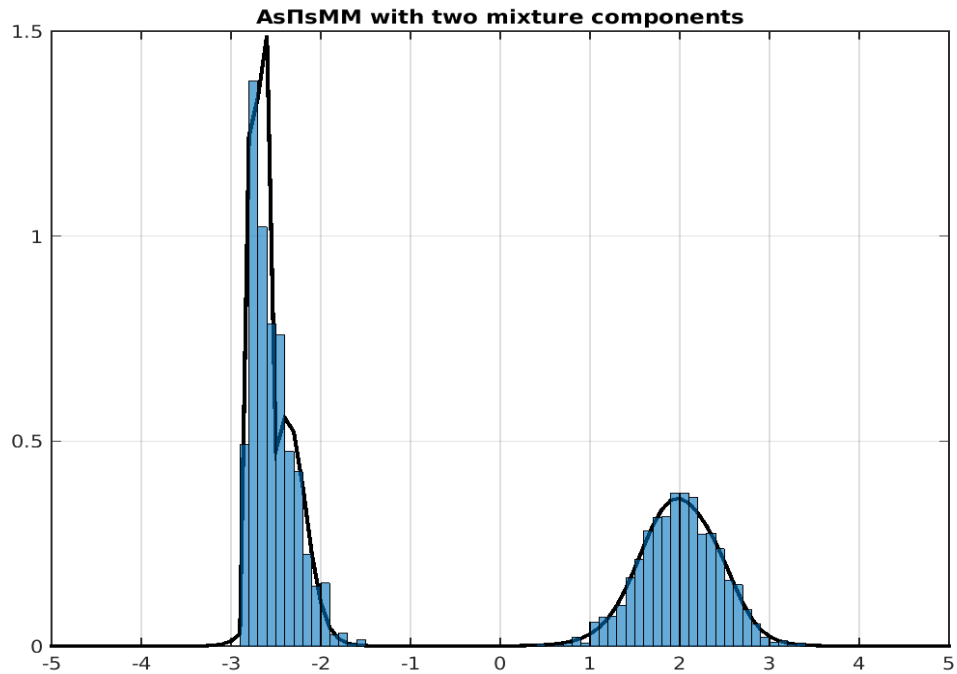
Ο πίνακας μερικών παραγώγων ανά διάσταση ισούται με:

$$\nabla f = \left[\frac{\partial LL(X|\theta)}{\partial g_{1j}}, \frac{\partial LL(X|\theta)}{\partial g_{2j}}, \frac{\partial LL(X|\theta)}{\partial a_j}, \frac{\partial LL(X|\theta)}{\partial r_{1j}}, \frac{\partial LL(X|\theta)}{\partial r_{2j}} \right]_{j=1 \dots D} \quad (3.49)$$

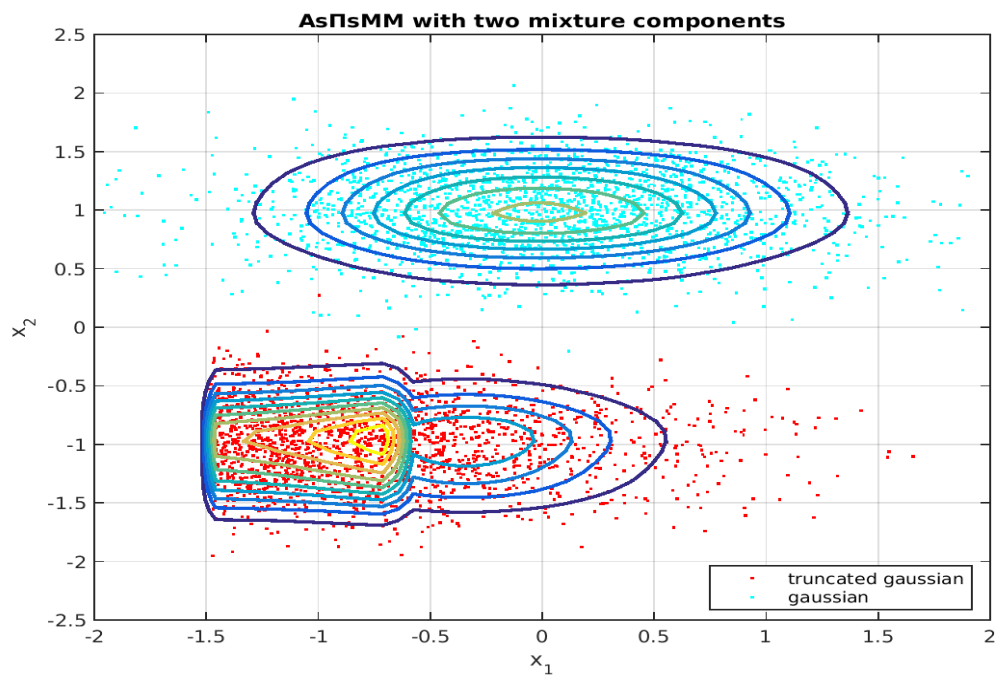
ενώ κάθε στοιχείο του παραπάνω πίνακα, προκύπτει αναλυτικά από τις σχέσεις (3.44)-(3.48).

Τέλος, η αρχικοποίηση των παραμέτρων της αντικειμενικής συνάρτησης f , ορίστηκε από την σχέση (3.43) οπότε ο αλγόριθμος αριθμητικής βελτιστοποίησης BFGS μπορεί πλέον να εκτελεστεί.

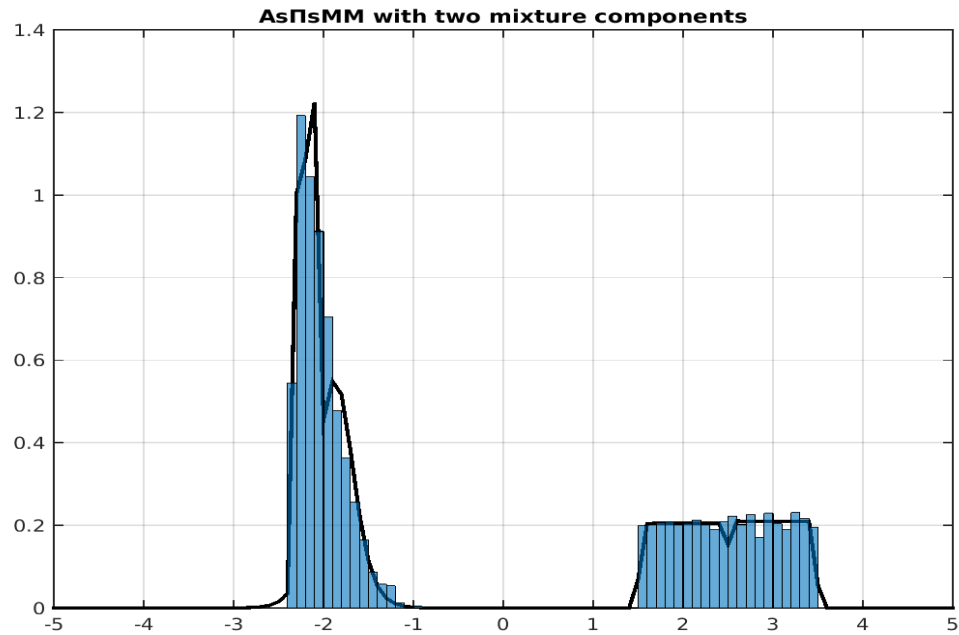
Στα Σχήματα 3.18 – 3.25 παρουσιάζονται τα αποτελέσματα της εφαρμογής του GEM για μοντέλα μικτών asymmetric Π-sigmoid κατανομών. Η μέθοδος εφαρμόζεται σε σύνθετα σύνολα παρατηρήσεων που σχηματίζουν ομοιογενείς ομάδες κανονικών, ομοιόμορφων και truncated κανονικών παρατηρήσεων αντίστοιχα. Πειραματικά αποτελέσματα παρουσιάζονται στο επόμενο κεφάλαιο.



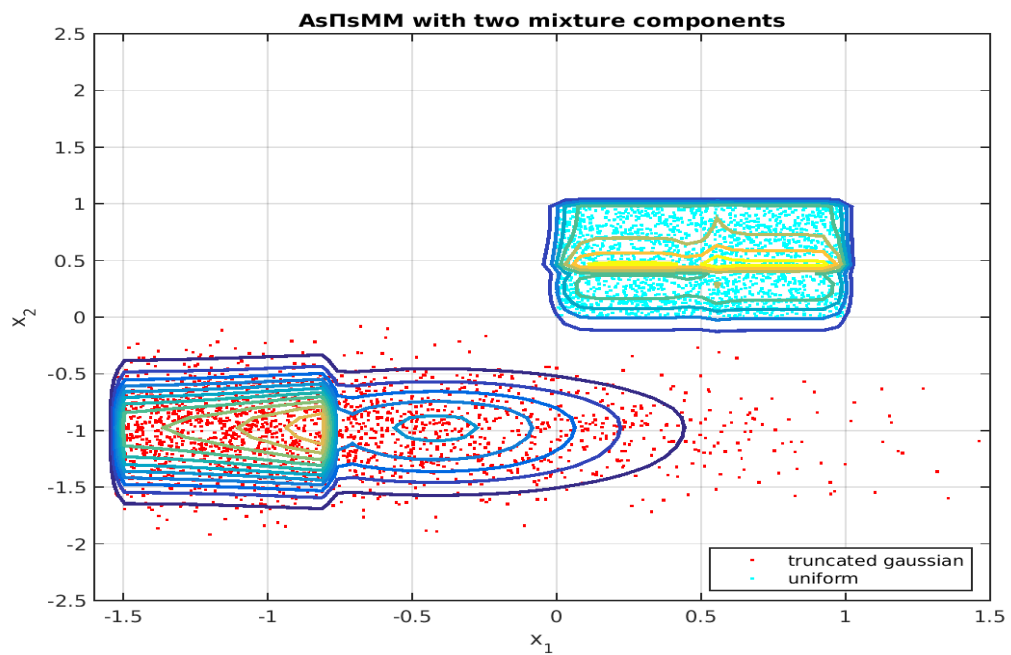
Σχήμα 3.18 Αποτέλεσμα της εκπαίδευσης ενός AsPisMM σε μονοδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων.



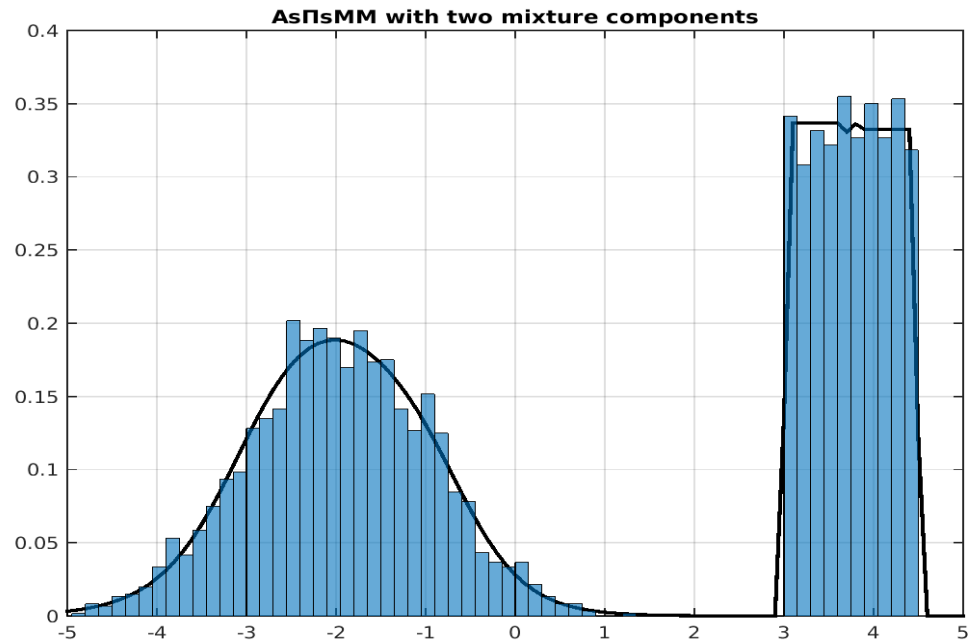
Σχήμα 3.19 Αποτέλεσμα της εκπαίδευσης ενός AsPisMM σε διδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων.



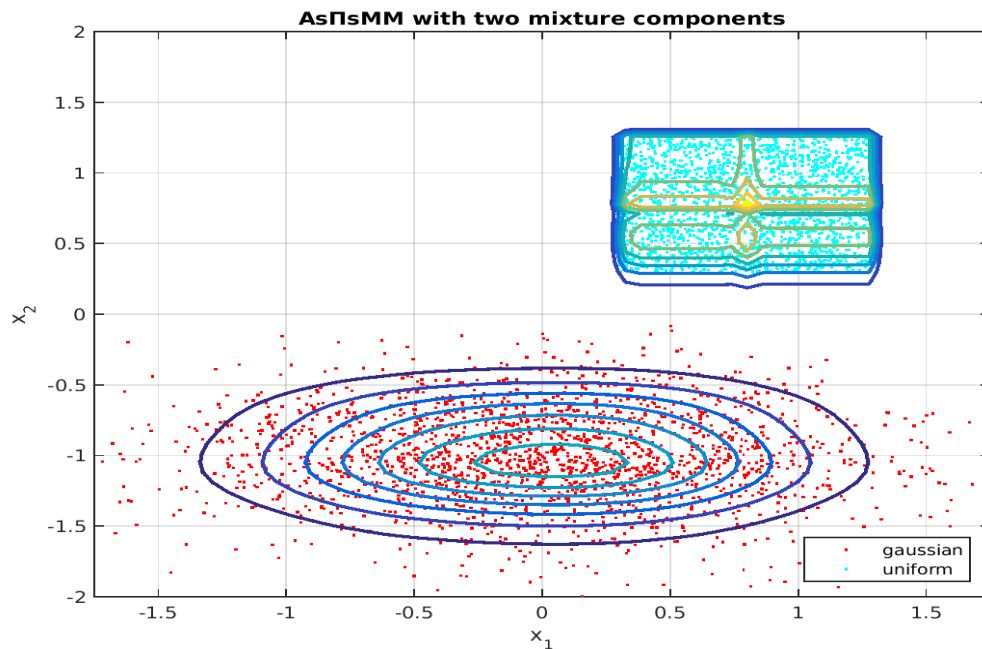
Σχήμα 3.20 Αποτέλεσμα της εκπαίδευσης ενός AsPIsMM σε μονοδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων και μια ομάδα ομοιόμορφων παρατηρήσεων.



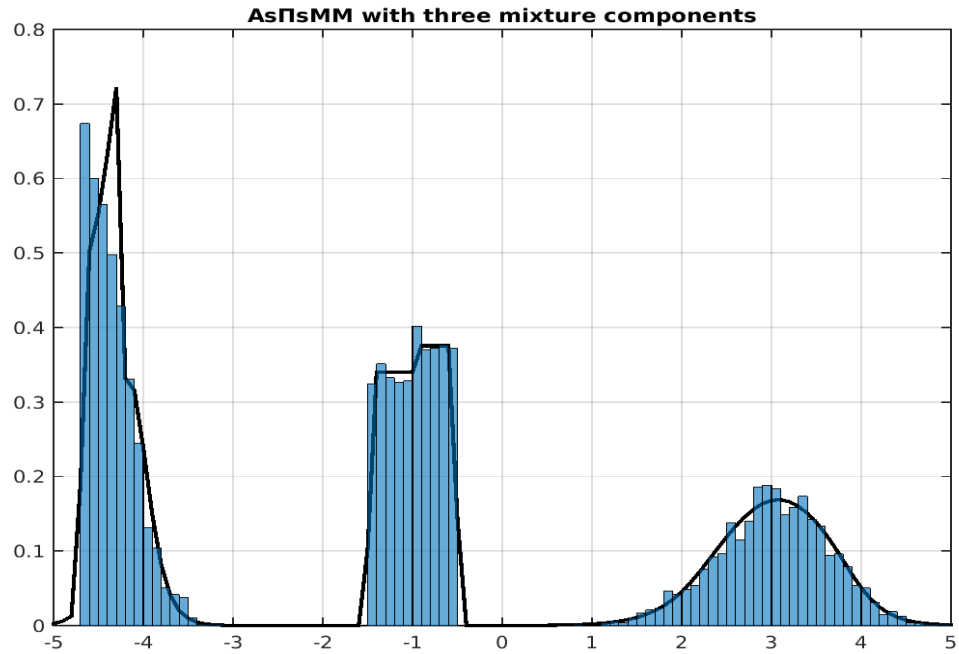
Σχήμα 3.21 Αποτέλεσμα της εκπαίδευσης ενός AsPIsMM σε διδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων.



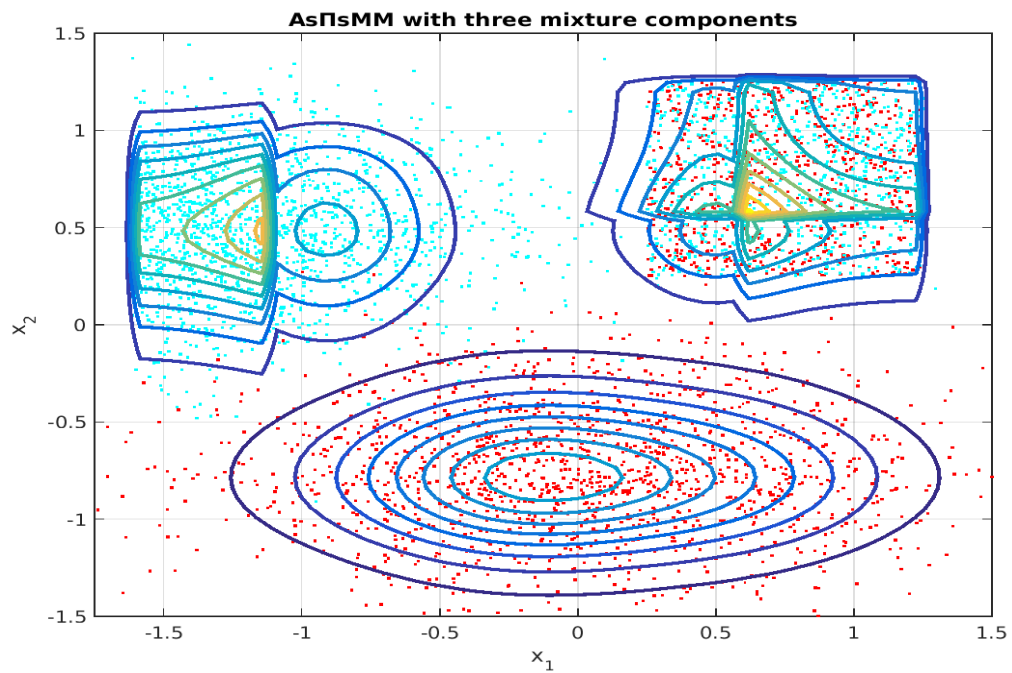
Σχήμα 3.22 Αποτέλεσμα της εκπαίδευσης ενός AsPIsMM σε μονοδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα κανονικών παρατηρήσεων και μια ομάδα ομοιόμορφων παρατηρήσεων.



Σχήμα 3.23 Αποτέλεσμα της εκπαίδευσης ενός AsPIsMM σε διδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων.



Σχήμα 3.24 Αποτέλεσμα της εκπαίδευσης ενός AsPIsMM σε μονοδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων, μια ομάδα ομοιόμορφων παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων.



Σχήμα 3.25 Αποτέλεσμα της εκπαίδευσης ενός AsPIsMM σε διδιάστατο σύνολο παρατηρήσεων που αποτελείται από μια ομάδα truncated κανονικών παρατηρήσεων, μια ομάδα ομοιόμορφων παρατηρήσεων και μια ομάδα κανονικών παρατηρήσεων.

3.8. Αντιμετώπιση θορύβου

Σε πολλά σύνολα παρατηρήσεων παρατηρείται συχνά η ύπαρξη ακραίων τιμών [7], καθώς επίσης και θορύβου που επηρεάζει άμεσα την απόδοση των αλγορίθμων ομαδοποίησης, προκαλώντας σημαντικές αλλοιώσεις στην ποιότητα των παραγόμενων λύσεων.

Για την αντιμετώπιση αυτών των προβλημάτων χρησιμοποιήθηκε μια ειδική τεχνική και ενσωματώθηκε στο μοντέλο των μικτών asymmetric Π-sigmoid κατανομών.

Πιο αναλυτικά, έστω ένα σύνολο παρατηρήσεων:

$$X = \{x_j^i\}, \quad i = 1 \dots N, \quad j = 1 \dots D$$

Για την περιγραφή των ακραίων τιμών, εισάγαμε μια επιπλέον συνιστώσα κατανομή, η οποία ονομάζεται background κατανομή και περιγράφεται από τον τύπο:

$$k_{BG} = \frac{1}{\max(X) - \min(X)} \quad (3.50)$$

Η background κατανομή είναι ομοιόμορφη και ορίζεται σε όλο το πεδίο τιμών (domain) του συνόλου των παρατηρήσεων. Σκοπός της είναι η απομόνωση και η περιγραφή όλων των ακραίων τιμών. Με αυτόν τον τρόπο, οι υπόλοιπες παρατηρήσεις θα περιγραφούν αποδοτικά από το AsPisMM με την διαδικασία που περιγράψαμε στην προηγούμενη ενότητα.

Η αρχικοποίηση της εκ των προτέρων πιθανότητας για την background κατανομή γίνεται σε μια μικρή τιμή (στην υλοποίηση ανατέθηκε $\pi_k = 0.01$) και δεν συμμετέχει στην διαδικασία βελτιστοποίησης που πραγματοποιείται στο M-βήμα του αλγορίθμου GEM. Παρόλα αυτά συμμετέχει κανονικά στο E-βήμα και υπολογίζονται οι εκ των υστέρων (posteriors) πιθανότητες $p(x|k_{BG})$. Σύμφωνα με τις τιμές αυτές μπορούμε εύκολα να εντοπίσουμε τις ακραίες παρατηρήσεις του συνόλου X , οι οποίες θα περιγραφούν τελικά από την background κατανομή.

Συμπερασματικά, πρέπει να αναφερθεί ότι η συνεισφορά της background κατανομής ως συνιστώσα του μικτού μοντέλου είναι πολύ μικρή καθώς η εκ των προτέρων πιθανότητά της π_k , αρχικοποιείται σε μια πολύ μικρή τιμή, οπότε η posterior πιθανότητες της συντριπτικής πλειοψηφίας των παρατηρήσεων είναι αμελητέες. Επιπρόσθετα, η background κατανομή δεν θα προσμετρηθεί στο συνολικό πλήθος K των συνιστωσών κατανομών που απαρτίζουν το AsPsMM, καθώς ο ρόλος της είναι υποστηρικτικός. Συνεπώς, η εισαγωγή της κατανομής αυτής, μπορεί να επηρεάσει μόνο θετικά το τελικό αποτέλεσμα του μοντέλου AsPsMM.

Αυτό που απομένει για την ολοκλήρωση της διαδικασίας εκπαίδευσης ενός AsPsMM, είναι ένας αποδοτικός τρόπος υπολογισμού του πλήθους των συνιστωσών asymmetric Π-sigmoid κατανομών και θα μας απασχολήσει στην επόμενη ενότητα.

3.9. Εκτίμηση του πλήθους των συνιστωσών κατανομών σε ένα μοντέλο μικτών asymmetric Π-sigmoid κατανομών.

Για την εκκίνηση της διαδικασίας εκπαίδευσης ενός AsPsMM πρέπει να προσδιοριστεί το πλήθος των συνιστωσών κατανομών του μοντέλου. Η τιμή αυτή, όπως αναλύθηκε και στα προηγούμενα κεφάλαια, αποτελεί μια υπερπαράμετρο του μοντέλου μικτών κατανομών. Στην προσπάθεια προσδιορισμού της, θα χρησιμοποιήσουμε το στατιστικό κριτήριο dip και θα δημιουργήσουμε έναν νέο αλγόριθμο εκπαίδευσης.

Το θεωρητικό σκέλος του κριτηρίου dip παρουσιάστηκε στο πρώτο κεφάλαιο. Η εφαρμογή του γίνεται στην εμπειρική συνάρτηση κατανομής που αντιστοιχεί σε ένα σύνολο μονοδιάστατων παρατηρήσεων και εξάγεται μια μη αρνητική τιμή.

Η τιμή του κριτηρίου dip , ορίζει εάν η εμπειρική συνάρτηση κατανομής έχει μια ή περισσότερες κορυφές. Μέσω της τιμής αυτής, ορίζεται και η p -value για τον στατιστικό έλεγχο που θα πραγματοποιηθεί.

Μια μικρή τιμή του dip υποδηλώνει μονοκόρυφη εμπειρική κατανομή. Αντίθετα, μια μικρή τιμή της p -value ($p - \text{value} < \alpha$), οδηγεί στην απόρριψη της υπόθεσης H_0 : “Η F_n είναι μονοκόρυφη”, με επίπεδο στατιστικής σημαντικότητας α .

Για την νέα διαδικασία εκπαίδευσης ενός PsMM μέσω του κριτηρίου dip , θέσαμε επίπεδο στατιστικής σημαντικότητας $\alpha = 0$.

Για την εφαρμογή του κριτηρίου, υποθέτουμε ότι μια μονοκόρυφη εμπειρική κατανομή περιγράφει ένα σύνολο παρατηρήσεων που δημιουργήθηκε από μια asymmetric Π-sigmoid κατανομή, ενώ αντίθετα, μια εμπειρική κατανομή με πολλαπλές κορυφές περιγράφει ένα σύνολο παρατηρήσεων που δημιουργείται από περισσότερες asymmetric Π-sigmoid κατανομές.

Οι ισοδύναμες υποθέσεις για την επίλυση του προβλήματος ομαδοποίησης μέσω του μοντέλου μικτών κατανομών, ορίζουν ότι μια μονοκόρυφη εμπειρική κατανομή περιγράφει μια ομοιογενή ομάδα παρατηρήσεων, ενώ αντίθετα μια κατανομή με πολλαπλές κορυφές περιγράφει ένα σύνθετο σύνολο παρατηρήσεων το οποίο περιλαμβάνει πολλές ομάδες.

Στην συνέχεια θα παρουσιαστεί η διαδικασία εκπαίδευσης ενός μοντέλου μικτών asymmetric Π-sigmoid κατανομών, με έμφαση στον υπολογισμό του πλήθους των συνιστωσών του κατανομών.

Έστω ένα σύνολο από N παρατηρήσεις, διάστασης D . Κάθε παρατήρηση έχει την μορφή:

$$x^i = \{x_1^i, x_2^i, \dots, x_D^i\}, \quad i = 1 \dots N$$

Συμβολίζουμε με Π^j το σύνολο μονοδιάστατων παρατηρήσεων, που προκύπτει από την προβολή ενός υπό μελέτη συνόλου, στον j -άξονα των παρατηρήσεων $j = 1 \dots D$.

Σύμφωνα με τα παραπάνω, όταν μια τυχαία παρατήρηση $x^i = \{x_1^i, x_2^i, \dots, x_D^i\}$, $i = 1 \dots N$, προβάλλεται στον j -άξονα, $j = 1 \dots D$, ισχύει $x_j^i \in \Pi^j$.

Ακόμα, συμβολίζουμε με dip_pvalue_j την p -value που προκύπτει από την εφαρμογή του κριτηρίου dip στα στοιχεία του συνόλου Π^j .

1. Αρχικά υποθέτουμε ότι το σύνολο των παρατηρήσεων X , παράγεται εξ ολοκλήρου από μια κατανομή οπότε αρχικοποιούμε το πλήθος συνιστωσών K στη τιμή ένα ($K=1$).

2. Στην συνέχεια, και ανεξάρτητα για την κάθε διάσταση των παρατηρήσεων του συνόλου X , υπολογίζουμε την τιμή dip_pvalue_j των συνόλων:

$$P^j = X^j = \{x_j^1, x_j^2 \dots x_j^N\}, j = 1 \dots D$$

(ο υπολογισμός των dip_pvalue_j γίνεται συνολικά D φορές).

Για κάθε σύνολο P^j συγκρίνουμε την αντίστοιχη τιμή της dip_pvalue_j με το μηδέν.

- Εάν $\exists j \in 1 \dots D : dip_pvalue_j = 0$, τότε η προβολή του X στην j διάσταση, περιγράφεται από μια εμπειρική συνάρτηση κατανομής με πολλαπλές κορυφές, οπότε συμπεραίνουμε ότι το υπό μελέτη σύνολο παρατηρήσεων X , παράγεται από περισσότερες (τουλάχιστον δύο) συνιστώσες κατανομές. Για τον σκοπό αυτό, το πλήθος των συνιστωσών κατανομών αυξάνεται κατά ένα και ο αλγόριθμος της εκπαίδευσης αρχικοποιείται στην τιμή $K=K+1$.
- Εάν $dip_pvalue_j > 0 \forall j \in 1 \dots D$, τότε η προβολή του X σε κάθε διάσταση περιγράφεται από μια μονοκόρυφη εμπειρική κατανομή, οπότε συμπεραίνουμε ότι το σύνολο των παρατηρήσεων X παράγεται εξ ολοκλήρου από μία κατανομή. Το τελικό συνολικό πλήθος συνιστωσών κατανομών του μοντέλου είναι ένα ($K=1$), δηλαδή το μικτό μοντέλο εκφυλίζεται στην asymmetric Π- sigmoid κατανομή.

Εάν $K \neq 1$, η διαδικασία εκπαίδευσης συνεχίζεται με τα παρακάτω βήματα:

3. Εκπαιδεύουμε ένα μικτό μοντέλο με K asymmetric Π-sigmoid κατανομές μέσω του GEM αλγορίθμου και προσδιορίζουμε τις εκτιμήσεις:

$$\pi_k, g_{1k}, g_{2k}, a_k, r_{1k}, r_{2k} \quad k = 1 \dots K.$$

4. Κάνουμε ομαδοποίηση των πολυδιάστατων παρατηρήσεων, μέσω των εκ των υστέρων πιθανοτήτων που υπολογίστηκαν στο τελευταίο Ε-βήμα του αλγορίθμου. Υπολογίζουμε K -ομάδες $C_k, k = 1 \dots K$

5. Για κάθε μια από τις ομάδες $C_k, k = 1 \dots K$ πολυδιάστατων παρατηρήσεων που προσδιορίστηκαν στο προηγούμενο βήμα, υπολογίζουμε ανεξάρτητα σε κάθε διάσταση $j \in 1 \dots D$, την τιμή dip_pvalue_j των συνόλων:

$$\Pi^j = C_k^j$$

(ο υπολογισμός των dip_pvalue_j γίνεται συνολικά $D * K$ φορές)

Για κάθε σύνολο Π^j συγκρίνουμε την τιμή της dip_pvalue_j με το μηδέν:

- Εάν $\exists j \in 1 \dots D : dip_pvalue_j = 0$, τότε η προβολή της ομάδας C_k στην j διάσταση, περιγράφεται από μια εμπειρική συνάρτηση κατανομής με πολλαπλές κορυφές, οπότε συμπεραίνουμε ότι η υπό μελέτη ομάδα παρατηρήσεων C_k , παράγεται από περισσότερες (τουλάχιστον δύο) συνιστώσες κατανομές. Για τον σκοπό αυτό το πλήθος των συνολικών συνιστωσών κατανομών του μοντέλου, αυξάνεται κατά ένα και ο αλγόριθμος της εκπαίδευσης επανεκκινείται από το βήμα τρία με νέο πλήθος συνιστωσών: $K = K + 1$.
- Εάν $dip_pvalue_j > 0 \forall j \in 1 \dots D, \forall k = 1 \dots K$, τότε η προβολή κάθε ομάδας C_k σε κάθε διάσταση περιγράφεται από μια μονοκόρυφη εμπειρική κατανομή, οπότε συμπεραίνουμε ότι κάθε ομάδα παρατηρήσεων C_k παράγεται εξ ολοκλήρου από μία κατανομή. Το συνολικό πλήθος συνιστωσών κατανομών του μοντέλου, καθώς και οι τελικές εκτιμήσεις των παραμέτρων είναι οι τρέχουσες.

Τα παραπάνω βήματα, οδηγούν στην ολοκλήρωση της νέας διαδικασίας εκπαίδευσης ενός πολυδιάστατου μικτού μοντέλου asymmetric Π-sigmoid κατανομών με ταυτόχρονη εκτίμηση του συνόλου των παραμέτρων και του πλήθους των συνιστωσών asymmetric Π-sigmoid κατανομών.

Η πλήρης εκπαίδευση ενός AsΠsMM θα εφαρμοστεί στο επόμενο κεφάλαιο σε πειραματικά σύνολα παρατηρήσεων με στόχο αφενός την εκτίμηση του πλήθους των συνιστωσών κατανομών του μοντέλου και αφετέρου την εκτίμηση των παραμέτρων για κάθε συνιστώσα asymmetric Π-sigmoid κατανομή.

ΚΕΦΑΛΑΙΟ 4. ΠΕΙΡΑΜΑΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ

- 4.1. Εισαγωγή
 - 4.2. Τεχνητά σύνολα παρατηρήσεων
 - 4.3. Πειραματική αξιολόγηση τεχνητών συνόλων παρατηρήσεων
 - 4.3.1. Ομοιογενή σύνολα μονοδιάστατων παρατηρήσεων
 - 4.3.2. Σύνθετα σύνολα μονοδιάστατων παρατηρήσεων
 - 4.3.3. Ομοιογενή σύνολα πολυδιάστατων παρατηρήσεων
 - 4.3.4. Σύνθετα σύνολα πολυδιάστατων παρατηρήσεων
 - 4.4. Συμπεράσματα
 - 4.5. Πειραματική αξιολόγηση πραγματικών συνόλων παρατηρήσεων
-

4.1. Εισαγωγή

Σε αυτό το κεφάλαιο θα παρουσιαστούν πειραματικά αποτελέσματα από την μοντελοποίηση τεχνητών και πραγματικών συνόλων παρατηρήσεων μέσω της asymmetric Π-sigmoid κατανομής. Για λόγους γενικότητας, το στατιστικό μοντέλο που αποτελείται από μια κατανομή, θεωρείται υποπερίπτωση ενός μικτού μοντέλου με μια συνιστώσα. Στόχος των πειραμάτων, είναι η σύγκριση της απόδοσης ενός AsΠsMM, έναντι των ΠsMM και GMM. Η πειραματική διαδικασία που θα παρουσιαστεί χωρίζεται σε δύο διακριτές ενότητες. Στην πρώτη ενότητα, περιλαμβάνονται τεχνητά σύνολα παρατηρήσεων με σκοπό την μοντελοποίηση τους από τα τρία μικτά μοντέλα. Πιο αναλυτικά, κάθε σύνολο παρατηρήσεων διαμερίζεται με τυχαίο τρόπο σε δύο ξένα υποσύνολα που περιλαμβάνουν ίσο πλήθος παρατηρήσεων: το σύνολο εκπαίδευσης (training set) και το σύνολο ελέγχου (test set). Μέσω του συνόλου εκπαίδευσης γίνεται εκτίμηση του συνόλου των παραμέτρων

του μοντέλου, ενώ στη συνέχεια, μέσω του συνόλου ελέγχου υπολογίζεται η λογαριθμική πιθανοφάνεια του μοντέλου, με δεδομένες τις εκτιμήσεις των παραμέτρων που υπολογίστηκαν μέσω της διαδικασίας της εκπαίδευσης. Ως μέτρο σύγκρισης των αποτελεσμάτων μοντελοποίησης χρησιμοποιείται η τιμή της αρνητικής λογαριθμικής πιθανοφάνειας. Το μοντέλο με την μικρότερη τιμή στην αρνητική λογαριθμική πιθανοφάνεια του συνόλου ελέγχου μοντελοποιεί με τον βέλτιστο τρόπο το σύνολο των παρατηρήσεων. Τέλος, γίνεται εφαρμογή του κριτηρίου dip , με σκοπό την εκτίμηση του πλήθους των συνιστωσών κατανομών για τα μοντέλα AsPsMM και PsMM. Στην δεύτερη ενότητα, περιλαμβάνονται πραγματικά σύνολα παρατηρήσεων από εικόνες, με σκοπό αφενός την μοντελοποίηση τους (σύγκριση των τιμών της αρνητικής λογαριθμικής πιθανοφάνειας των μικτών μοντέλων και ανάδειξη του βέλτιστου μοντέλου) και αφετέρου, την κατάτμηση των εικόνων μέσω της ομαδοποίησης. Στην κατάτμηση των εικόνων, το πλήθος των ομάδων καθορίζεται εκ των προτέρων.

4.2. Τεχνητά σύνολα παρατηρήσεων

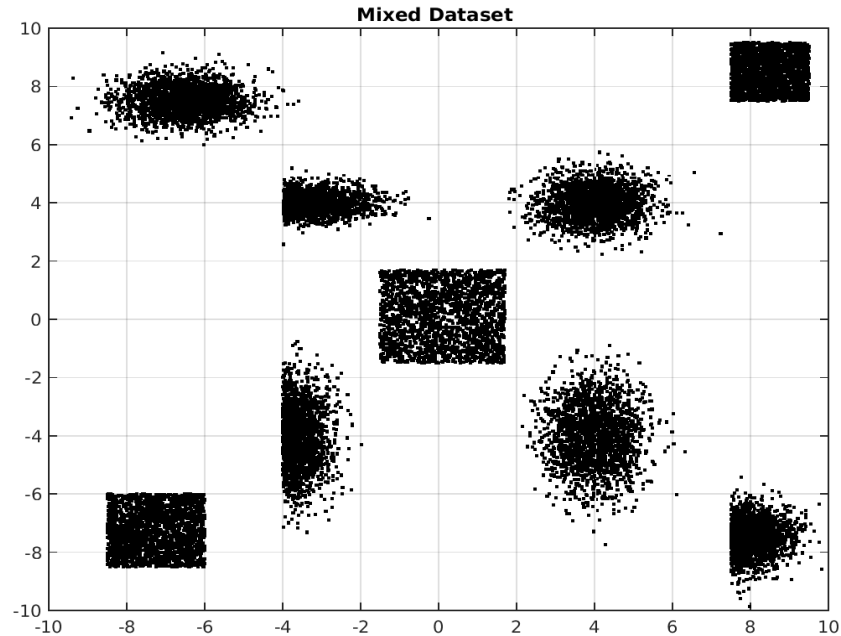
Για την εκτέλεση ενός πειράματος σε τεχνητά σύνολα παρατηρήσεων, τα AsPsMM, PsMM εκπαιδεύτηκαν μέσω του αλγορίθμου GEM ενώ το GMM μέσω του αλγορίθμου EM. Η διαδικασία της εκπαίδευσης των AsPsMM, PsMM, ενισχύθηκε με την εφαρμογή του κριτηρίου dip , όπως περιγράφηκε στα προηγούμενα κεφάλαια. Αποτέλεσμα της ενισχυμένης εκπαίδευσης είναι η ταυτόχρονη εκτίμηση του συνόλου των παραμέτρων κάθε μοντέλου και η εκτίμηση του πλήθους των συνιστωσών του κατανομών. Αντίθετα, η εκπαίδευση ενός GMM γίνεται με δεδομένη την πραγματική τιμή του πλήθους των συνιστωσών κατανομών.

Εφόσον το ζητούμενο για την αποδοτική μοντελοποίηση ενός συνόλου παρατηρήσεων είναι η μεγιστοποίηση της συνάρτησης της πιθανοφάνειας, μια ισοδύναμη έκφραση ορίζει την ελαχιστοποίηση της συνάρτησης της αρνητικής λογαριθμικής πιθανοφάνειας. Στην πειραματική διαδικασία, ως κριτήριο σύγκρισης της απόδοσης των τριών μοντέλων, χρησιμοποιήθηκε η τιμή της αρνητικής λογαριθμικής πιθανοφάνειας στο σύνολο ελέγχου.

Συνδέοντας το πρόβλημα της εκτίμησης συνάρτησης πυκνότητας πιθανότητας με το πρόβλημα της ομαδοποίησης, υποθέτουμε ότι οι παρατηρήσεις που παράγονται από την ίδια κατανομή σχηματίζουν μια ενιαία ομάδα. Υπό αυτήν την παραδοχή, τα τεχνητά σύνολα διακρίνονται σε ομοιογενή, περιλαμβάνοντας παρατηρήσεις που σχηματίζουν μια ομάδα, αλλά και σύνθετα, όπου ο αριθμός των ομάδων είναι μεγαλύτερος. Στα πειράματα που θα αναλυθούν στην συνέχεια, τα τεχνητά σύνολα δεδομένων περιλαμβάνουν μια, τρεις, έξι ή εννέα ανεξάρτητες ομάδες παρατηρήσεων ($K=1,3,6,9$). Εφόσον το πλήθος των ομάδων στα τεχνητά σύνολα κυμαίνεται, μέσω της πειραματικής διαδικασίας θα διαπιστωθεί η ακρίβεια της εκτίμησης του πλήθους των συνιστωσών κατανομών των AsPsMM και PsMM. Τονίζεται ότι η συγκεκριμένη εκτίμηση είναι πρωτεύουσας σημασίας για την επιτυχή εκπαίδευση των μοντέλων μικτών κατανομών και επηρεάζει άμεσα την τιμή της αρνητικής λογαριθμικής πιθανοφάνειας. Τέλος, μια ακόμη σημαντική παράμετρος που κυμαίνεται είναι η διάσταση των παρατηρήσεων. Τα τεχνητά σύνολα διακρίνονται σε μονοδιάστατα και πολυδιάστατα ($D=1,2,4,6$). Η εκπαίδευση των μικτών μοντέλων στα πολυδιάστατα σύνολα παρατηρήσεων έγινε με την υπόθεση τις ανεξαρτησίας των παρατηρήσεων ανά διάσταση. Παρακάτω παρουσιάζονται οι κατηγορίες των τεχνητών συνόλων παρατηρήσεων που χρησιμοποιήθηκαν στην πειραματική διαδικασία.

- Gaussian: Στην κατηγορία αυτή περιλαμβάνονται παρατηρήσεις που παράχθηκαν από την κανονική κατανομή και σχηματίζουν μια ομοιογενή ομάδα παρατηρήσεων.
- Uniform: Στην κατηγορία αυτή περιλαμβάνονται παρατηρήσεις που παράχθηκαν από την ομοιόμορφη κατανομή και σχηματίζουν μια ομοιογενή ομάδα παρατηρήσεων.
- Truncated Gaussian: Στην κατηγορία αυτή περιλαμβάνονται παρατηρήσεις που παράχθηκαν από τμήματα της κανονικής κατανομής και σχηματίζουν μια ομοιογενή ομάδα παρατηρήσεων.
- Mixed: Η τελευταία αυτή σύνθετη κατηγορία περιλαμβάνει ένα μίγμα παρατηρήσεων που παράγονται από τις παραπάνω κατανομές και σχηματίζουν μη επικαλυπτόμενες ομάδες από gaussian, uniform και truncated gaussian παρατηρήσεις. Για την δίκαιη σύγκριση της απόδοσης των μικτών μοντέλων στην συγκεκριμένη κατηγορία, η αναλογία των κατηγοριών των ομάδων είναι ίση, ενώ κάθε ομάδα περιλαμβάνει περίπου το ίδιο πλήθος παρατηρήσεων.

Στο Σχήμα 4.1 παρουσιάζεται ένα διδιάστατο σύνολο παρατηρήσεων μικτού τύπου.



Σχήμα 4.1 Μικτό σύνολο που αποτελείται από εννέα ομάδες διδιάστατων παρατηρήσεων. Περιλαμβάνει τρεις Gaussian με ελλειπτικό σχήμα, τρεις uniform με ορθογώνιο σχήμα και τρεις truncated Gaussian ως τμήματα ελλείψεων. Η αναλογία των κατηγοριών των ομάδων είναι σταθερή και οι ομάδες περιέχουν ίσο πλήθος παρατηρήσεων. Το αρχικό σύνολο διαμερίζεται τυχαία σε ένα σύνολο εκπαίδευσης και σε ένα σύνολο ελέγχου. Μετά το τέλος της εκπαίδευσης προκύπτουν οι εκτιμήσεις για το σύνολο των παραμέτρων του κάθε μοντέλου καθώς και του πλήθους των συνιστωσών κατανομών για τα AsPsMM, PsMM. Τέλος, υπολογίζεται η τιμή της αρνητικής λογαριθμικής πιθανοφάνειας για κάθε μοντέλο. Το μοντέλο με την ελάχιστη τιμή, πετυχαίνει την βέλτιστη περιγραφή του συνόλου παρατηρήσεων.

4.3. Πειραματική αξιολόγηση τεχνητών συνόλων παρατηρήσεων

Στις παρακάτω ενότητες (4.3.1 – 4.3.4) παρουσιάζονται πίνακες με πειραματικά αποτελέσματα από ανεξάρτητα σύνολα παρατηρήσεων τα οποία σχηματίζουν μη-επικαλυπτόμενες ομάδες. Τα σύνολα παρατηρήσεων διακρίνονται σε ομοιογενή, όταν οι παρατηρήσεις σχηματίζουν μια μοναδική ομάδα αλλά και σε σύνθετα, με περισσότερες ομάδες παρατηρήσεων. Ακόμη, διακρίνονται σε μονοδιάστατα και πολυδιάστατα. Στο πάνω μέρος του κάθε πίνακα αναγράφεται η κατηγορία, η διάσταση και ο πραγματικός αριθμός των ομάδων που σχηματίζουν οι παρατηρήσεις. Η απόδοση των τριών μοντέλων μικτών κατανομών ορίζεται μέσω της τιμής της αρνητικής λογαριθμικής πιθανοφάνειας σε κάθε σύνολο ελέγχου. Συμπερασματικά,

το μοντέλο που περιγράφει καλύτερα ένα σύνολο παρατηρήσεων, είναι αυτό που έχει την μικρότερη τιμή στην αρνητική λογαριθμική πιθανοφάνεια.

Στα πειράματα με μονοδιάστατα σύνολα παρατηρήσεων, οι τιμές της αρνητικής λογαριθμικής πιθανοφάνειας και της εκτίμησης του αριθμού των ομάδων κάθε συνόλου, προέκυψαν από τον μέσο όρο δέκα επαναλήψεων αντίστοιχων πειραμάτων. Επίσης, συνοδεύονται από γραφικές αναπαραστάσεις των PsMM και AsPsMM. Στα πολυδιάστατα σύνολα παρατηρήσεων, οι αντίστοιχες τιμές προέκυψαν από τον μέσο όρο δύο επαναλήψεων αντίστοιχων πειραμάτων.

4.3.1. Ομοιογενή σύνολα μονοδιάστατων παρατηρήσεων

Πίνακας 4.1 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Gaussian, $D=1$, $K=1$, $N_{test\ set}=4500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	1/1	6575
PsMM	1/1	6576
GMM	1	6554

Πίνακας 4.2 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

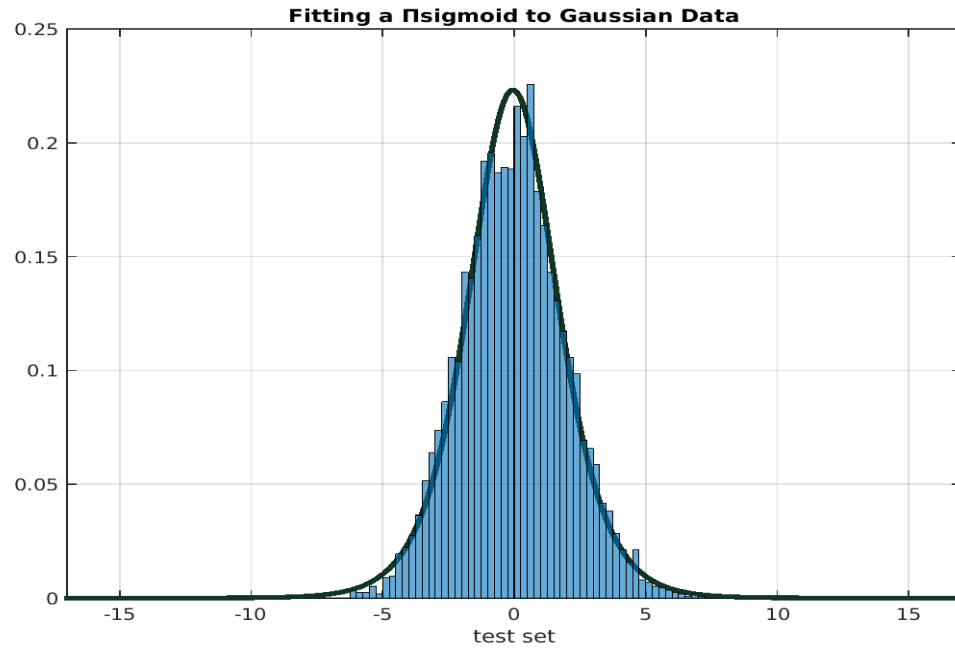
Dataset: Uniform, $D=1$, $K=1$, $N_{test\ set}=4500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	1/1	-2176
PsMM	1/1	-2018
GMM	1	-1437

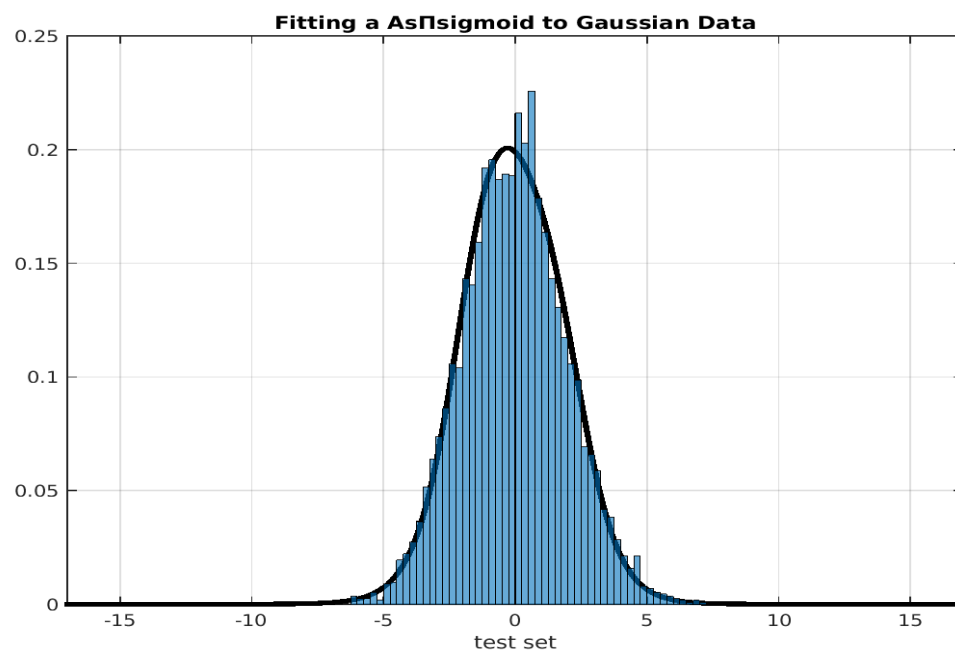
Πίνακας 4.3 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Truncated Gaussian, $D=1$, $K=1$, $N_{test\ set}=4500$

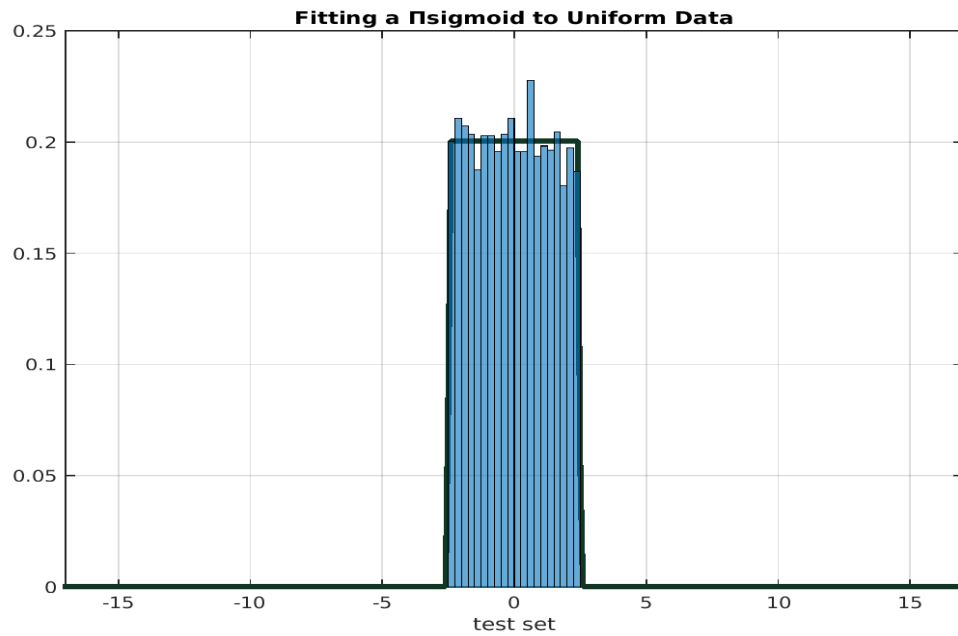
Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	1/1	3402
PsMM	1/1	4004
GMM	1	4069



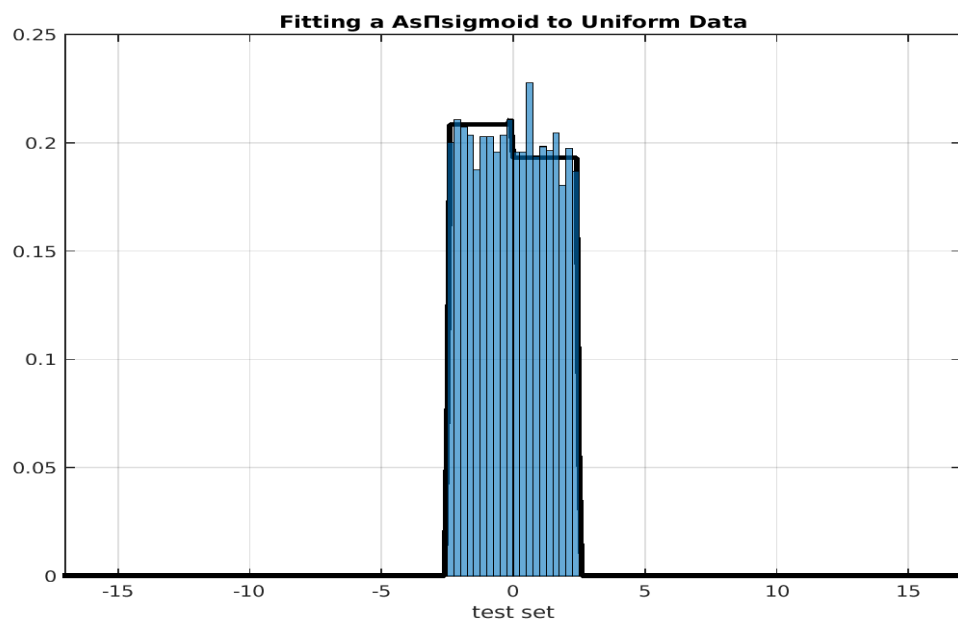
Σχήμα 4.2 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία gaussian ομάδα μέσω της κατανομής Π -sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης.



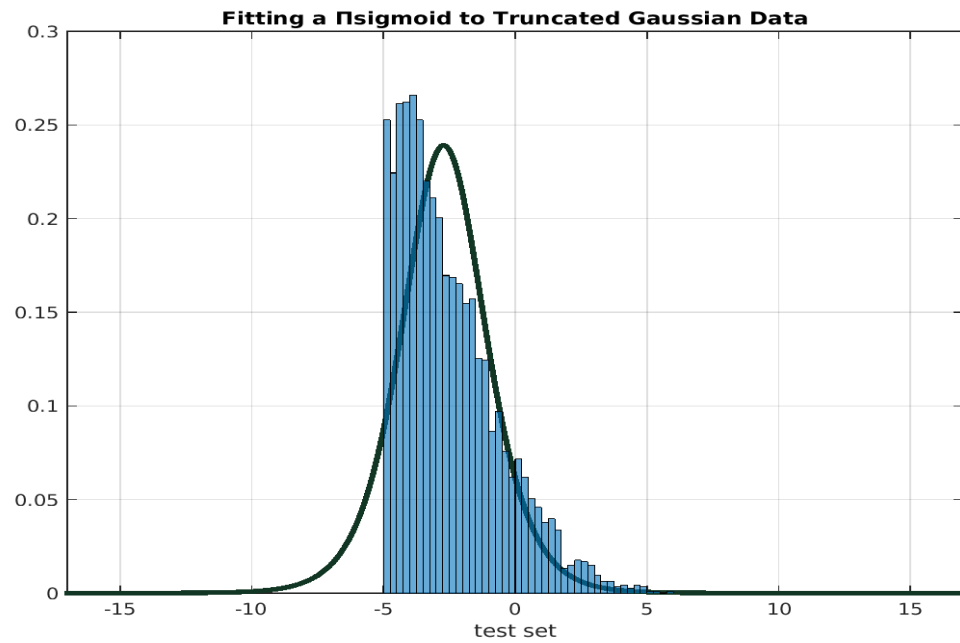
Σχήμα 4.3 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία gaussian ομάδα μέσω της κατανομής As Π -sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης.



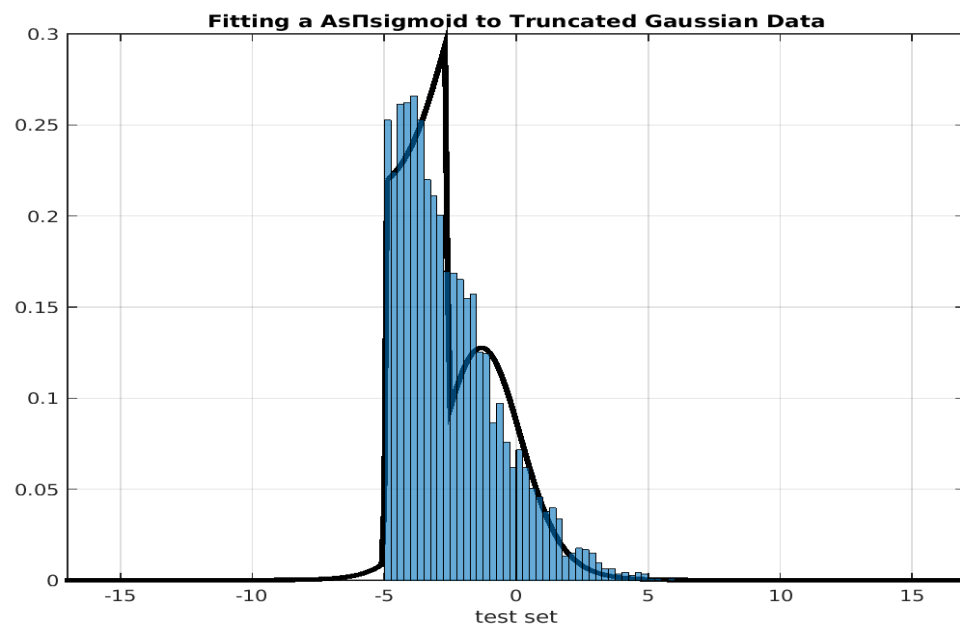
Σχήμα 4.4 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία uniform ομάδα μέσω της κατανομής Π-sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης.



Σχήμα 4.5 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία uniform ομάδα μέσω της κατανομής AsΠ-sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης.



Σχήμα 4.6 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία truncated gaussian ομάδα μέσω της κατανομής Π -sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης.



Σχήμα 4.7 Η μοντελοποίηση ενός συνόλου ελέγχου που αποτελείται από μια ενιαία truncated gaussian ομάδα μέσω της κατανομής As Π -sigmoid. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου MLE στο σύνολο εκπαίδευσης.

4.3.2. Σύνθετα σύνολα μονοδιάστατων παρατηρήσεων

Πίνακας 4.4 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

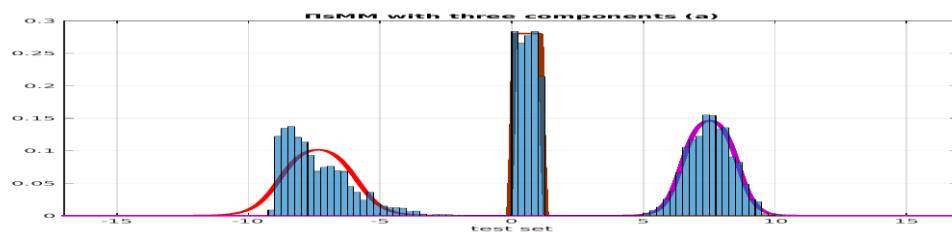
Dataset: Mixed, $D=1$, $K=3$, $N_{test\ set}=4500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	3/3	9517
PsMM	3/3	9727
GMM	3	11882

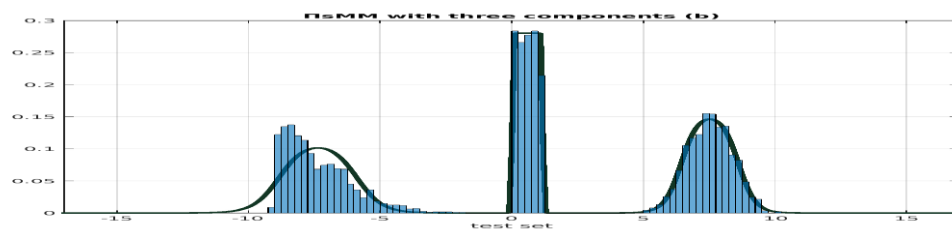
Πίνακας 4.5 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Mixed, $D=1$, $K=6$, $N_{test\ set}=4500$

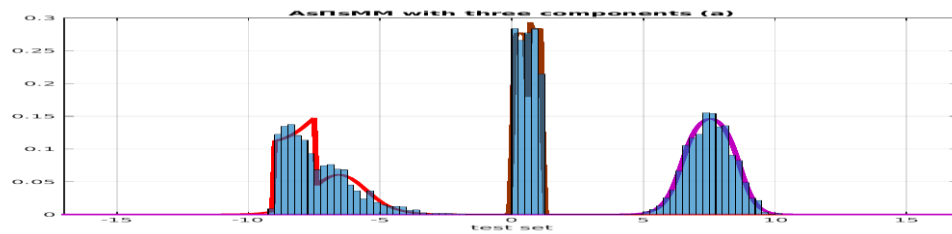
Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	6/6	12845
PsMM	6.1/6	13033
GMM	6	14072



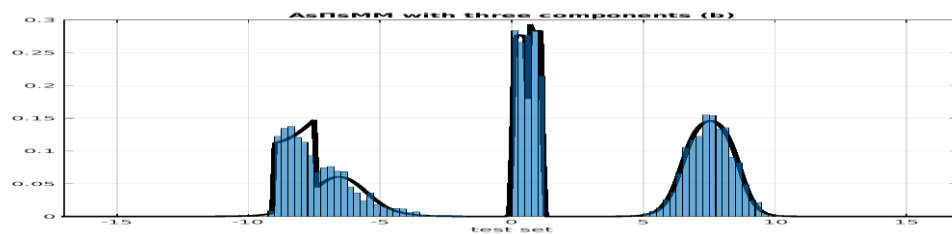
Σχήμα 4.8 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από τρεις ομάδες παρατηρήσεων μέσω τριών Π-sigmoid κατανομών. Κάθε κατανομή μοντελοποιεί μια ενιαία ομάδα παρατηρήσεων.



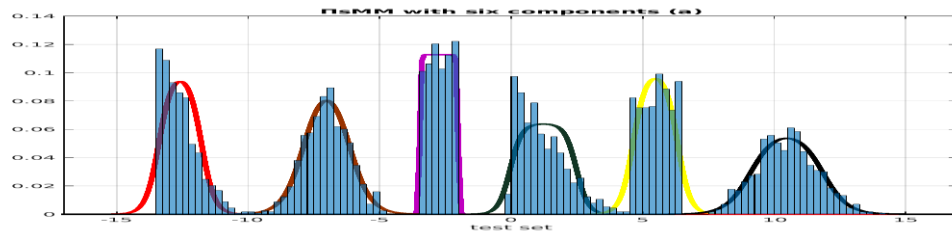
Σχήμα 4.9 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από τρεις ομάδες παρατηρήσεων μέσω του PsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου GEM στο σύνολο εκπαίδευσης.



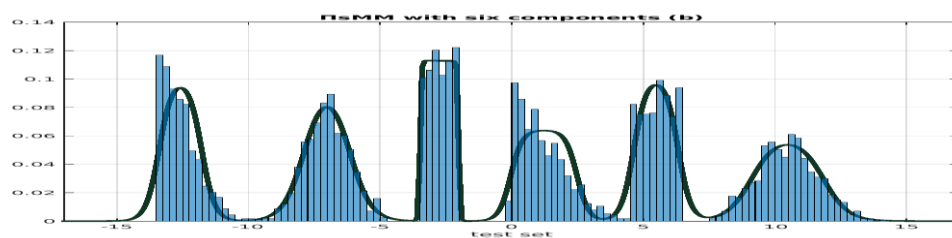
Σχήμα 4.10 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από τρεις ομάδες παρατηρήσεων μέσω τριών AsΠ-sigmoid κατανομών. Κάθε κατανομή μοντελοποιεί μια ενιαία ομάδα παρατηρήσεων.



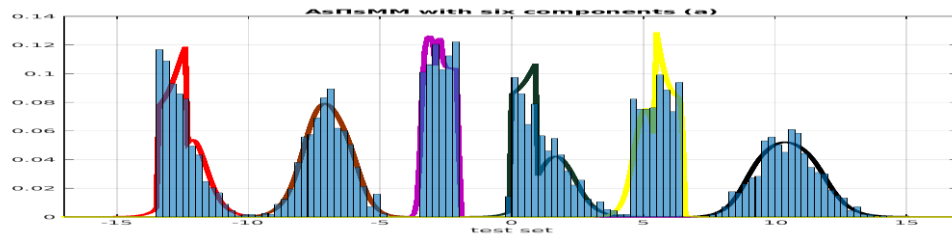
Σχήμα 4.11 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από τρεις ομάδες παρατηρήσεων μέσω του AsΠsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου GEM στο σύνολο εκπαίδευσης.



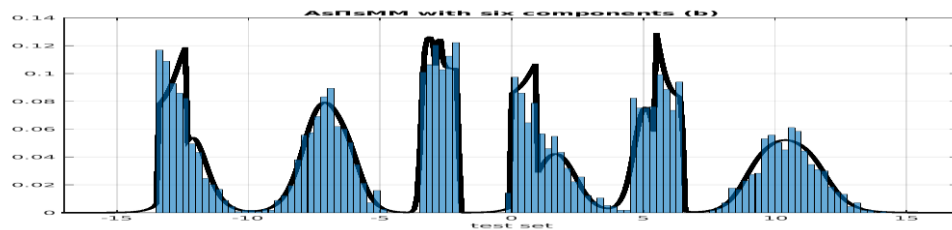
Σχήμα 4.12 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από έξι ομάδες παρατηρήσεων μέσω έξι Π-sigmoid κατανομών. Κάθε κατανομή μοντελοποιεί μια ενιαία ομάδα παρατηρήσεων.



Σχήμα 4.13 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από έξι ομάδες παρατηρήσεων μέσω του ΠsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου GEM στο σύνολο εκπαίδευσης.



Σχήμα 4.14 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από έξι ομάδες παρατηρήσεων μέσω έξι AsPi-sigmoid κατανομών. Κάθε κατανομή μοντελοποιεί μια ενιαία ομάδα παρατηρήσεων.



Σχήμα 4.15 Η μοντελοποίηση ενός mixed συνόλου ελέγχου που αποτελείται από έξι ομάδες παρατηρήσεων μέσω του AsPiMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε με εφαρμογή του αλγορίθμου GEM στο σύνολο εκπαίδευσης

4.3.3. Ομοιογενή σύνολα πολυδιάστατων παρατηρήσεων

Πίνακας 4.6 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPiMM, PiMM, GMM.

Dataset: Gaussian, $D=2, 4, 6$, $K=1$, $N_{test\ set}=1500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood $D=2$	Test -Log Likelihood $D=4$	Test -Log Likelihood $D=6$
AsPiMM	1/1	2640	5267	8156
PiMM	1/1	2647	5268	8143
GMM	1	2631	5255	8131

Πίνακας 4.7 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPiMM, PiMM, GMM.

Dataset: Uniform, $D=2, 4, 6$, $K=1$, $N_{test\ set}=1500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood $D=2$	Test -Log Likelihood $D=4$	Test -Log Likelihood $D=6$
AsPiMM	1/1	2093	4185	68395
PiMM	1/1	2086	4174	62550
GMM	1	2566	5213	78397

Πίνακας 4.8 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Truncated Gaussian, $D=2, 4, 6$, $K=1$, $N_{test\ set}=1500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood $D=2$	Test -Log Likelihood $D=4$	Test -Log Likelihood $D=6$
AsPsMM	1/1	4060	6478	7193
PsMM	1/1	4285	6675	7396
GMM	1	4277	6676	7400

4.3.4. Σύνθετα σύνολα πολυδιάστατων παρατηρήσεων

Πίνακας 4.9 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Mixed, $D=2$, $K=3$, $N_{test\ set}=2500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	3/3	5329
PsMM	3/3	5409
GMM	3	5637

Πίνακας 4.10 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Mixed, $D=2$, $K=6$, $N_{test\ set}=2500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	6/6	6030
PsMM	6/6	6122
GMM	6	6758

Πίνακας 4.11 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Mixed, $D=2$, $K=9$, $N_{test\ set}=2500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	9/9	8620
PsMM	9/9	8831
GMM	9	9232

Πίνακας 4.12 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Mixed, D=4, K=3, $N_{test\ set}=2500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	3/3	2891
PsMM	3/3	2921
GMM	3	3061

Πίνακας 4.13 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Mixed, D=4, K=6, $N_{test\ set}=2500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	6/6	6936
PsMM	6/6	6992
GMM	6	7778

Πίνακας 4.14 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Mixed, D=4, K=9, $N_{test\ set}=2500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	9/9	11894
PsMM	9/9	12543
GMM	9	13035

Πίνακας 4.15 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Mixed, D=6, K=3, $N_{test\ set}=2500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	3/3	10666
PsMM	3/3	10711
GMM	3	11246

Πίνακας 4.16 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Mixed, D=6, K=6, $N_{test\ set}=2500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	6/6	19387
PsMM	6/6	19437
GMM	6	21087

Πίνακας 4.17 Η εκτίμηση του συνολικού αριθμού των ομάδων και ο υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας για τα AsPsMM, PsMM, GMM.

Dataset: Mixed, D=6, K=9, $N_{test\ set}=2500$

Model	Estimated Clusters/Real clusters	Test -Log Likelihood
AsPsMM	9/9	18786
PsMM	9/9	19080
GMM	9	19736

4.4. Συμπεράσματα

Σύμφωνα με τα αποτελέσματα της πειραματικής διαδικασίας προκύπτουν τα παρακάτω συμπεράσματα:

- Η AsΠ-sigmoid κατανομή μοντελοποιεί βέλτιστα τα truncated Gaussian σύνολα παρατηρήσεων. Η μη συμμετρική της ιδιότητα, εξασφαλίζει την αποδοτική περιγραφή κάθε μη-συμμετρικού συνόλου παρατηρήσεων με αντιπροσωπευτική κατηγορία τα truncated Gaussian σύνολα. Αντίθετα, η συμμετρικότητα της Π-sigmoid και της Gaussian κατανομής οδηγεί σε μια ελλειμματική περιγραφή αντίστοιχων συνόλων παρατηρήσεων.
- Η Π-sigmoid κατανομή μοντελοποιεί βέλτιστα τα uniform σύνολα παρατηρήσεων. Επίσης, η μοντελοποίηση τους μέσω της AsΠ-sigmoid κατανομής κρίνεται ιδιαίτερα αποδοτική. Η μικρή απόκλιση στην απόδοση των δύο κατανομών, προκύπτει από την μεγάλη ευαισθησία της AsΠ-sigmoid κατανομής στην περιγραφή τιμών γύρω από την μέση τιμή των παρατηρήσεων, αλλά και την δυσκολία της να προσαρμόσει μια πλήρως συμμετρική συμπεριφορά στα uniform σύνολα παρατηρήσεων. Παρόλα αυτά, η απόδοση της είναι ιδιαίτερα αξιόπιστη και άμεσα συγκρίσιμη με αυτήν της Π-sigmoid κατανομής. Αντίθετα, η Gausssian κατανομή μειονεκτεί αισθητά στην μοντελοποίηση των uniform συνόλων παρατηρήσεων, καθώς η γεωμετρική της προσαρμογή (ελλειπτικό σχήμα) στις ομοιόμορφες ομάδες παρατηρήσεων (ορθογώνιο σχήμα) κρίνεται ελλειμματική. Συμπερασματικά, παρουσιάζει κατώτερη επίδοση σε σχέση με τα άλλα μοντέλα.
- Η Gaussian κατανομή μοντελοποιεί – προφανώς – με βέλτιστο τρόπο τα gaussian σύνολα παρατηρήσεων. Ταυτόχρονα, η μοντελοποίηση τους μέσω των κατανομών Π-sigmoid και AsΠ-sigmoid κρίνεται αποτελεσματική και ισοδύναμη. Οι κατανομές Π-sigmoid και AsΠ-sigmoid μέσω κατάλληλων τιμών των παραμέτρων τους, προσαρμόζουν μια συμμετρική καμπανοειδή περιγραφή στα Gaussian σύνολα παρατηρήσεων, ανάλογη με αυτήν της Gaussian κατανομής. Το συμπέρασμα αυτό αναδεικνύει την αυξημένη τους ευελιξία έναντι της Gaussian κατανομής.

Στην ανάλυση της απόδοσης των μοντέλων, διαπιστώθηκε αφενός η σαφής υπεροχή της AsΠ-sigmoid κατανομής στη περιγραφή μη-συμμετρικών συνόλων παρατηρήσεων (truncated Gaussian) και αφετέρου μια ιδιαίτερα αξιόπιστη περιγραφή συμμετρικών συνόλων παρατηρήσεων (gaussian, uniform). Συμπερασματικά, η ευελιξία και η καλή γενικευτική ικανότητα της AsΠ-sigmoid κατανομής της προσδίδει την ικανότητα της βέλτιστης περιγραφής των mixed συνόλων παρατηρήσεων όπως αναδείχτηκε και στην πειραματική διαδικασία.

Αναφορικά με την δυνατότητα εκτίμησης τους πλήθους των συνιστωσών κατανομών των μικτών μοντέλων μέσω της ενίσχυσης της διαδικασίας εκπαίδευσης τους με το κριτήριο dip, η πειραματική διαδικασία ανέδειξε την ακρίβεια της εκτίμησης. Τα πειραματικά σύνολα περιελάμβαναν μη επικαλυπτόμενες ομάδες παρατηρήσεων οι οποίες σχεδόν στο σύνολο των περιπτώσεων εντοπίστηκαν επιτυχώς στην διαδικασία εκπαίδευσης των μικτών μοντέλων.

4.5. Πειραματική αξιολόγηση πραγματικών συνόλων παρατηρήσεων

Στην ενότητα αυτή θα γίνει η εφαρμογή των τριών μικτών μοντέλων (GMM, ΠsMM, AsΠsMM) με στόχο την κατάτμηση εικόνων. Αντικειμενικός σκοπός της εφαρμογής αυτής είναι η ανίχνευση ομογενών περιοχών (τμημάτων) σε κάθε εικόνα με βάση την απόχρωση του γκρι που αναπαριστά κάθε εικονοστοιχείο (pixel). Για την υλοποίηση της εφαρμογής, κάθε εικόνα αρχικά κανονικοποιείται μέσω του μετασχηματισμού:

$$z = \frac{x - \mu}{\sigma} \quad (4.1)$$

Όπου μ και σ είναι η μέση τιμή και η τυπική απόκλιση του συνόλου των εικονοστοιχείων (εικόνας). Στην συνέχεια, οι νέες τιμές ορίζουν ένα μονοδιάστατο σύνολο που αποτελείται από $W * H$ παρατηρήσεις όπου με W και H συμβολίζονται οι διαστάσεις της εικόνας. Στην υλοποίηση, χρησιμοποιήθηκαν εικόνες ανάλυσης 140*210, οπότε κάθε εικόνα ορίζει ένα σύνολο 29400 μονοδιάστατων παρατηρήσεων.

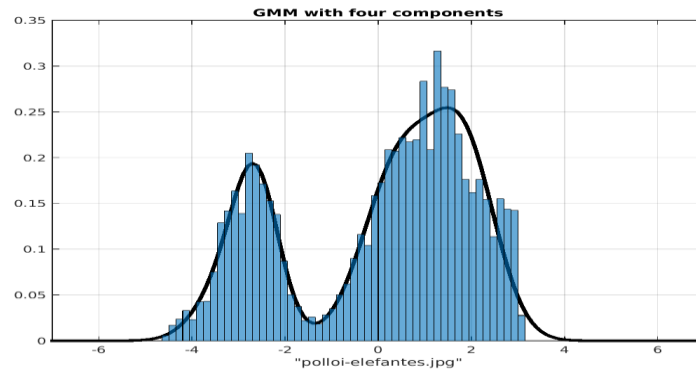
Σε αντιστοιχία με την μέθοδο μοντελοποίησης τεχνητών συνόλων παρατηρήσεων που αναλύθηκε στην προηγούμενη ενότητα, κάθε σύνολο μονοδιάστατων παρατηρήσεων διαμοιράζεται με τυχαίο τρόπο σε δύο ξένα υποσύνολα που περιλαμβάνουν ίσο πλήθος παρατηρήσεων: το σύνολο εκπαίδευσης (training set) και το σύνολο ελέγχου (test set). Μέσω του συνόλου εκπαίδευσης, γίνεται αρχικά εκτίμηση του συνόλου των παραμέτρων του μοντέλου, ενώ στη συνέχεια, μέσω του συνόλου ελέγχου υπολογίζεται η λογαριθμική πιθανοφάνεια του μοντέλου, με δεδομένες τις εκτιμήσεις των παραμέτρων που υπολογίστηκαν μέσω της διαδικασίας της εκπαίδευσης. Ως μέτρο σύγκρισης των αποτελεσμάτων μοντελοποίησης της κάθε εικόνας χρησιμοποιείται η τιμή της αρνητικής λογαριθμικής πιθανοφάνειας. Το μοντέλο με την μικρότερη τιμή στην αρνητική λογαριθμική πιθανοφάνεια του συνόλου ελέγχου μοντελοποιεί με τον βέλτιστο τρόπο το σύνολο των παρατηρήσεων.

Για την κατάτμηση μιας εικόνας μέσω των μικτών μοντέλων απαιτείται μια διαδικασία ομαδοποίησης. Αρχικά κάθε μικτό μοντέλο εκπαιδεύεται στο κανονικοποιημένο σύνολο παρατηρήσεων, με στόχο την εκτίμηση του συνόλου των παραμέτρων, για ένα προκαθορισμένο πλήθος συνιστωσών κατανομών. Στην συνέχεια, υπό την παραδοχή ότι οι παρατηρήσεις που παράγονται από την ίδια συνιστώσα κατανομή σχηματίζουν μια ενιαία ομάδα, προκύπτει η ομαδοποίηση του συνόλου των παρατηρήσεων στο προκαθορισμένο αριθμό ομάδων, μέσω των posterior πιθανοτήτων. Ενδεικτικό πλήθος που χρησιμοποιήθηκε στην υλοποίηση είναι οι τέσσερις ομάδες. Μετά το πέρας της ανάθεσης όλων των παρατηρήσεων στις τέσσερις ομάδες και για τον υπολογισμό της απόχρωσης του γκρι που θα χρωματιστεί η κάθε ομάδα εφαρμόζεται ο αντίστροφος μετασχηματισμός της σχέσης (4.1)

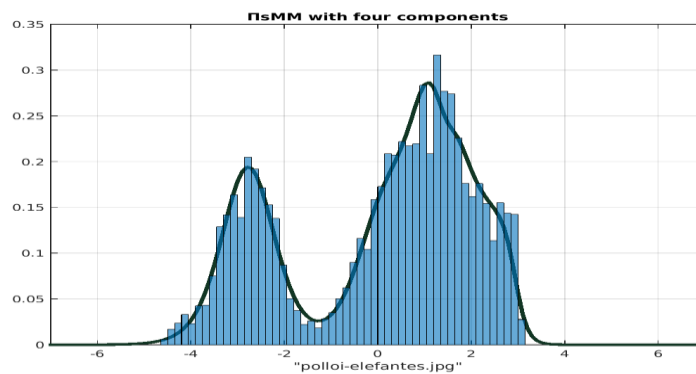
$$x = z * \sigma + \mu \quad (4.2)$$

Στη συνέχεια, ως τιμή χρωματισμού κάθε ομάδας επιλέγεται η μέση τιμή των παρατηρήσεων που την συνιστούν. Συμπερασματικά, τα εικονοστοιχεία (pixels) που ανήκουν στην ίδια ομάδα, «χρωματίζονται» με την ίδια απόχρωση του γκρι, οριοθετώντας με αυτόν τον τρόπο και μια ομοιογενή περιοχή στην αρχική εικόνα.

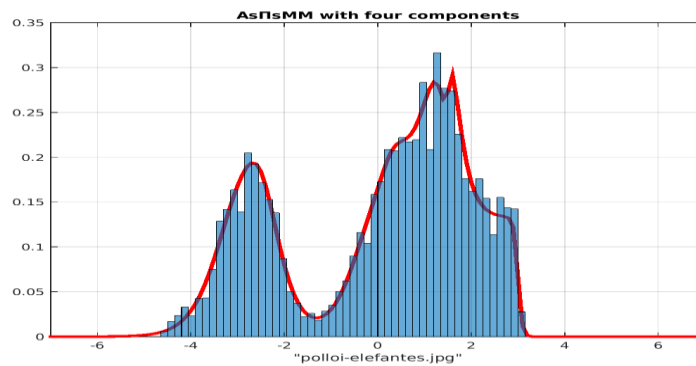
Στην συνέχεια παρουσιάζονται τα αποτελέσματα της μοντελοποίησης των εικονοστοιχείων (pixels) τεσσάρων εικόνων από τα GMM, PsMM και AsPsMM καθώς επίσης και τα αποτελέσματα της κατάτμησης των εικόνων.



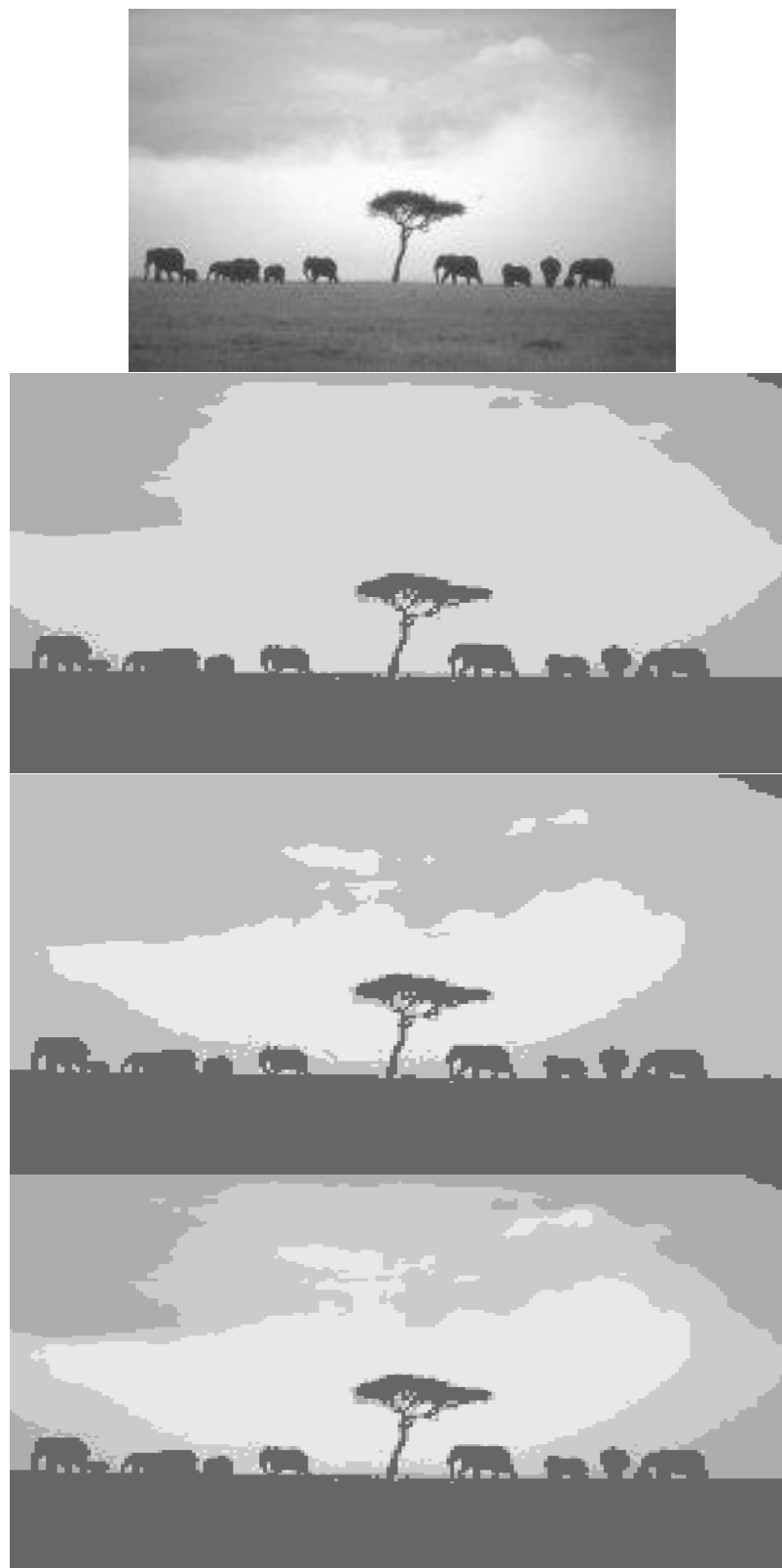
Σχήμα 4.16 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “elephants.jpg” μέσω του GMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



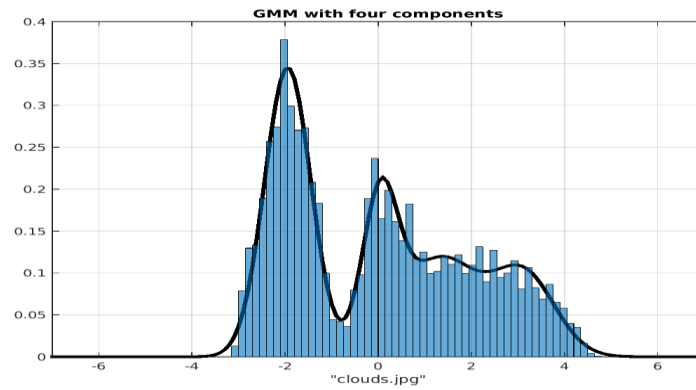
Σχήμα 4.17 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “elephants.jpg” μέσω του PsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



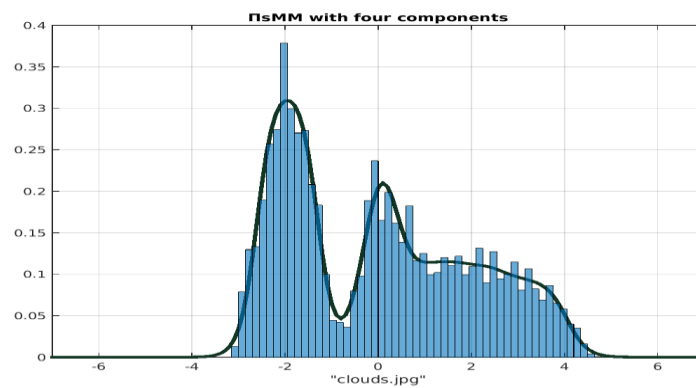
Σχήμα 4.18 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “elephants.jpg” μέσω του AsPsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



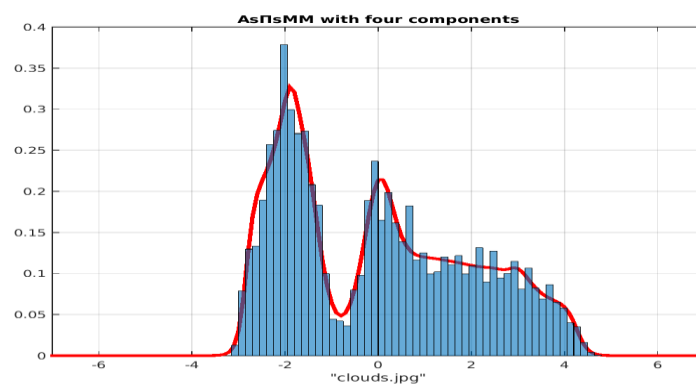
Σχήμα 4.19 Το αποτέλεσμα της κατάτμησης της αρχικής εικόνας “elephants.jpg” μέσω των τριών μοντέλων: GMM, PsMM, AsPsMM. Κάθε μοντέλο αποτελείται από τέσσερις συνιστώσες κατανομές (ομάδες).



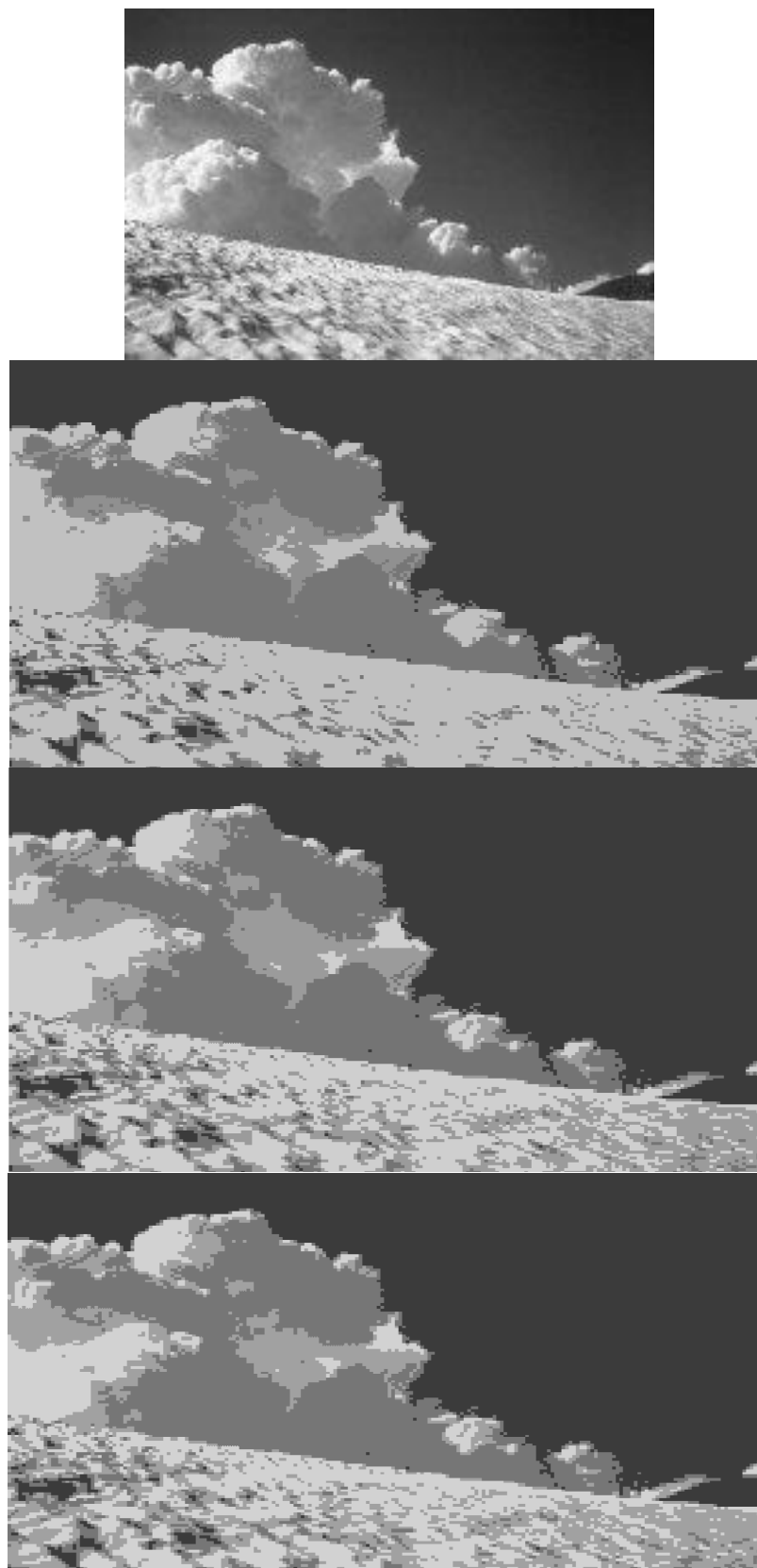
Σχήμα 4.20 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “clouds.jpg” μέσω του GMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



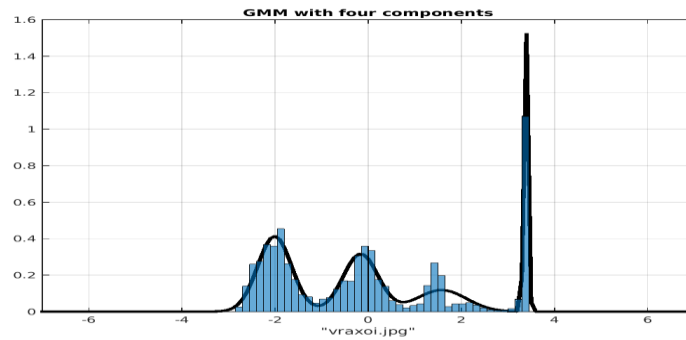
Σχήμα 4.21 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “clouds.jpg” μέσω του PIsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



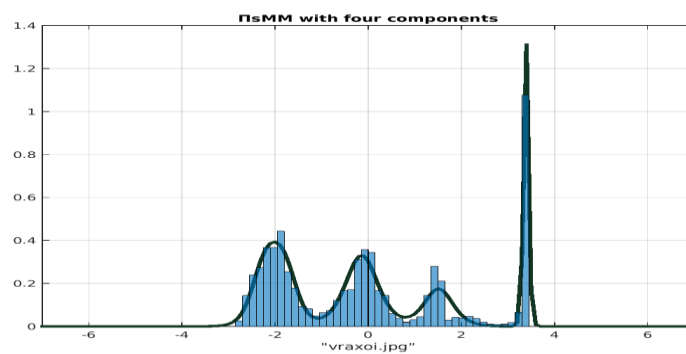
Σχήμα 4.22 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “clouds.jpg” μέσω του AsPIsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



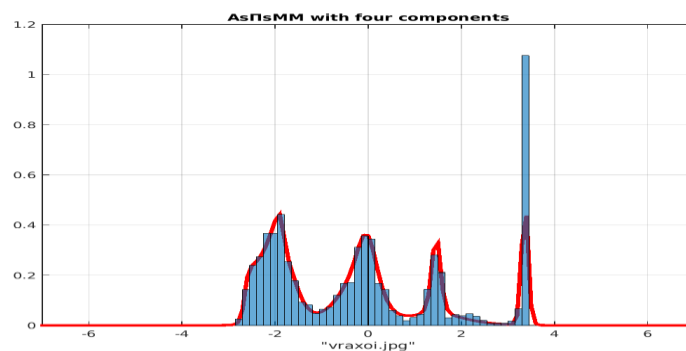
Σχήμα 4.23 Το αποτέλεσμα της κατάτμησης της αρχικής εικόνας “clouds.jpg” μέσω των τριών μοντέλων: GMM, PsMM, AsPsMM. Κάθε μοντέλο αποτελείται από τέσσερις συνιστώσες κατανομές (ομάδες).



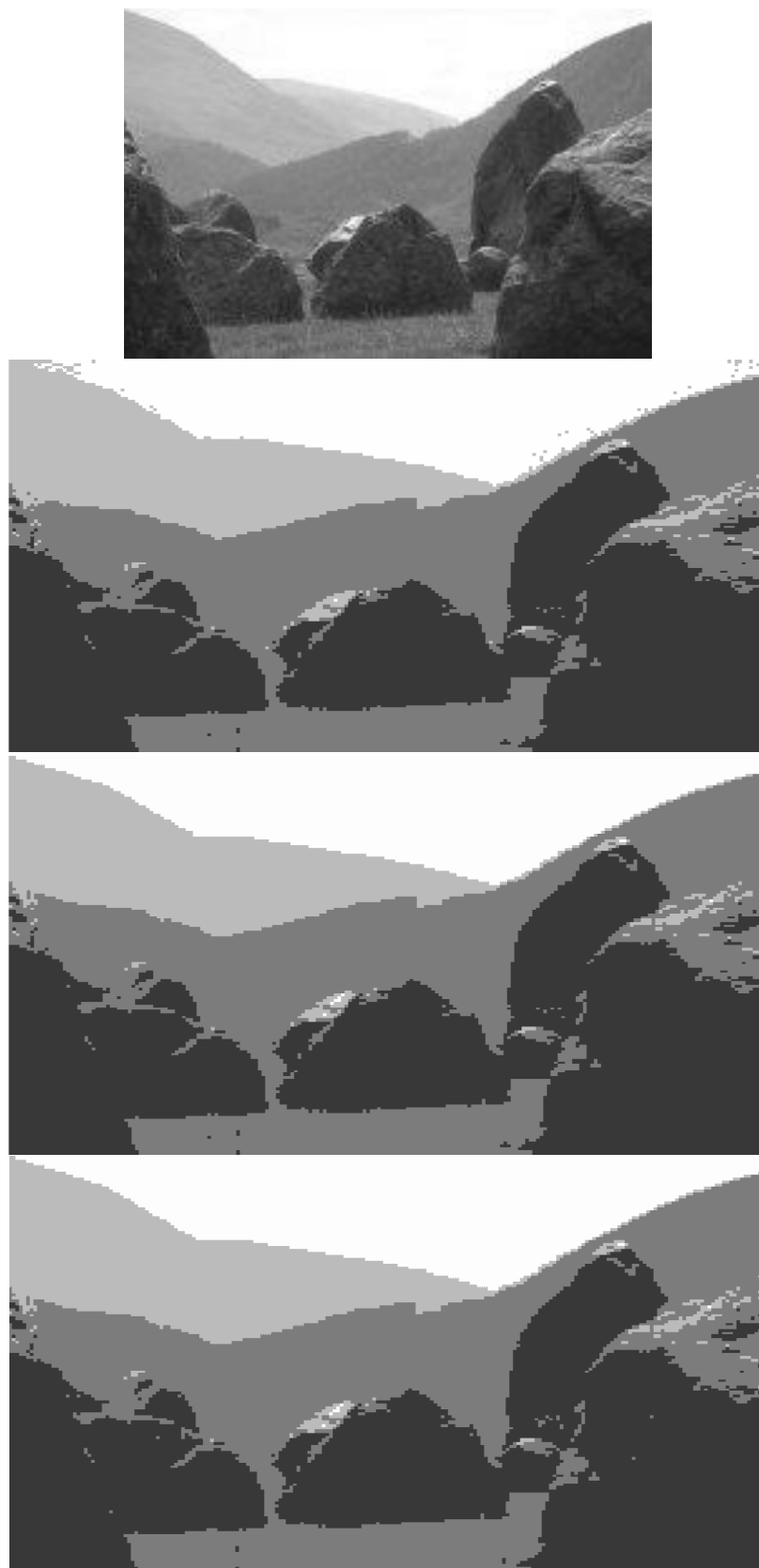
Σχήμα 4.24 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rocks.jpg” μέσω του GMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



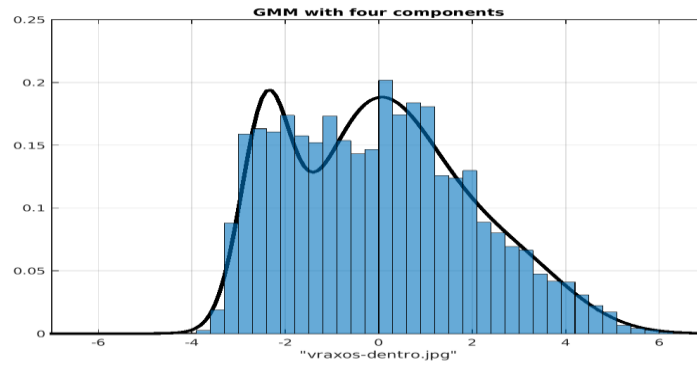
Σχήμα 4.25 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rocks.jpg” μέσω του PsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



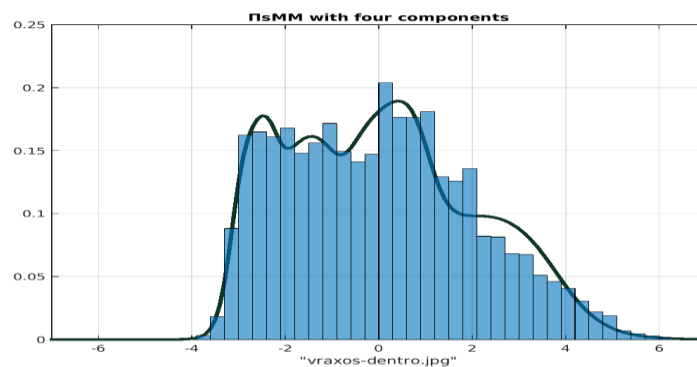
Σχήμα 4.26 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rocks.jpg” μέσω του AsPsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



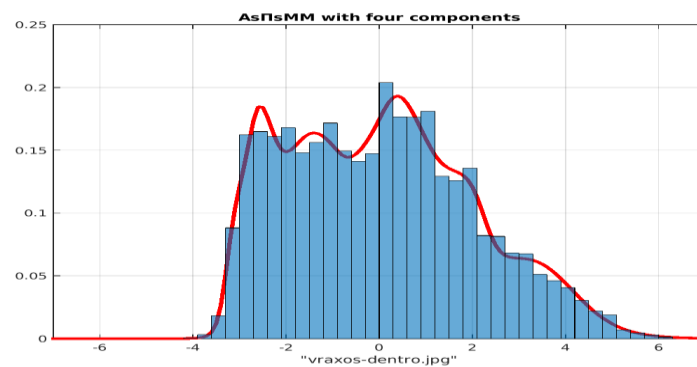
Σχήμα 4.27 Το αποτέλεσμα της κατάτμησης της αρχικής εικόνας “rocks.jpg” μέσω των τριών μοντέλων: GMM, PsMM, AsPsMM. Κάθε μοντέλο αποτελείται από τέσσερις συνιστώσες κατανομές (ομάδες).



Σχήμα 4.28 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rock_and_tree.jpg” μέσω του GMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



Σχήμα 4.29 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rock_and_tree.jpg” μέσω του AsPsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



Σχήμα 4.30 Η μοντελοποίηση του συνόλου ελέγχου της εικόνας “rock_and_tree.jpg” μέσω του AsPsMM. Η εκτίμηση του συνόλου των παραμέτρων έγινε στο σύνολο εκπαίδευσης.



Σχήμα 4.31 Το αποτέλεσμα της κατάτμησης της αρχικής εικόνας “rock_and_tree.jpg” μέσω των τριών μοντέλων: GMM, PsMM, AsPsMM. Κάθε μοντέλο αποτελείται από τέσσερις συνιστώσες κατανομές (ομάδες).

Τα αποτελέσματα της κατάτμησης υποδεικνύουν πως το AsPsMM βελτιώνει την ποιότητα των παραγόμενων λύσεων σε σχέση με τα μοντέλα PsMM και GMM. Η κατάτμηση στην εικόνα “elephants.jpg” μέσω του AsPsMM είναι μια ενδεικτική περίπτωση στην οποία το μοντέλο ομαδοποιεί πολύ ικανοποιητικά την αρχική εικόνα και οι τελικές ομάδες που προκύπτουν αναπαριστούν αποδοτικά τις διαφορετικές περιοχές της εικόνας.

Στον παρακάτω πίνακα παρουσιάζονται τα συνολικά αποτελέσματα της αρνητικής λογαριθμικής πιθανοφάνειας για κάθε μοντέλο, στο σύνολο ελέγχου (test set) της κάθε εικόνας:

Πίνακας 4.18 Σύγκριση των μοντέλων GMM,PsMM και AsPsMM σε gray-scale εικόνες. Ισχύει: $N_{train} = N_{test} = 14700$

Image	Clusters	GMM Test -Log Likelihood	PsMM Test -Log Likelihood	AsPsMM Test -Log Likelihood
elephants	4	27280	27084	26958
clouds.jpg	4	27732	27704	27637
rocks.jpg	4	20477	20442	11069
rock_and_tree.jpg	4	30032	30011	29934

Τα αποτελέσματα του πίνακα αποδεικνύουν την σαφή υπεροχή του AsPsMM έναντι των μοντέλων GMM και PsMM στην μοντελοποίηση (και την κατάτμηση) εικόνων. Συγκεκριμένα, η μοντελοποίηση των εικονοστοιχείων (pixels) της εικόνας “rocks.jpg” μέσω του AsPsMM, είναι μια ενδεικτική περίπτωση στην οποία το μοντέλο παρουσιάζει σημαντικά βελτιωμένη επίδοση. Το γεγονός αυτό αναδεικνύει την γενικότερη ευελιξία του ασύμμετρου μοντέλου στην περιγραφή διάφορων συνόλων παρατηρήσεων, καθιστώντας το ως ένα ιδιαίτερα αποδοτικό μοντέλο για την επίλυση αντίστοιχων προβλημάτων.

Σύμφωνα με τις παραπάνω παρατηρήσεις αλλά και τα συνολικά πειραματικά αποτελέσματα από τα πραγματικά και τα τεχνητά σύνολα παρατηρήσεων, προκύπτει ότι το AsPsMM οδηγεί στην καλύτερη περιγραφή συνόλων παρατηρήσεων με άγνωστες στατιστικές ιδιότητες.

ΚΕΦΑΛΑΙΟ 5. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

5.1. Συμπεράσματα

5.2. Μελλοντική έρευνα

5.1. Συμπεράσματα

Στην παρούσα εργασία προτάθηκε η ασύμμετρη Π-σιγμοειδής κατανομή που ορίζεται ως ένα μικτό μοντέλο από δύο Π-σιγμοειδείς κατανομές με διαφορετική κλίση η καθεμία. Μελετήθηκαν οι ιδιότητες της νέας κατανομής και ορίστηκε το μοντέλο μικτών ασύμμετρων Π-σιγμοειδών κατανομών. Επιπλέον, προτάθηκε ένας αλγόριθμος εκπαίδευσης του μικτού μοντέλου που βασίζεται στον αλγόριθμο GEM και εμπεριέχει το στατιστικό κριτήριο dip . Αποτέλεσμα της διαδικασίας εκπαίδευσης είναι ο προσδιορισμός του συνόλου των παραμέτρων του μοντέλου και ταυτόχρονα ο υπολογισμός του πλήθους των συνιστωσών κατανομών. Η επίδοση των προτεινόμενων στατιστικών μοντέλων και των μεθόδων εκπαίδευσής τους εξετάστηκε πειραματικά σε τεχνητά και πραγματικά σύνολα παρατηρήσεων. Η πειραματική αξιολόγηση ανέδειξε την βελτιωμένη επίδοση του μικτού μοντέλου ασύμμετρων Π-σιγμοειδών κατανομών στην μοντελοποίηση συνόλων παρατηρήσεων και στην κατάτμηση εικόνων. Η βελτιωμένη επίδοση προκύπτει από την δυνατότητα της ασύμμετρης Π-σιγμοειδής κατανομής να μοντελοποιεί παρατηρήσεις που προέρχονται από γενικές κατανομές. Συμπερασματικά, το μικτό μοντέλο ασύμμετρων Π-σιγμοειδών κατανομών αποτελεί ένα αποδοτικό παραμετρικό μοντέλο για τη στατιστική μοντελοποίηση συνόλων παρατηρήσεων.

5.2. Μελλοντική έρευνα

Ως κατευθύνσεις για μελλοντική έρευνα που αξιοποιούν την προτεινόμενη ασύμμετρη Π-σιγμοειδή κατανομή, προτείνονται τα παρακάτω:

- **Στην διαδικασία εκπαίδευσης του AsPsMM μέσω του αλγορίθμου GEM, η αρχικοποίηση του μοντέλου μέσω του PsMM.**

Όπως τονίστηκε στο κεφάλαιο τρία η απόδοση του αλγορίθμου GEM εξαρτάται σε πολύ μεγάλο βαθμό από τις αρχικές τιμές από τις οποίες θα ξεκινήσει η διαδικασία της βελτιστοποίησης. Για την αντιμετώπιση αυτού του προβλήματος, αρχικά εκπαιδεύσαμε ένα GMM και στην συνέχεια εντοπίσαμε συσχετισμούς του συνόλου των παραμέτρων του AsPsMM με το σύνολο των παραμέτρων του GMM επιτυγχάνοντας μια αποδοτική αρχικοποίηση. Η διαδικασία αυτή μπορεί να βελτιωθεί, εκπαιδεύοντας αρχικά ένα PsMM και εντοπίζοντας κατάλληλους συσχετισμούς μεταξύ του συνόλου των παραμέτρων του AsPsMM και του PsMM.

- **Εφαρμογή της ασύμμετρης Π-σιγμοειδούς κατανομής στον ταξινομητή Naïve Bayes.**

Για την υλοποίηση του ταξινομητή Naïve Bayes υποθέτουμε την ανεξαρτησία των παρατηρήσεων σε κάθε διάσταση. Η διαδικασία μπορεί να τροποποιηθεί αντικαθιστώντας την κανονική κατανομή με την ασύμμετρη Π-σιγμοειδή κατανομή.

- **Εκπαίδευση ενός AsPsMM για την μοντελοποίηση εξαρτημένων χαρακτηριστικών.**

Εάν οι παρατηρήσεις δεν είναι ανεξάρτητες ανά διάσταση, τότε σχηματίζουν περιστραμμένες ως προς τους άξονες ελλείψεις και απαιτείται ειδικός μετασχηματισμός στον αλγόριθμο εκπαίδευσης για την αποδοτική μοντελοποίηση τους.

ΑΝΑΦΟΡΕΣ

- [1] C. M. Bishop, Pattern recognition and machine learning, Springer 2006.
- [2] A. P. Dempster, N. M. Laird and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B, 39:1-38, 1977
- [3] C. Blake and Mertz C. (7998). UCI repository of machine learning databases.
[HTTP://WWW.ICS.UCI.EDU/~MLEARN/DATABASES](http://www.ics.uci.edu/~mllearn/databases)
- [4] J. A Hartigan and P. M. Hartigan, The dip test of unimodality, The Annals of Statistics, 13(1), pp 70-84, 1985
- [5] A. Alivanoglou and A. Likas, Probabilistic models based on the Π-Sigmoid distribution, Proc. 3rd IAPR Workshop on Artificial Neural Networks in Pattern Recognition (ANNPR), 2008
- [6] J. Nocedal and S. J. Wright, Numerical Optimization, Springer 1999
- [7] A. Kalogeratos and A. Likas, Dip-means: an incremental clustering method for estimating the number of clusters, Proc. Neural Information Processing Systems (NIPS'12), Lake Tahoe, Nevada, USA, 2012
- [8] Α. Αλιβάνογλου, Μικτά μοντέλα Π-σιγμοειδών κατανομών, Μεταπτυχιακή Εργασία Εξειδίκευσης, Παν. Ιωαννίνων, Τμήμα Πληροφορικής, 2007

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Ο Μπαλίκας Ευστάθιος γεννήθηκε στην Καβάλα το 1989. Το 2013 ολοκλήρωσε τις σπουδές του στο Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων και από τον Οκτώβριο του 2015 είναι μεταπτυχιακός φοιτητής του Τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Ιωαννίνων.

Τα ερευνητικά του ενδιαφέροντα εστιάζονται στην θεωρία πιθανοτήτων και στατιστικής, την μηχανική μάθηση, καθώς και την μαθηματική ανάλυση.

