ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

Σχολή ΗΜΜΥ

Ομάδα 5 – Αγγελόπουλος Δημήτριος 2020030038

Στατιστική Μοντελοποίηση και Αναγνώριση Προτύπων (ΤΗΛ311)

1η Σειρά Ασκήσεων

Θέμα 1

Σε αυτό το θέμα θα κατηγοριοποιήσουμε 2D δείγματα $\underline{x}$ σε δύο κλάσεις ω_1 και ω_2. Οι κλάσεις περιγράφονται από τις ακόλουθες κατανομές

$$p(\underline{x}|\omega_1) \sim \mathcal{N}(\mu_1, \Sigma_1), \qquad p(\underline{x}|\omega_2) \sim \mathcal{N}(\mu_2, \Sigma_2)$$

Αρχικά θα βρούμε ένα όριο απόφασης με βάση τις posterior πιθανότητες των κλάσεων.

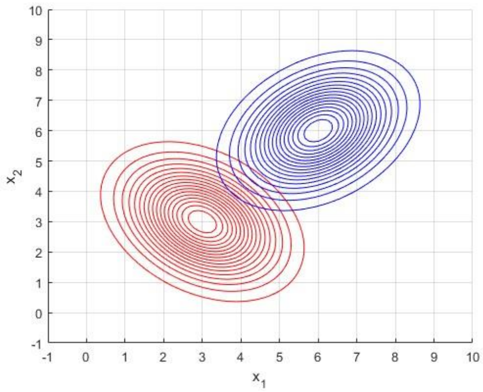
$$P(\omega_1|\underline{x}) = P(\omega_2|\underline{x}) <=> P(\omega_1)p(\underline{x}|\omega_1) = P(\omega_2)p(\underline{x}|\omega_2)$$

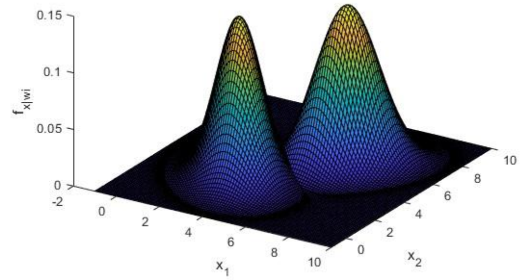
Μετά από πράξεις καταλήγουμε στο εξής όριο απόφασης :

$$x_1x_2 + 1.6\ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right) - 18 = 0$$

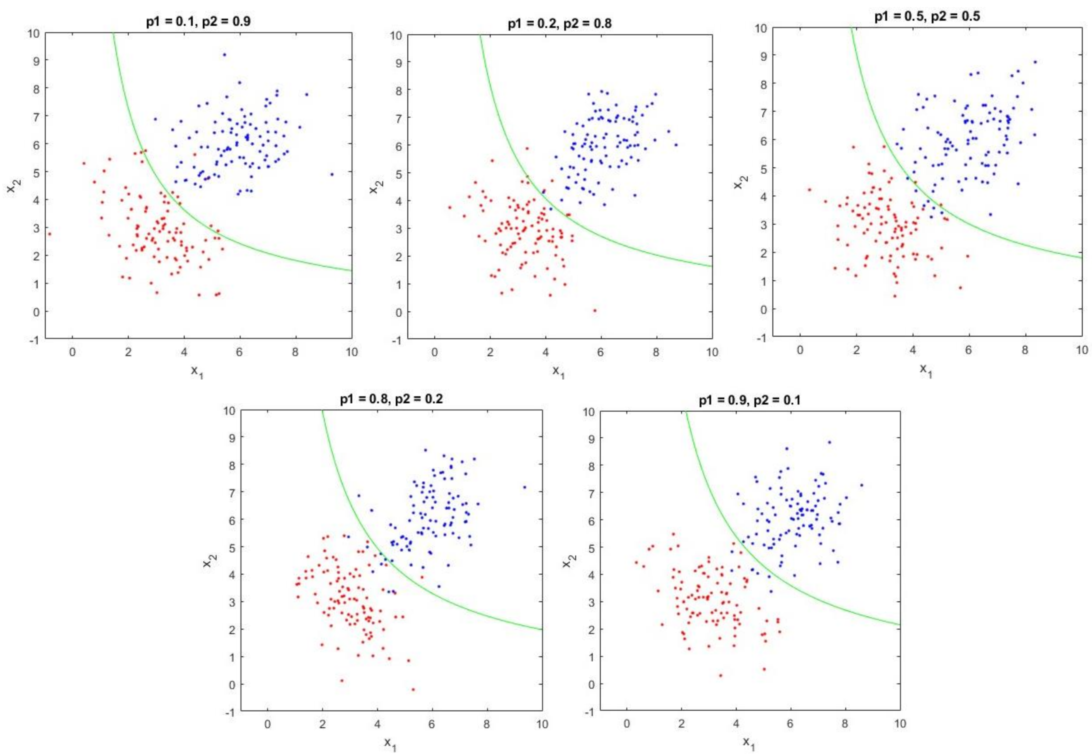
Το σύνορο περιγράφει μία υπερβολή και για τυχαίο δείγμα $\underline{x}$ αν το παραπάνω όριο είναι μικρότερο του μηδενός τότε το δείγμα προήλθε από την κλάση ω_1 αλλιώς ανήκει στην ω_2. Το γεγονός ότι προκύπτει υπερβολή ευθύνεται στο ότι οι πίνακες συνδιασποράς είναι διαφορετικοί μεταξύ τους (αυτό προκύπτει από τις πράξεις που επισυνάπτονται σε ξεχωριστό pdf).

Στην συνέχεια θα σχεδιάσουμε σε κοινό plot μερικές ισοϋψείς καμπύλες για κάθε κατανομή αλλά και σε ένα άλλο τις αντίστοιχες κατανομές στον χώρο.



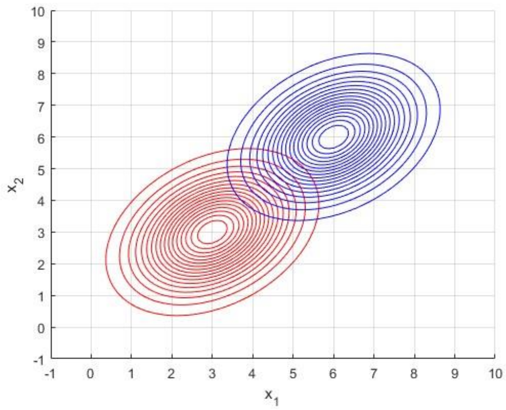


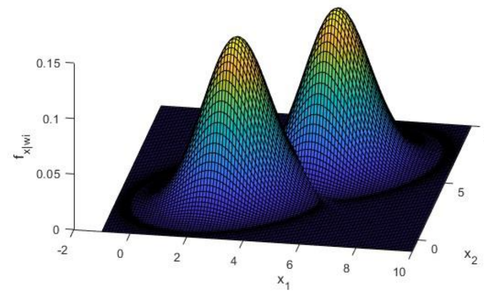
Τώρα θα δημιουργήσουμε 200 τυχαία δείγματα, 100 από την κάθε κατανομή και θα σχεδιάσουμε τα όρια απόφασης για τις διάφορες prior πιθανότητες.



Τα όρια απόφασης είναι υπερβολές που εξαρτώνται από τις εκ των προτέρων πιθανότητες των κλάσεων. Επομένως όσο αυξάνεται η prior πιθανότητα της ω_1 το όριο απομακρύνεται από την μέση τιμή της pdf της πρώτης κλάσης.

Επαναλαμβάνουμε όλα τα παραπάνω για ίδιους πίνακες συνδιασποράς και έχουμε

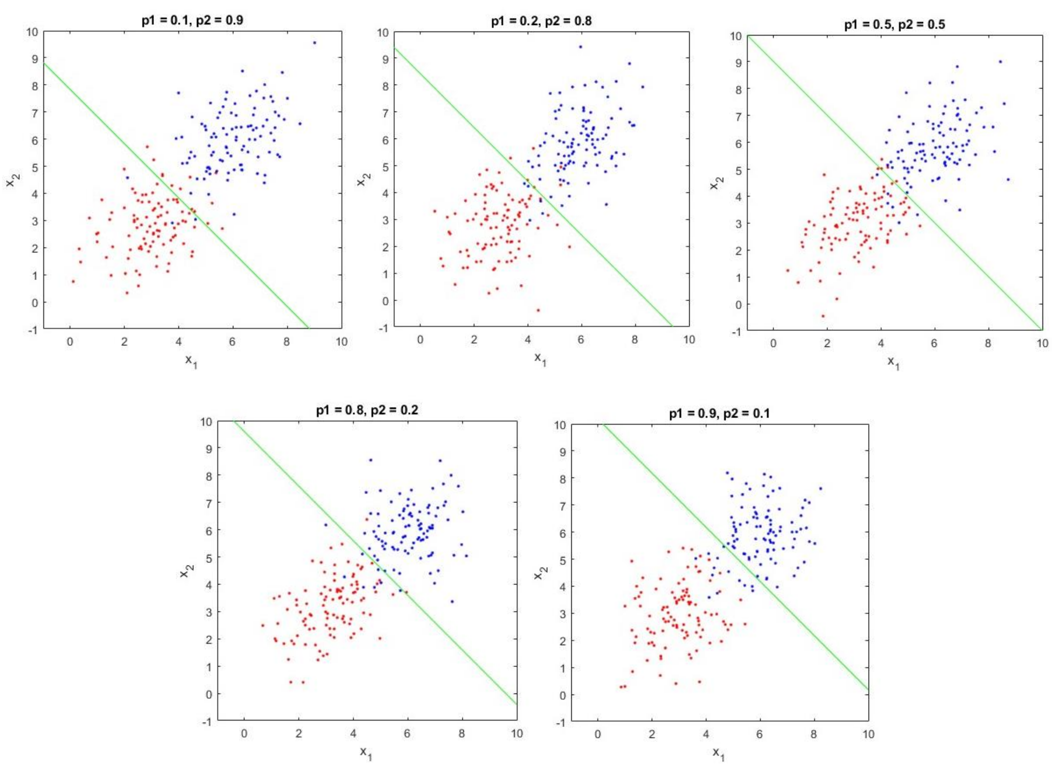




Μετά από πράξεις βρίσκουμε το εξής όριο απόφασης

$$x_1 + x_2 = 9 - 0.533\ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

Το οποίο είναι εξίσωση ευθείας. Επομένως πράγματι για ίσους πίνακες συνδιασποράς το σύνορο απόφασης παύει να είναι καμπύλη.



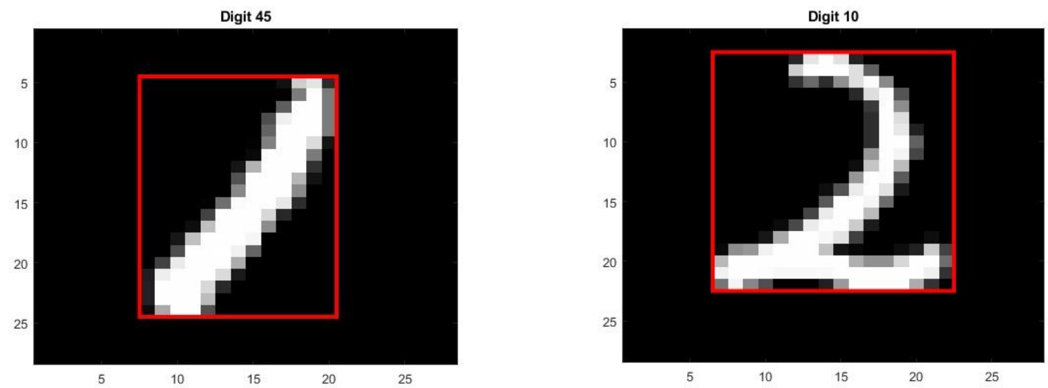
Θέμα 2

Σε αυτό το θέμα ο σκοπός είναι να δημιουργήσουμε έναν ταξινομητή Bayes ο οποίος θα ταξινομεί τα δείγματα στις κλάσεις C1 και C2. Η ταξινόμηση θα γίνει με βάσει το aspect ratio κάθε ψηφίου.

Αρχικά θα δημιουργήσουμε κώδικα που θα υπολογίζει τον λόγο όψεως για κάθε στοιχείο και θα βρίσκει την μέγιστη και ελάχιστη τιμή αυτών.

$$minAspectRatio = 0.1 \quad maxAspectRatio = 2.22$$

Ενώ στην συνέχεια θα σχεδιάσουμε δύο τυχαία δείγματα από το train set μαζί με ένα παραλληλόγραμμο στα όρια του λόγου όψεως.



Αφού ορίσαμε την τυχαία μεταβλητή, πριν αρχίσουμε το classification θα υπολογίσουμε τις prior πιθανότητες κάθε κλάσης.

$$P(C_1) = 0.53, P(C_2) = 0.47$$

Υποθέτουμε ότι η κατανομή κάθε χαρακτηριστικού, δεδομένου μίας κλάσης, ακολουθεί κανονική κατανομή με μέση τιμή και διασπορά :

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} x_j \ \& \ \sigma_i^2 = \frac{1}{N}\sum_{j=1}^{N}\left(x_j - \mu_j\right)^2$$

Έτσι υπολογίζουμε ότι $\mu_1 = 0.467,\ \sigma_1^2 = 0.0399$ και $\mu_2 = 0.96,\ \sigma_2^2 = 0.0355$ με βάση τα δείγματα εκπαίδευσης.

Στην συνέχεια υπολογίζουμε το aspect ratio για κάθε δείγμα των δεδομένων ελέγχου και με βάση την posterior πιθανότητα το ταξινομούμε στην αντίστοιχη κλάση. Ταυτόχρονα ελέγχουμε αν η ταξινόμηση έχει γίνει σωστά και υπολογίζουμε το συνολικό classification error.

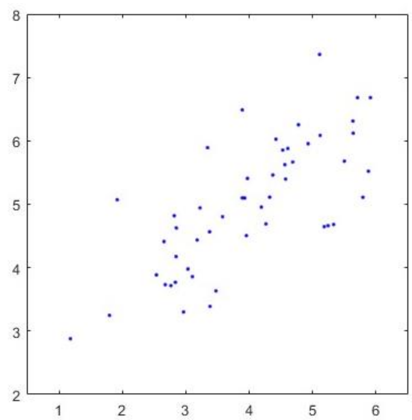
$$Error = 10.93\ \%$$

<u>Θέμα 5</u>

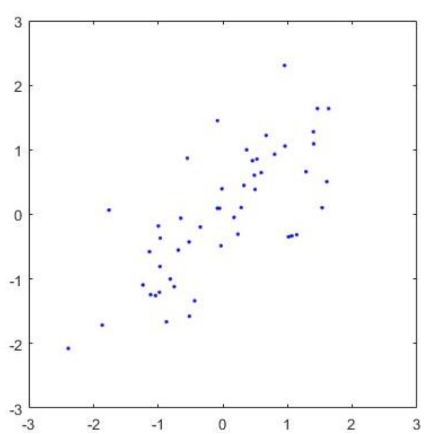
Σε αυτό το θέμα θα εφαρμόσουμε την μέθοδο Principal Components Analysis για να κάνουμε Dimensionality Reduction αρχικά σε 2D data και στην συνέχεια σε ένα dataset εικόνων προσώπων με πολύ μεγάλο αριθμό features.

Μέρος 1

Αρχικά θα εφαρμόσουμε PCA σε 2D data όπως φαίνονται παρακάτω.

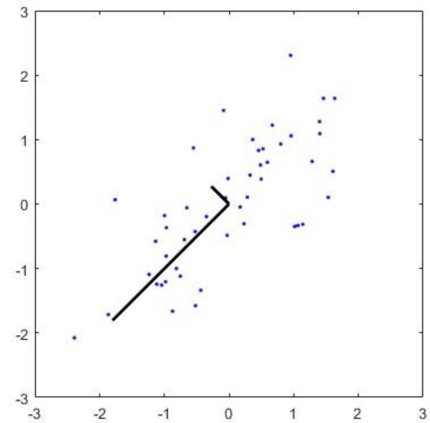


Κάνουμε standardization των data αφαιρώντας αρχικά την μέση τιμή και διαιρώντας με την τυπική απόκλιση και παίρνουμε:



Παρατηρούμε ότι δεν άλλαξε η σχετική θέση μεταξύ των δειγμάτων.

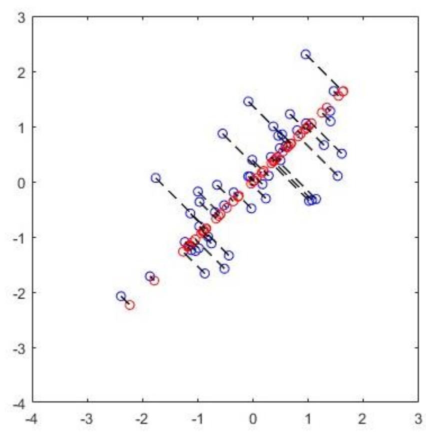
Θα υπολογίσουμε τις κύριες συνιστώσες του αλγορίθμου PCA. Αρχικά υπολογίζουμε τον πίνακα συνδιασποράς. Ύστερα υπολογίζουμε τις ιδιοτιμές του και σχεδιάζουμε στα κανονικοποιημένα δείγματα τα ιδιοδιανύσματα που αντιστοιχούν σε κάθε ιδιοτιμή.



Τώρα θα εφαρμόσουμε μείωση διάστασης ως προς το ιδιοδιάνυσμα που αντιστοιχεί στην μεγαλύτερη ιδιοτιμή. Δηλαδή θα πάρουμε δείγματα της μορφής

$z = w^T x$, όπου το w είναι το ιδιοδιάνυσμα αυτό.

Προβάλλοντας τα δεδομένα μας πάνω σε αυτό έχουμε :



Και τα z είναι τα δείγματα μας σε 1D.

Στην συνέχεια επιθυμούμε να ανακτήσουμε το μειωμένης διάστασης x και επομένως παίρνουμε μία εκτίμηση του με βάση μόνο την κύρια ιδιοτιμή :

$x_{rec} = zw$

Μέρος 2

Επαναλαμβάνουμε την παραπάνω διαδικασία σε δείγματα 5000 εικόνων προσώπων.

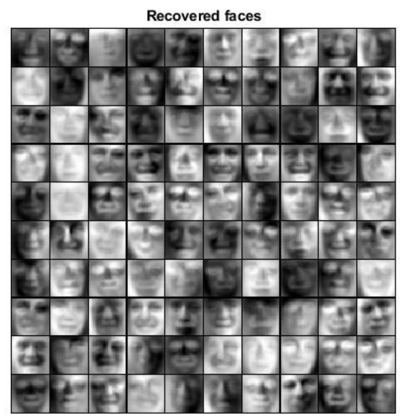
Τυπώνουμε τα 100 πρώτα πρώτα πρόσωπα από το dataset και έχουμε :

Αφού εφαρμόσουμε standardization κάνουμε PCA και τυπώνουμε τα 36 κυριότερα χαρακτηριστικά των προσώπων. Αυτά τα 36 χαρακτηριστικά είναι τα αντίστοιχα ιδιοδιανύσματα των 36 κύριων ιδιοτιμών του πίνακα συνδιασποράς και περιέχουν τις αριθμητικές τιμές των 1024x1024 pixel εικόνων.

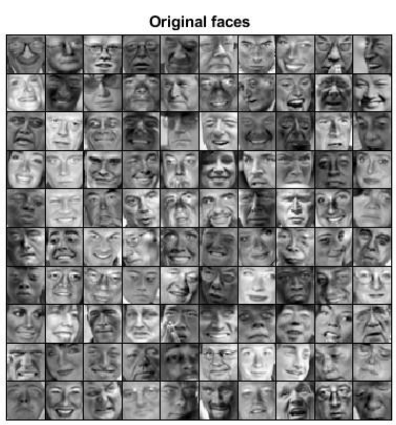
Παρατηρούμε ότι κάθε εικόνα έχει πληροφορία για μερικά βασικά χαρακτηριστικά των προσώπων. Για παράδειγμα η 3^{η} εικόνα φαίνεται να περιγράφει το σχήμα των προσώπων ενώ η 14^{η} και η 15^{η} περιγράφουν τα left και right edges των μυτών. Επειδή τα πρόσωπα είναι πολυδιάστατα δεδομένα έχουν πολλές κύριες συνιστώσες που είναι απαραίτητες για την περιγραφή τους.

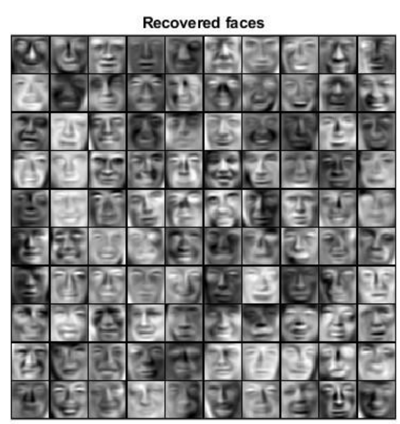
Θα επαληθεύσουμε αυτό το συμπέρασμα εφαρμόζοντας dimensionality reduction στις K κύριες συνιστώσες και στην συνέχεια εφαρμόζοντας ανάκτηση αυτών.

- ➔ K = 10

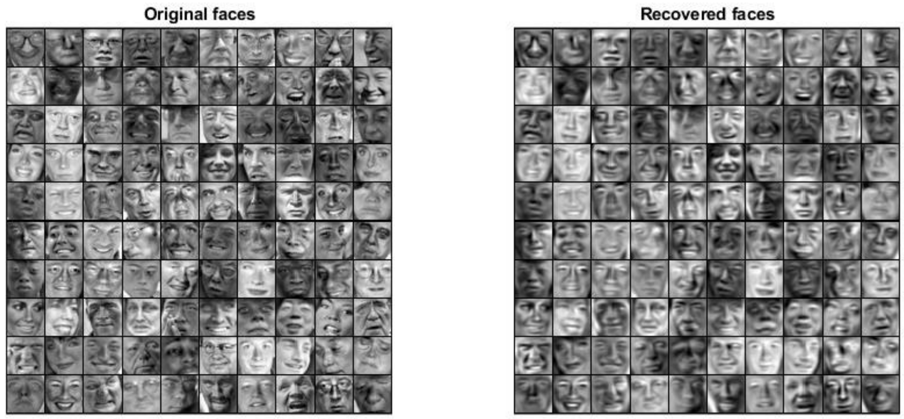


- ➔ K = 50

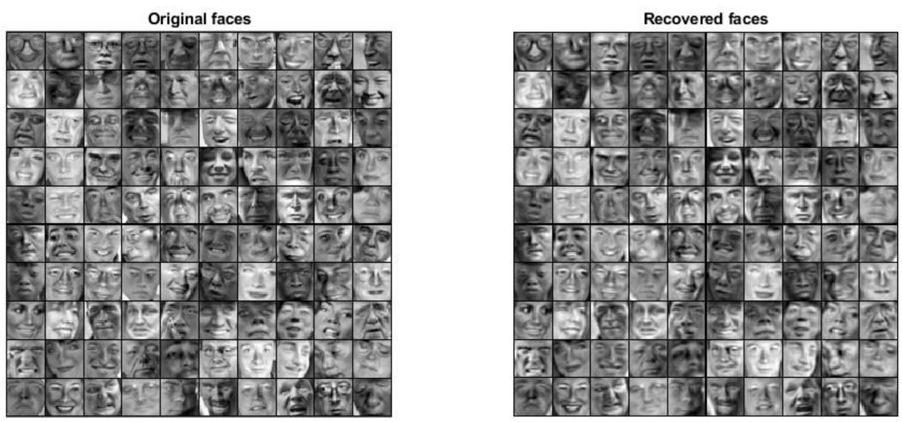




➔ K = 100



➔ K = 200



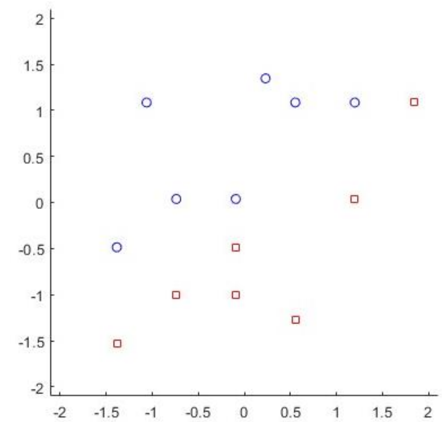
Πράγματι όπως περιμέναμε για μείωση διάστασης με βάση των πρώτων 10 συνιστωσών τα recovered πρόσωπα προβάλουν μόνο μερικά βασικά χαρακτηριστικά. Όσο αυξάνεται το K οι recovered εικόνες πλησιάζουν όλο και περισσότερο τις αρχικές μέχρι που είναι αδύνατο να καταλάβουμε ότι αυτές είναι διαφορετικές μεταξύ τους.

<u>Θέμα 5</u>

Σε αυτήν την άσκηση θα εφαρμόσουμε έναν νέο αλγόριθμο μείωσης διάστασης, τον αλγόριθμο του Fisher (LDA). Αυτή η μέθοδος χρησιμοποιεί παραπάνω πληροφορία των data και μεγιστοποιεί μία συνάρτηση κόστους με βάση αυτά. Στην συνέχεια θα τον συγκρίνουμε με τον PCA.

Μέρος 1

Αρχικά εφαρμόζουμε τον αλγόριθμο σε 2D data, αφού πρώτα τα έχουμε κανονικοποιήσει όπως κάναμε και στο Θέμα 5.

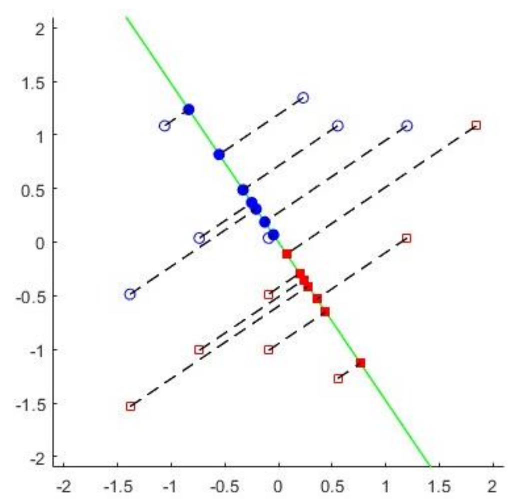


Για 2D data που θέλουμε να ταξινομήσουμε σε δύο κλάσεις ο βέλτιστος κανόνας με βάση το LDA είναι :

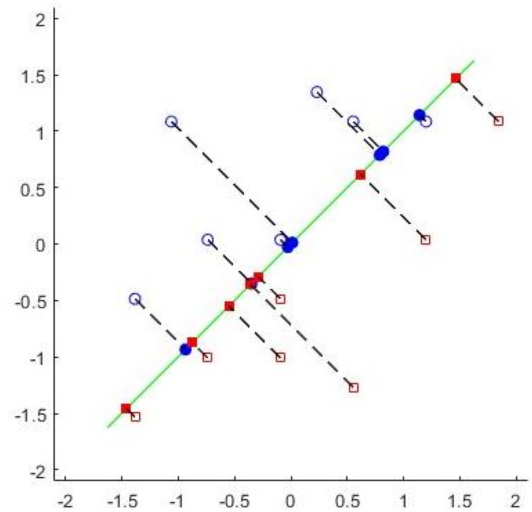
$$z = w^T x, \quad w = \beta(\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2)$$

Για κάθε β το αποτέλεσμα της προβολής είναι ίδιο.

Προβάλουμε τα δεδομένα μας πάνω στην ευθεία που ορίζει το w και έχουμε :



Στην συνέχεια προβάλουμε τα δεδομένα μας χρησιμοποιώντας τον αλγόριθμο PCA και έχουμε :

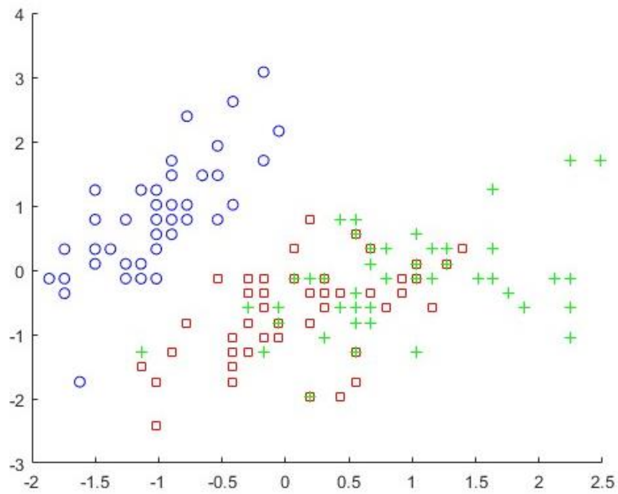


Παρατηρούμε ότι χρησιμοποιώντας τον LDA η προβολή των δεδομένων στον μονοδιάστατο χώρο είναι βέλτιστη όσον αφορά την διαφορά μεταξύ των μέσων τιμών και της διασποράς των δεδομένων μας ανά κλάση. Ενώ αγνοώντας την πληροφορία που μας δίνουν οι κλάσεις, δηλαδή εφαρμόζοντας PCA τα δεδομένα στον μονοδιάστατο χώρο είναι μπλεγμένα μεταξύ τους και άρα θα περιμένουμε ότι η ταξινόμηση θα έχει πολύ μεγάλο σφάλμα.

Μέρος 2

Σε αυτό το μέρος θα εφαρμόσουμε LDA σε ένα πολύ γνωστό dataset στον χώρο της Μηχανικής Μάθησης, του database Iris.

Αρχικά θα σχεδιάσουμε τα 2 πρώτα features κάθε κλάσης, αφού πρώτα εφαρμόσουμε standardization και θα έχουμε :



Για να εφαρμόσουμε τον αλγόριθμο του Fisher στην συνάρτηση myLDA θα πρέπει να υπολογίσουμε τις εκ των προτέρων πιθανότητες των κλάσεων, τις μέσες τιμές τους, το ολικό μέσο και τους πίνακες σκέδασης.

- ➔ Within-class scatter Matrix

$$S_w = \sum_{i=1}^{c} P_i \Sigma_i$$

- ➔ Within-class scatter Matrix

$$S_b = \sum_{i=1}^{c} P_i (\mu_i - \mu_0)(\mu_i - \mu_0)^T$$

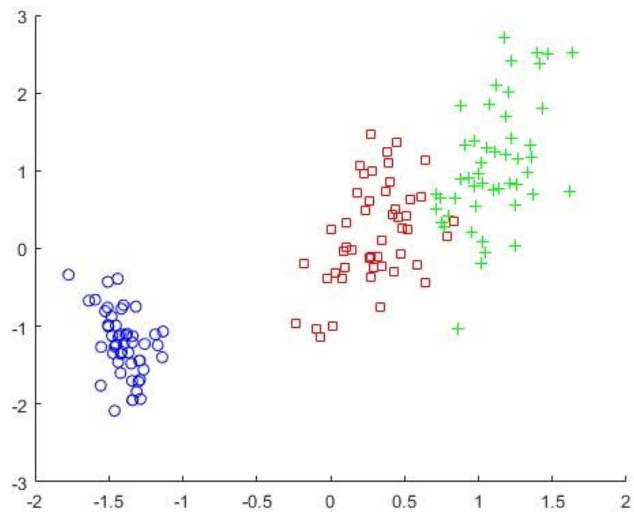
- ➔ Global mean

$$\mu_0 = \frac{1}{N} \sum_{i=1}^{c} n_i \mu_i$$

Όπου c ο αριθμός των κλάσεων και μ_0 το ολικό μέσο.

Αφού συμπληρώσουμε τον κώδικα και εφαρμόσουμε την μείωση διάστασης $z = W^T x$, όπου ο W θα περιέχει τα d μεγαλύτερα ιδιοδιανύσματα του $S_w^{-1} S_b$, θα έχουμε dimensionality reduction από 4 σε 2 διαστάσεις.

Τυπώνουμε τα αποτελέσματα και έχουμε :



Πράγματι ο αλγόριθμος έχει προβάλει τα δεδομένα μας με απομακρυσμένες μέσες τιμές και μικρή διασπορά. Επομένως η ταξινόμηση θα έχει μικρό σφάλμα.