



Technical  
University  
of Crete

# Στατιστική Μοντελοποίηση και Αναγνώριση Προτύπων (ΤΗΛ311)

## 1<sup>Η</sup> ΣΕΙΡΑ ΑΣΚΗΣΕΩΝ

ΒΙΤΤΗΣ ΒΑΣΙΛΗΣ 2015030164

### Θέμα 1: Principal Component Analysis (PCA)

#### 1.1 Εισαγωγή

Το Principal Component Analysis είναι μια στατιστική ανάλυση για dimensionality reduction δεδομένων. Το PCA είναι μια unsupervised μέθοδος και σκοπός μας είναι η εύρεση του συνόλου των κατάλληλων συνιστωσών οι οποίες καθώς θα περιστρέφονται και θα αλλάζουν μέτρο θα μας παρέχουν προβολές που μεγιστοποιούν την διασπορά των δεδομένων με γνώμονα και την αποφυγή απώλειας της πληροφορίας. Με απλή παρατήρηση το πρόβλημα της ελαχιστοποίησης της απόστασης του σημείου ενός στοιχείου από κάθε παραγόμενο ιδιοδιάνυσμα, μετατρέπεται σε μεγιστοποίηση της προβολής του διανύσματος του σημείου αυτού πάνω στην προέκταση του ιδιοδιανύσματος. Έτσι, ακολουθώντας τα παρακάτω βήματα πετυχαίνουμε την μείωση του διανύσματος χαρακτηριστικών με την ελάχιστη απώλεια πληροφορίας.

#### 1.2 Διαδικασία και Αποτελέσματα

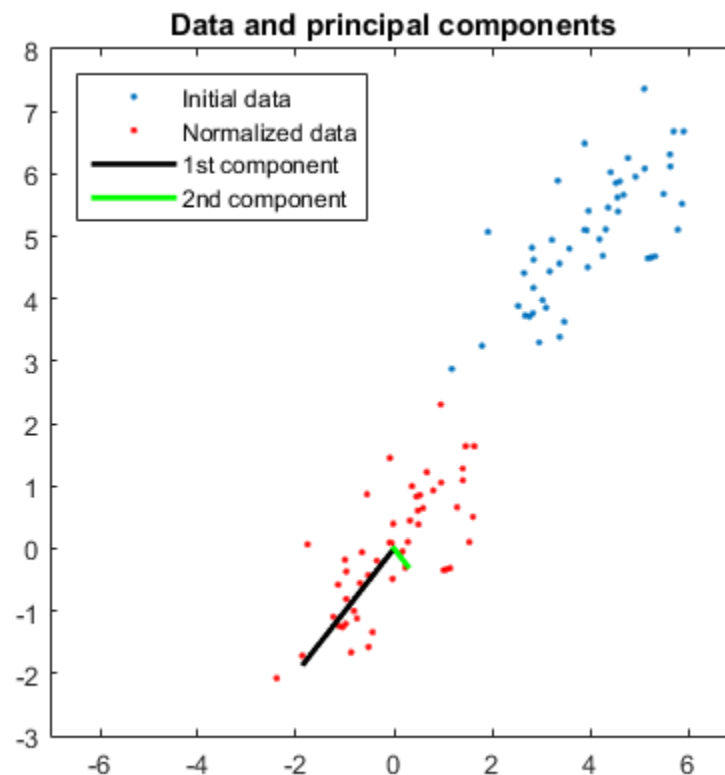
##### 1.2.1: PCA on 2D Samples

Στην παρακάτω άσκηση εφαρμόζονται τα κάτωθι βήματα: Αφαίρεση του μέσου όρου, υπολογισμός του πίνακα συν διασποράς, ιδιοδιανυσμάτων και ιδιοτιμών του πίνακα συν διασποράς και τέλος την ανάκτηση των δεδομένων.

Στο πλαίσιο του preprocessing και ύστερα από την λήψη των δεδομένων ξεκινάμε την διαδικασία του normalization και standardization καθώς θέλουμε τα δεδομένα μας να είναι όλα συγκεντρωμένα κοντά

σε μια τιμή (συγκεκριμένα στο 0) έτσι ώστε να αποφύγουμε μεγάλες δυσαναλογίες κατά την διαδικασία επεξεργασίας τους αλλά και απεικόνισής τους.

Έπειτα, υπολογίζουμε τον πίνακα συν διασποράς των δεδομένων μας και με βάση αυτό βρίσκουμε τις ιδιοτιμές και τα ιδιοδιανύσματα. Σκοπός μας όπως αναφέρθηκε και παραπάνω είναι η μείωση των χαρακτηριστικών άρα επιδιώκουμε να κρατήσουμε ένα υποσύνολο χαρακτηριστικών (συνιστώσες) με βάση την συνεισφορά τους στην συνολική διακύμανση και που θα μας βοηθήσουν στην καλύτερη προσέγγιση κατά την ανάκτηση των τιμών των αρχικών μας δεδομένων.

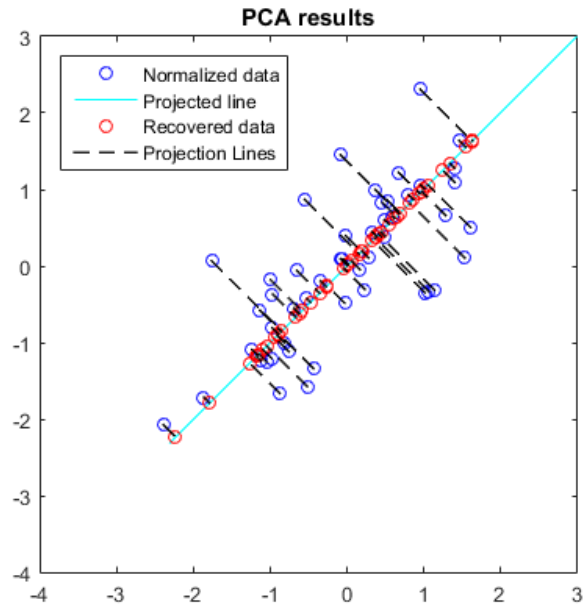


Εικόνα 1

Άρα, έχοντας ταξινομήσει τις ιδιοτιμές και τα ιδιοδιανύσματα των συνιστωσών και έχοντας κρατήσει τις συνιστώσες που μας βοηθάνε περισσότερο στον σκοπό μας συνεχίζουμε στην διαδικασία προβολής των δεδομένων στις λιγότερες διαστάσεις. Η διαδικασία της προβολής είναι απλή καθώς υλοποιεί μόνο το εσωτερικό γινόμενο μεταξύ των ισχυρών ιδιοδιανυσμάτων με τα αρχικά δεδομένα.

Τελευταίο βήμα για την ολοκλήρωση της διαδικασίας είναι η ανάκτηση των δεδομένων μας. Η διαδικασία αυτή είναι ανάλογη αυτής της προβολής και αποτελείται από το εσωτερικό γινόμενο των ισχυρών ιδιοδιανυσμάτων με τα προβαλλόμενα δεδομένα που υπολογίσαμε στο προηγούμενο βήμα.

Στην παρακάτω απεικόνιση παρατηρούμε με κυανό χρώμα την συνιστώσα με την μεγαλύτερη προσφερόμενη διακύμανση που πάνω σε αυτή θα προβληθούν τα αρχικά μας δεδομένα. Η κατεύθυνση της όπως παρατηρούμε είναι ταυτόσημη με την κατεύθυνση της παραπάνω εικόνας.



Εικόνα 2

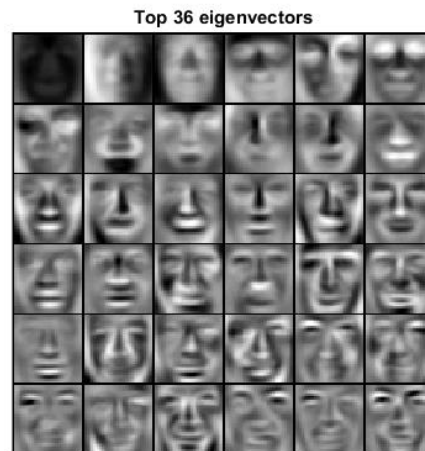
### 1.2.2: PCA on Face Data

Έχοντας κατανοήσει την συμπεριφορά της PCA στις 2 διαστάσεις, το βήμα προς την γενίκευση είναι απλό. Στην συγκεκριμένη περίπτωση δεν έχουμε να κάνουμε με περισσότερα δεδομένα και με περισσότερα χαρακτηριστικά. Έτσι, ακολουθώντας τα αντίστοιχα βήματα με το πρώτο μέρος της άσκησης, απεικονίζουμε τα δεδομένα μας.



Εικόνα 3

Στην συνέχεια εφαρμόσαμε normalization πάνω στα αρχικά μας δεδομένα και υπολογίσαμε τις κύριες συνιστώσες χρησιμοποιώντας την συνάρτηση PCA που υλοποιήσαμε στο πρώτο μέρος. Στην παρακάτω εικόνα παρατηρούμε πως διαμορφώνονται τα ιδιοδιανύσματα με βάση την ταξινόμηση των ιδιοτιμών τους. Επειδή το αντικείμενο μας είναι εικόνες αυτά τα 36 ιδιοδιανύσματα ,που αντιστοιχούν σε 36 διαφορετικά χαρακτηριστικά, είναι φωτεινότητα , βάθος χρώματος (άσπρου μαύρου) ,ένταση αλλαγής χρωμάτων, αντίθεση.



Εικόνα 3

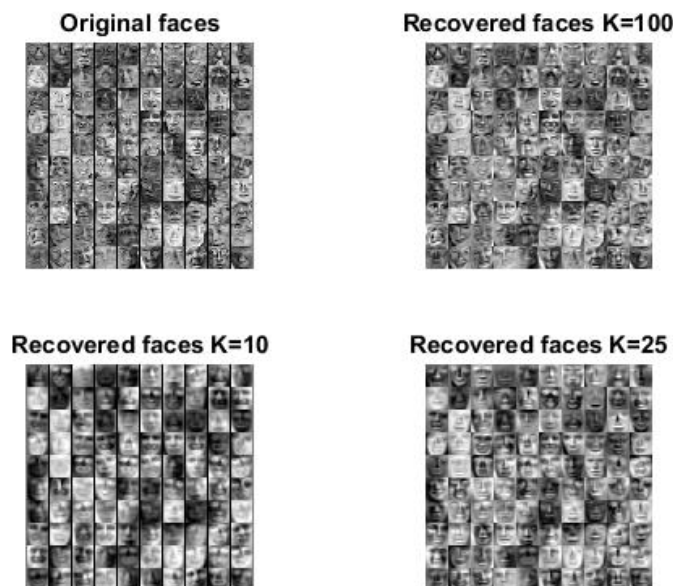
Το επόμενο βήμα μετά την εξαγωγή των ιδιοδιανυσμάτων και των ιδιοτιμών συνεχίσαμε όπως και στο πρώτο μέρος με την προβολή των δεδομένων μας πάνω στα ιδιοδιανύσματα αλλά αυτή την φορά το μέγεθος δεν είναι ένα αλλά 100. Άρα προβάλαμε τα δεδομένα μας πάνω σε 100 διαφορετικά ιδιοδιανύσματα.

Τέλος, έχοντας καταφέρει την μείωση των διαστάσεων από 5000 εικόνες με 1024 χαρακτηριστικά σε 5000 εικόνες με 100 χαρακτηριστικά, κάνουμε την ανάκτηση και την σύγκριση των τελικών δεδομένων με τα αρχικά για να δούμε οπτικά τουλάχιστον αν υπήρχε απώλεια πληροφορίας και πόσο αυτή επηρεάζει το τελικό αποτέλεσμα.

Στην Εικόνα 5 παρατηρούμε ότι με την περεταίρω μείωση των χαρακτηριστικών χάνουμε πολύτιμη πληροφορία που δεν μας ικανοποιεί κατά την διαδικασία της ανάκτησης. Όσο λιγότερα χαρακτηριστικά τόσο λιγότερη ικανότητα διαχωρισμού των δεδομένων.



Εικόνα 4



Εικόνα 5

### 1.3 Παρατηρήσεις

Με βάση όλα τα παραπάνω θα μπορούσε κάποιος να καταλήξει ότι η τεχνική Principal Component Analysis είναι καλή ως προς το να μειώνει τις διαστάσεις του διανύσματος χαρακτηριστικών με μεγάλη επιτυχία χωρίς να χάνει τόσο μεγάλη πληροφορία. Το παράδειγμα με τα πρόσωπα των ανθρώπων ισχυροποιεί την παραπάνω άποψη ακόμα περισσότερο καθώς δεν μπορούμε να διακρίνουμε μεγάλες αλλαγές μεταξύ των εικόνων με 1024 χαρακτηριστικά και με 100. Όμως ανάλυση αυτή, επειδή δεν λαμβάνει υπόψη της πολλές παραμέτρους, είναι κατάλληλη για συγκεκριμένο φάσμα εφαρμογών, ενώ αντίστοιχες μέθοδοι αποτελούν καλύτερη επιλογή καθώς προσφέρουν υψηλότερη απόδοση.

## Θέμα 2: LDA Classifier (Linear Discriminant Analysis)

Σκοπός αυτής της άσκησης είναι να υπολογιστεί αναλυτικά το διάνυσμα προβολής  $w$ . Γνωρίζουμε ότι το

$$w = S_W^{-1} (\mu_2 - \mu_1)$$

Άρα ξεκινώντας την ανάλυση μας υπολογίζουμε το within-class matrix ως εξής:

$$S_W = \sum_{i=1}^C P_i \cdot \Sigma_i$$

Αναλυτικά πίνακας του  $S_W$  προκύπτει ότι είναι:

$$S_W = \sum_{i=1}^C P_i \cdot \Sigma_i = \frac{1}{2} \begin{bmatrix} 11 & 9 \\ 9 & 11 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 13 & 9 \\ 9 & 13 \end{bmatrix}$$

Προκύπτει ότι ο αντίστροφος πίνακας του  $S_W$  χρησιμοποιώντας τον παρακάτω τύπο προκύπτει ότι είναι:

$$\left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right)^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \text{ και } (cA)^{-1} = c^{-1} \cdot A^{-1}$$

$$S_W^{-1} = \frac{1}{44} \begin{bmatrix} 13 & -9 \\ -9 & 13 \end{bmatrix}$$

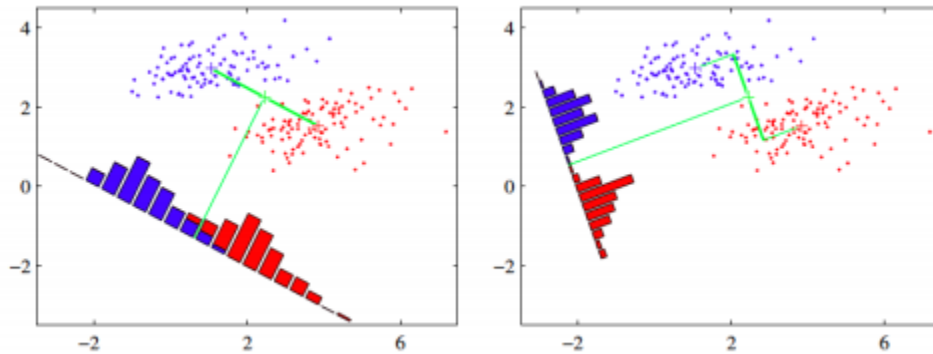
Άρα το διάνυσμα προβολής είναι:

$$w = \frac{1}{44} \begin{bmatrix} 13 & -9 \\ -9 & 13 \end{bmatrix} \left( \begin{bmatrix} -5 \\ 5 \end{bmatrix} - \begin{bmatrix} 10 \\ 15 \end{bmatrix} \right) = \frac{1}{44} \begin{bmatrix} 13 & -9 \\ -9 & 13 \end{bmatrix} \cdot \begin{bmatrix} -15 \\ -10 \end{bmatrix} = \frac{1}{44} \begin{bmatrix} -105 \\ 5 \end{bmatrix}$$

## Θέμα 3: Linear Discriminant Analysis (LDA) vs PCA

### 3.1 Εισαγωγή

Το Linear Discriminant Analysis είναι μια στατιστική ανάλυση για dimensionality reduction δεδομένων και για classification. Το LDA είναι μια supervised μέθοδος και σκοπός μας είναι η εύρεση του συνόλου των κατάλληλων συνιστωσών οι οποίες καθώς θα περιστρέφονται και θα αλλάζουν μέτρο θα μας παρέχουν προβολές που μεγιστοποιούν την κανονικοποιημένη διαφορά των μέσων των διαφορετικών κλάσεων χωρίς απώλεια της πληροφορίας. Η σχηματική απεικόνιση της διαδικασίας LDA γίνεται ξεκάθαρα με την παρακάτω εικόνα.

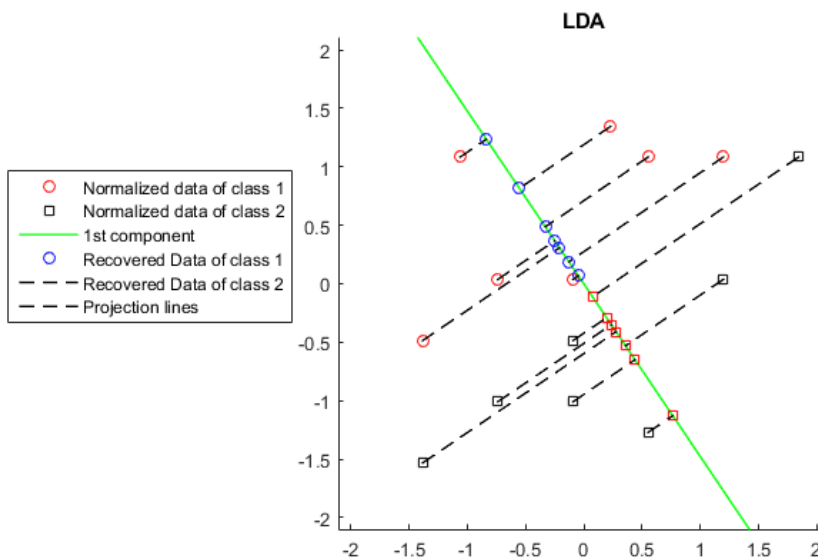


Εικόνα 6

Με την τεχνική LDA θέλουμε όσο το δυνατό απομακρυσμένες μέσες τιμές και όσο το δυνατό μικρότερες διασπορές. Το αριστερό μέρος παραπάνω εικόνα μας δείχνει ότι το ιδιοδιάνυσμα στο οποίο πάνω είναι προβαλλόμενα τα δεδομένα δεν κάνει καλό classification μεταξύ των κλάσεων, ενώ το δεξί μέρος μας δείχνει ότι αν περιστρέψουμε κατάλληλα το διάνυσμα προβολής τότε μπορούμε με μεγαλύτερη ακρίβεια πάνω στα testing data να αποφασίσουμε σε ποια κλάση ανήκει το στοιχείο.

### 3.2 Διαδικασία και Αποτελέσματα

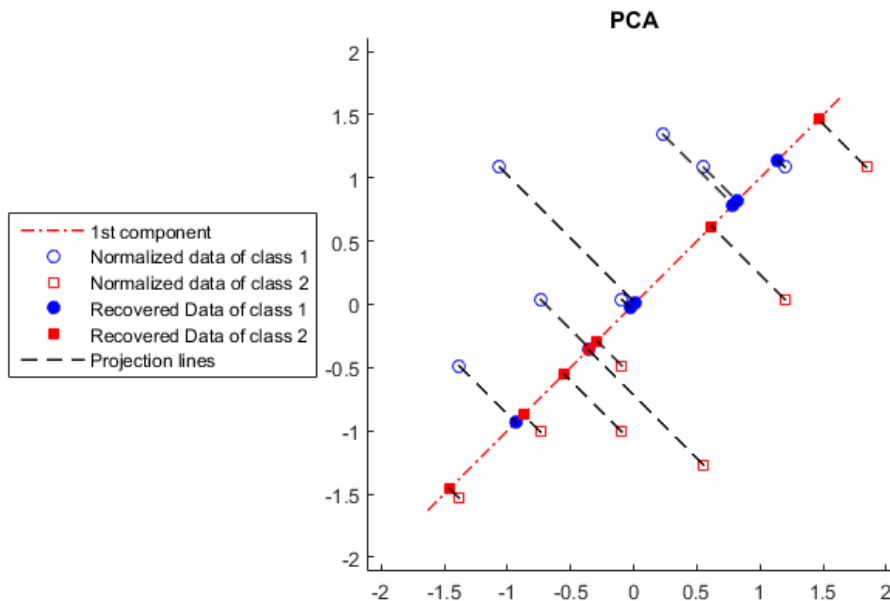
#### 3.2.1: LDA on 2D Samples With 2 Classes



Εικόνα 7

Στην παραπάνω εικόνα βλέπουμε ότι τα recovered data της μια κλάσης είναι ξεκάθαρα διαχωρισμένα από τα recovered data της άλλης κλάσης. Αυτό σημαίνει ότι κατά την διαδικασία του LDA επιλέχθηκε ένα ιδιοδιάνυσμα που είχε την μεγαλύτερη τιμή διαφοράς μέσων και μικρού αθροίσματος διασπορών. Αυτό το ιδιοδιάνυσμα παρατηρείται με την πράσινη γραμμή και πάνω σε αυτό βλέπουμε τα προβαλλόμενα στοιχεία των δυο κλάσεων.

Στην παρακάτω εικόνα παρατηρούμε την επίδραση του αλγόριθμου PCA πάνω στα ίδια δεδομένα με πριν. Αρχικά ο αλγόριθμος PCA εκ φύσεως του ομαδοποίησε τις δύο κλάσεις σε μια και προσπάθησε να βρει εκείνο το ιδιοδιάνυσμα που μεγιστοποιεί την διακύμανση των δεδομένων. Το PCA έκανε στο 100% την δουλεία του απλά δεν ήταν αυτό που αναζητούσαμε. Άρα το PCA για τον διαχωρισμό κλάσεων δεν αποτελεί καλή μέθοδος πλην εξαιρέσεων. Υπάρχουν περιπτώσεις όπου ο αλγόριθμος PCA μπορεί να βγάλει καλύτερο αποτέλεσμα από LDA αλλά αυτό εξαρτάται καθαρά από την φύση του προβλήματος και των δεδομένων του.



Εικόνα 8

### 3.2.2: Fisher Iris DATA

Έχοντας κατανοήσει την συμπεριφορά της LDA μεταξύ 2 κλάσεων στις 2 διαστάσεις, το βήμα προς την γενίκευση είναι εύκολο. Στην συγκεκριμένη περίπτωση έχουμε να κάνουμε όχι με περισσότερα δεδομένα αλλά και με περισσότερα χαρακτηριστικά. Έτσι, ακολουθώντας τα αντίστοιχα βήματα με το πρώτο μέρος της άσκησης, απεικονίζουμε τα δεδομένα μας.





Εικόνα 9

Το συγκεκριμένο παράδειγμα όπως και φαίνεται παραπάνω έχει τρεις διαφορετικές κλάσεις. Στο παραπάνω διάγραμμα τα στοιχεία είναι τοποθετημένα στον χώρο με βάση τα δύο από αυτά. Σκοπός μας είναι αρχικά να μειώσουμε τις διαστάσεις από έξι σε δυο ψάχνοντας χαρακτηριστικά που δεν έχουν μεγάλη συσχέτιση μεταξύ τους και είναι ικανά να διαχωρίσουν τις κλάσεις, που όπως φαίνεται με βάση την παραπάνω εικόνα τα δεδομένα είναι ανακατεμένα και τα δυο χαρακτηριστικά αυτά έχουν αρκετή συσχέτιση μεταξύ τους.

Τα βήματα που ακολουθήσαμε για την υλοποίηση της συνάρτησης LDA είναι τα εξής: Υπολογίζουμε τις εκ των προτέρων πιθανότητες των κλάσεων. Υπολογίζουμε το μέσο κάθε κλάσης και ύστερα το γενικό μέσο μεταξύ των μέσων κάθε κλάσης. Υπολογίζουμε τους πίνακες διασποράς κάθε κλάσης και τον πίνακα συν διασποράς των κλάσεων μεταξύ τους. Τέλος, εξάγουμε τα ιδιοδιανύσματα και τις ιδιοτιμές και κρατάμε τα ιδιοδιανύσματα με βάση την επιθυμητή διάσταση που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές.



Εικόνα 10

## Θέμα 4: Bayes

### 4.1 Διαδικασία

Γνωρίζουμε ότι η Gaussian κατανομή δεδομένης μιας κλάσης  $i$  είναι

$$p(x|\omega_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{-\frac{1}{2}}} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

Άρα για να βρούμε το σύνορο διαχωρισμού μεταξύ των δύο κλάσεων  $\omega_1$  και  $\omega_2$  πρέπει να λύσουμε την ισότητα:

$$p(x|\omega_1) p(\omega_1) = p(x|\omega_2) p(\omega_2) \Rightarrow \quad (1)$$

$$p(\omega_2) \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_2|^{-\frac{1}{2}}} \cdot e^{-\frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2)} = p(\omega_1) \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_1|^{-\frac{1}{2}}} \cdot e^{-\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)} \Rightarrow$$

Όπου  $d = 2$  και  $|\Sigma_1| = |\Sigma_2|$

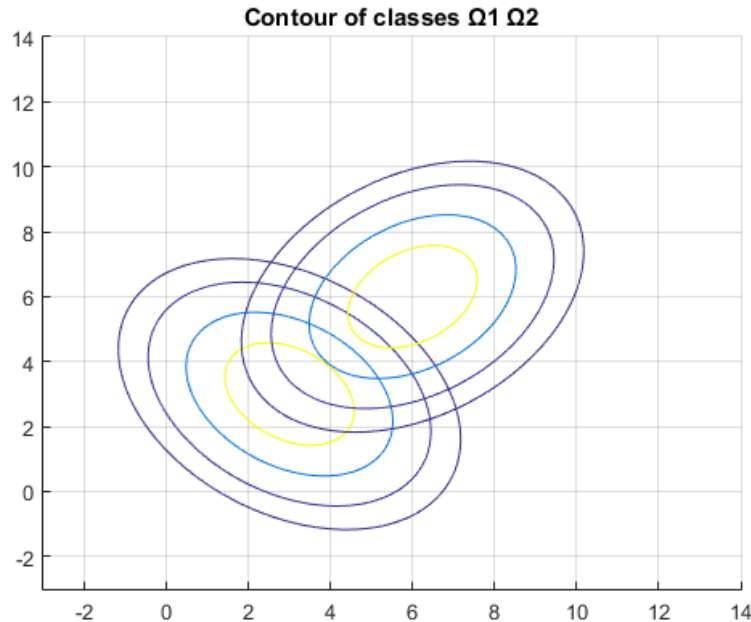
$$p(\omega_2) e^{-\frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2)} = p(\omega_1) e^{-\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)} \Rightarrow$$

$$\ln(p(\omega_2)) + \ln\left(e^{-\frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2)}\right) = \ln(p(\omega_1)) + \ln\left(e^{-\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)}\right) \Rightarrow$$

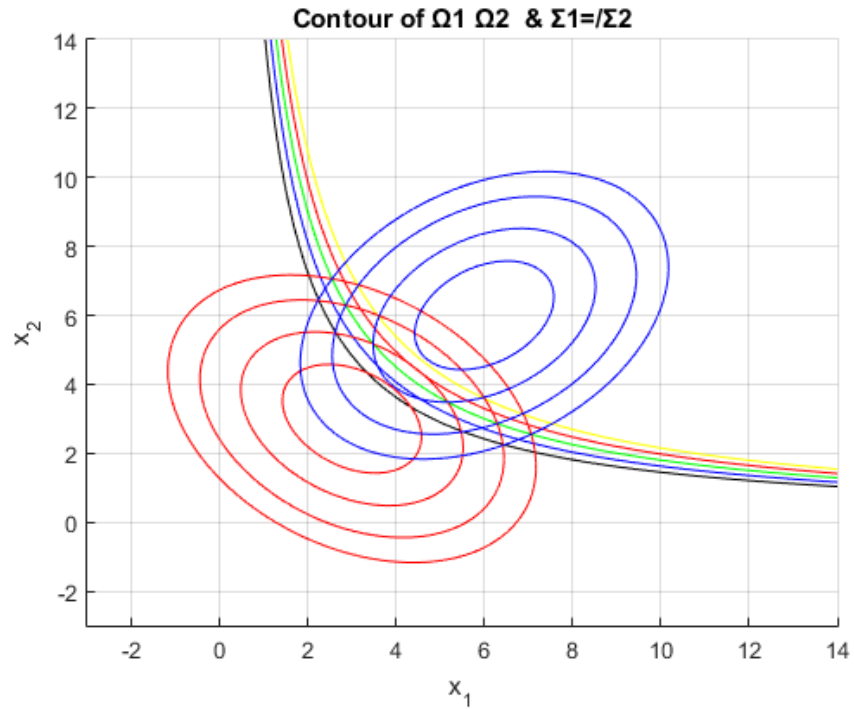
$$-\frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2) + \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) = \ln\left(\frac{p(\omega_1)}{p(\omega_2)}\right) \Rightarrow$$

#### 4.2 Παρατηρήσεις

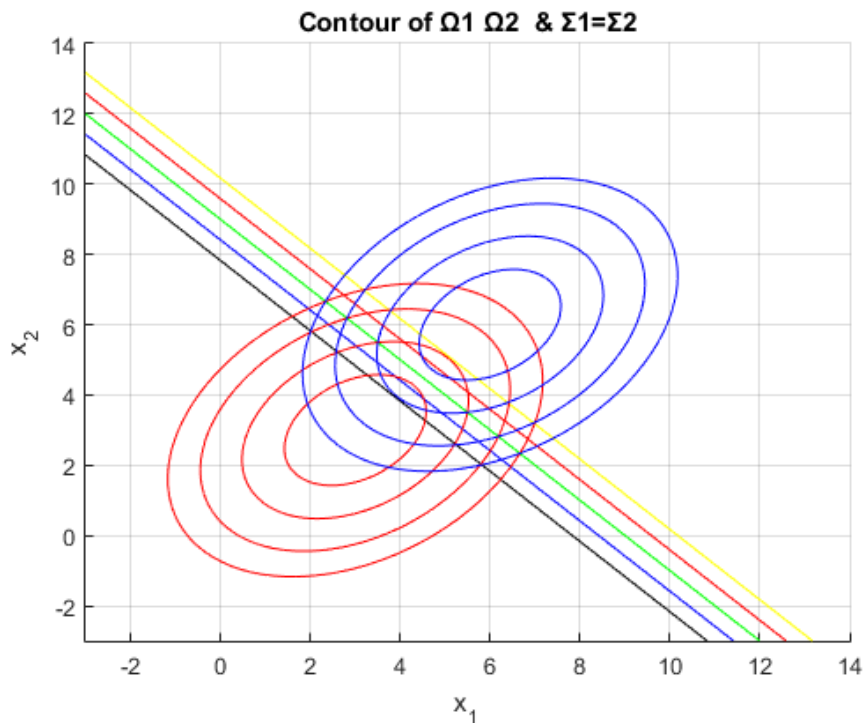
Με βάση την παραπάνω ανάλυση προκύπτουν οι τομές της καμπάνας στα διάφορα ύψη που ορίζονται από τις πιθανότητες δημιουργώντας ομόκεντρες ελλείψεις. Η διεύθυνση τους ορίζεται από τα μη διαγώνια στοιχεία του πίνακα της συν διασποράς που επί της ουσίας περιστρέφουν τις ελλείψεις κατά κάποιες μοίρες. Ενώ στο σχήμα τους επηρεάζουν και τα διαγώνια στοιχεία του πίνακα συν διασποράς που. Παρακάτω βλέπουμε τις τομές των κλάσεων με βάση διάφορες πιθανότητες.



Υποθέτοντας τώρα ότι ισχύει  $P(\omega_2) = 1 - P(\omega_1)$  και με βάση την ανάλυση που κάναμε παραπάνω μπορούμε να σχεδιάσουμε τα σύνορα απόφασης για διάφορες τιμές των εκ των προτέρων πιθανοτήτων  $P(\omega_2)$  &  $P(\omega_1)$ . Οι κόκκινες ελλείψεις είναι αποτέλεσμα της κλάσης  $\omega_1$  και οι μπλε της κλάσης  $\omega_2$ . Ο γενικός κανόνας που παρατηρείται μέσω της παρακάτω εικόνας είναι ότι όταν η εκ των προτέρων πιθανότητα μια κλάσης είναι μεγαλύτερη από μια άλλη τότε το όριο απόφασης μετακινείται μακριά από αυτή δηλαδή αποκτά περισσότερο χώρο και άρα έχει αυξημένη πιθανότητα ένα καινούργιο δεδομένο να κατηγοριοποιηθεί με βάση εκείνη.



Επαναλαμβάνοντας την διαδικασία για πίνακες συνδιασποράς αυτή την φορά ίσους μεταξύ τους. Παρατηρούμε ότι οι τομές των κατανομών από ελλείψεις έγιναν κύκλοι και τα όρια απόφασης από παραβολές έγιναν ευθείες, αυτό είναι αναμενόμενο καθώς κατά την επίλυση της ισότητας (1) από ένα σύστημα 2<sup>ης</sup> τάξης μεταφερόμαστε σε ένα γραμμικό σύστημα 1<sup>ης</sup> καθώς υπάρχουν απλοποιήσεις. Ο κανόνας, τον οποίο παρατηρήσαμε και παραπάνω ισχύει και σε αυτή την περίπτωση.



## Θέμα 5: Minimum Risk

Υπολογίζουμε το όριο απόφασης  $x_0$  που έχει το μικρότερο ρίσκο δεδομένου ποιων λάθους απόφασης δεδομένου ισοπίθανων κλάσεων.

$$l1 = \lambda_{11}P(\omega_1)P(x|\omega_1) + \lambda_{21}P(\omega_2)P(x|\omega_2) = \lambda_{21}P(\omega_2)P(x|\omega_2)$$

$$l2 = \lambda_{12}P(\omega_1)P(x|\omega_1) + \lambda_{22}P(\omega_2)P(x|\omega_2) = \lambda_{12}P(\omega_1)P(x|\omega_1)$$

$$l1 = l2 \Rightarrow \lambda_{21}P(\omega_2)P(x|\omega_2) = \lambda_{12}P(\omega_1)P(x|\omega_1) \Rightarrow$$

$$\lambda_{21}P(x|\omega_2) = \lambda_{12}P(x|\omega_1) \Rightarrow$$

$$\ln(P(x|\omega_2)) - \ln(P(x|\omega_1)) = \ln\left(\frac{\lambda_{12}}{\lambda_{21}}\right) \Rightarrow$$

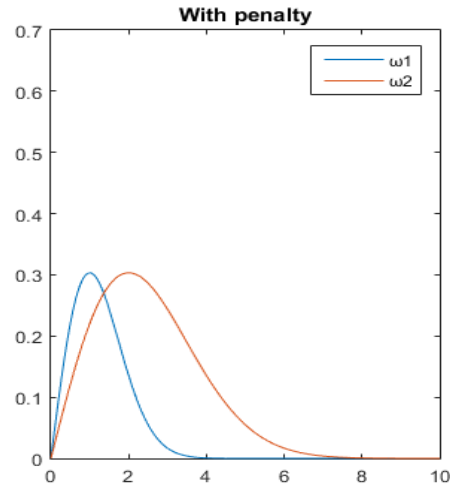
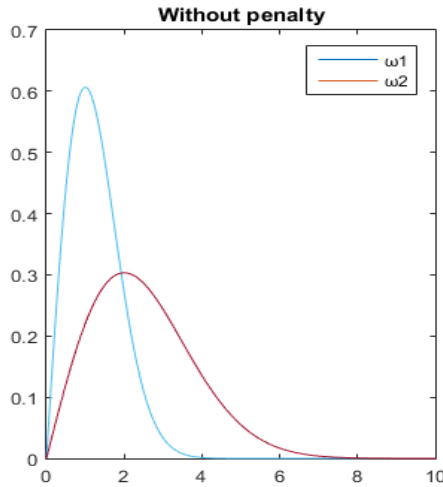
$$\ln\left(\frac{x}{\sigma_2^2}\right) - \frac{x^2}{2\sigma_2^2} - \ln\left(\frac{x}{\sigma_1^2}\right) + \frac{x^2}{2\sigma_1^2} = \ln\left(\frac{\lambda_{12}}{\lambda_{21}}\right) \Rightarrow$$

$$x^2\left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\right) = \ln\left(\frac{\lambda_{12}}{\lambda_{21}}\right) - \ln\left(\frac{\sigma_1^2}{\sigma_2^2}\right) \Rightarrow$$

$$x^2\left(\frac{2\sigma_2^2 - 2\sigma_1^2}{2\sigma_1^2\sigma_2^2}\right) = \ln\left(\frac{\lambda_{12}\sigma_2^2}{\lambda_{21}\sigma_1^2}\right) \Rightarrow$$

$$x^2\left(\frac{3}{8}\right) = \ln(2) \Rightarrow$$

$$x = \sqrt{1.848} = 1.359$$



Στο αριστερό μέρος της παραπάνω εικόνας παρατηρούμε ότι στην περίπτωση που δεν έχουμε ποινή το όριο απόφασης (δηλαδή σημείο τομής των δυο καμπυλών) βρίσκεται στο σημείο (1.92,0.3028). Όμως, με την εισαγωγή της ποινής στις δύο κλάσεις το όριο απόφασης μετακινήθηκε κατά αριστερά. Αυτό σημαίνει ότι η κλάση  $\omega_1$  παίρνει μικρότερο χώρο άρα το να αποφασίσω αυτή αντί της άλλης είναι λιγότερο πιθανό.