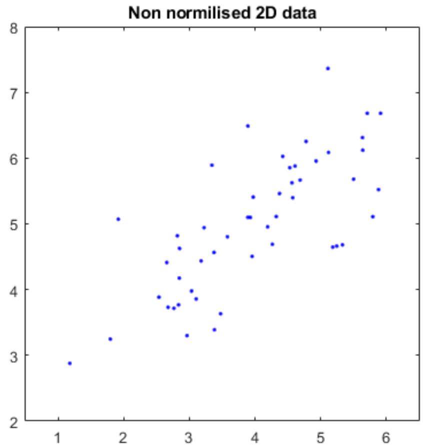
Κατσούπης
Ευάγγελος
2017030077

1^{η}Principal Component Analysis (PCA)

Αρχικά διαβάστηκαν τα παρακάτω δείγματα. Τα δείγματα κανονικοποιήθηκαν με μέση τιμή μηδέν

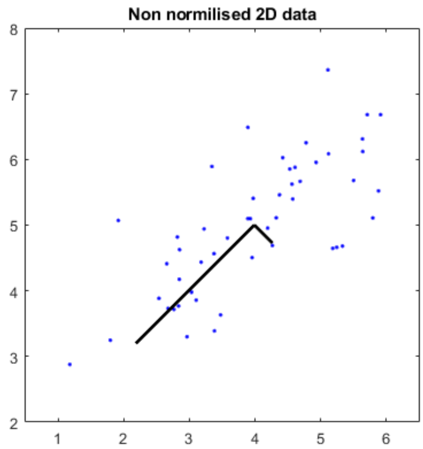


Σχήμα 1: Raw 2D data

και διασπορά 1, σύμφωνα με τον τύπο:

$$X_{norm} = \frac{(X_i - \mu_x)}{\sigma_i}$$

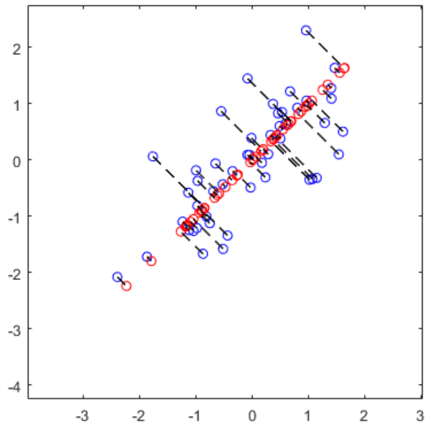
Στην συνέχεια υπολογίσονται οι κύριες συνηστώσεις PCA οι οποίες εξάγωνται από τον covariance matrix των δεδομένων και αφού ταξινομηθούν επιστρέφονται σε μορφή πίνακα και απικονίζονται στα αρχικά δεδομένα . Στην συνέχεια γίνεται η προβολή των $2D$ δεδομένων σε μια διάσταση



Σχήμα 2: Raw 2D data with PCA

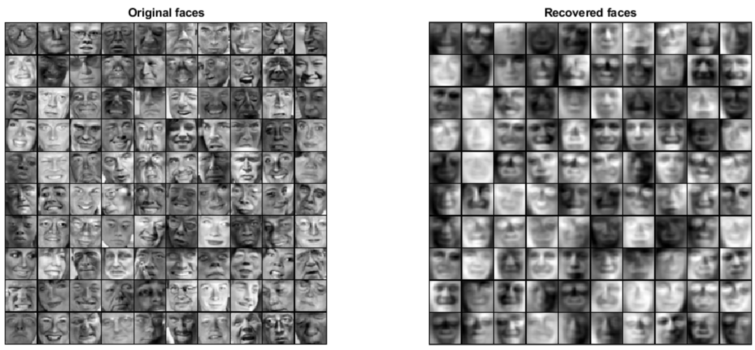
σύμφωνα με τις οδηγίες (γίνεται χρήση των eigenvectors που μας ενδιαφέρουν, στο παράδειγμα, μόνο του πρώτου).

Τα κανονικοποιημένα δεδομένα έχωντας γίνει η προβολή τους:



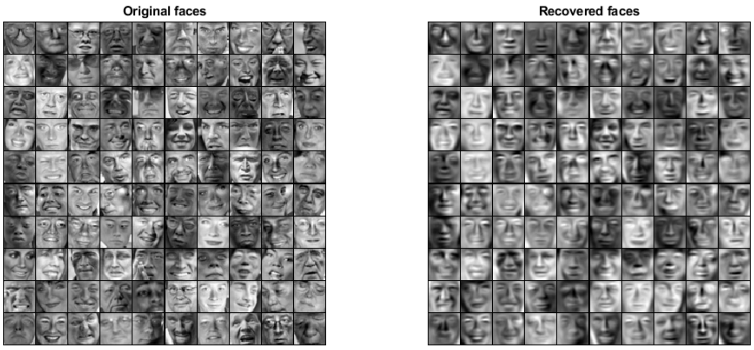
Σχήμα 3: Projection of normalized data

Στην συνέχεια πραγματοποιήθηκαν τα παραπάνω βήματα σε δεδομένα πόλλων διαστάσων. Για να γίνει αντιληπτός ο τρόπος με τον οποίο ο PCA λειτουργεί θα παρουσιαστούν δεδομένα με διαφορετικές διαστάσεις προβολής. Αρχικά για 10 μεταβλητές έχουμε



Σχήμα 4: Recovered data for K = 10

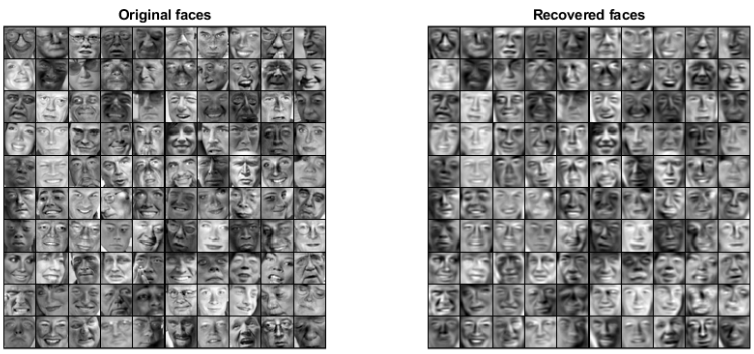
Παρατειρείται ίδη, ότι, από την πληροφορία που κρατείτε στις 10 πρώτες “μεταβλητές” μπορούν να σχηματιστούν τα χαρακτηιστικά των προσώπων μαζί με κάποιες λεπτομέριες. Για $K = 50$ αρχίζουν να εμφανίζονται λεπτομέριες ποιο ευδιάκριτες αλλά και που δεν υπήρχαν πριν. Για



Σχήμα 5: Recovered data for K = 50

$K = 100, 200$ οι εικόνες γίνονται ποιο λεπτομερείς και πλησιάζουν αρκετά της αρχικές.

Μπορούμε να καταλάβουμε πως μεγάλος όγκος πληροφορίας κρατείται στα πρώτα principal components.



Σχήμα 6: Recovered data for K = 100

Δεν έχει νόημα να εμφανιστούν παραπάνω περιπτώσεις μιας και η πληροφορία που ανακτούμε, κάθε φορά, είναι συγκριτικά μικρή σε σχέση με τις αρχικές.

2^{η} Σχεδιασμός ταξινομητή LDA

Για τις ισοπίθανες κατηγορίες:

$$\omega_1 = \mathcal{N}\left(\mu_1, \sigma_1^2\right) \;,\; \mu_1 = \begin{bmatrix} -5 \\ 5 \end{bmatrix} ,\; \sigma_1 = \begin{bmatrix} 11 & 9 \\ 9 & 11 \end{bmatrix}$$
$$\omega_2 = \mathcal{N}\left(\mu_2, \sigma_2^2\right) \;,\; \mu_2 = \begin{bmatrix} 10 \\ 15 \end{bmatrix} ,\; \sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

Για να θπολογιστεί το διάνυσμα προβολής w αρχικά πρέπει να υπολογιστεί ο πίνακας διασποράς εντός των κατηγοριών (μιας και μιλάμε για μόλις 2 κλάσεις), S_w:

$$S_w = \Sigma_{i=1}^{C} P_i \Sigma_i = \frac{1}{2}(\Sigma_1 + \Sigma_2) = \frac{1}{2}\left(\begin{bmatrix} 11 & 9 \\ 9 & 11 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right) = \begin{bmatrix} 6.5 & 4.5 \\ 4.5 & 6.5 \end{bmatrix}$$

Με βάση τον S_w μπορεί να υπολογιστεί το διάνυσμα ως εξής:

$$w = S_w^{-1}(\mu_1 - \mu_2) = \frac{1}{6.5^2 - 4.5^2}\begin{bmatrix} 6.5 & -4.5 \\ -4.5 & 6.5 \end{bmatrix} \cdot \begin{bmatrix} -15 \\ -10 \end{bmatrix} \Rightarrow$$
$$w = \frac{1}{22}\begin{bmatrix} -52.5000 \\ 2.5 \end{bmatrix} = \begin{bmatrix} -2.3864 \\ 0.1136 \end{bmatrix}$$

Τέλος θα κανονικοποιηθεί το w και θα έχουμε:

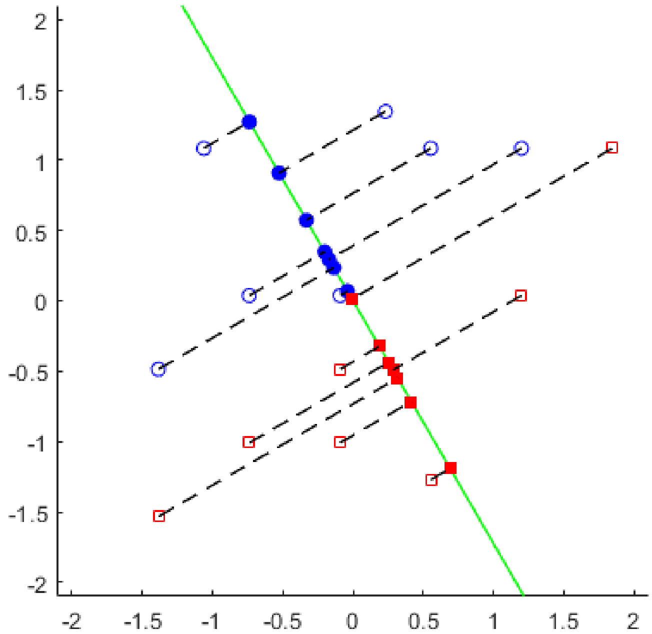
$$w_{normalised} = \frac{w}{||w||} = \frac{w}{\sqrt{(-2.3864)^2 + (0.1136)^2}} = \frac{w}{2.3891} \Rightarrow$$

$$w_{normalised} = \begin{bmatrix} -0.9988 \\ 0.0476 \end{bmatrix}$$

3^η **LDA vs PCA**

Σε πρώτη φάση έγινε η κανονικοποίηση των δεδομένων (όμοια με άσκηση 1) και έπειτα υπολογίστηκε η βέλτηστη διέυθηνση για την οποία τα δεδομένα παρουσιάζουν την μέγιστη διακριτικότητα μεταηξύ των κλάσεων. Ουσιαστικά προσπαθούμε να μεγιστοποιήσουμε την μέση τιμή καθώς και να μειώσουμε την διασπρά της κάθε κλάσης. Αυτή η πληροφορία αποθηκέυεται σε διάνισμα.

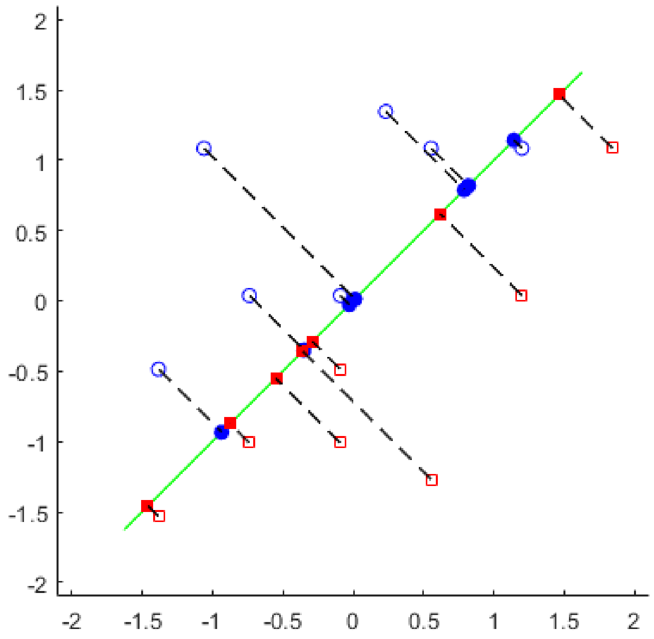
Για τα αρχικά δεδομένα, έχουμε την εξής προβολή στον στην μια διάσταση. Μπορούμε να πα-



Σχήμα 7: LDA 2D to 1D

ρατηρήσουμε εύκολα, την διακρητικότητα μεταξύ των κλάσεων στην μια διάσταση.

Αν προσπαθήσουμε να εφασρμόσουμε PCA στα ίδια δεδομένα παρατηρούμε χειρότερη διακριτικότητα.

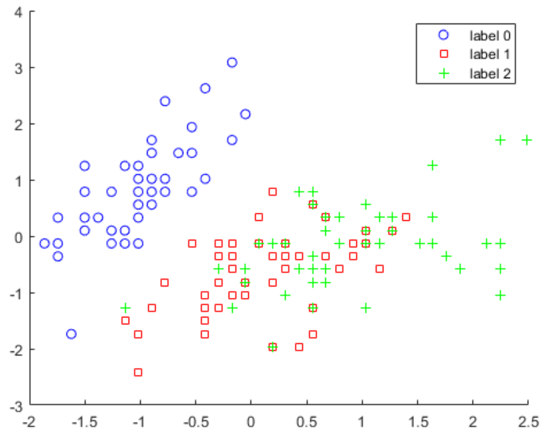


Σχήμα 8: PCA 2D to 1D

Οι κλάσεις στην μια διάσταση, παρουσιάζουν χειρότερη διακριτικότητα σε σχέση με πριν. Χονδρικά αυτό συμβαίνει λόγο του ότι ο PCA κάνει σύνοψη δεδομένων. Δουλεύοντας με τα ιδιοδιανύσματα και τις ιδιοτιμές προπαθή να περιγράψει καλύτερα μια κλάση χωρίς να την πάρει υπόωψην του για την εξαγωγή των principal components.Αντίθετα, ο LDA προσπαθεί να τροποιήσει τα δεδομένα ώστε οι κλάσεις, εκτώς των άλλον, να απέχουν μεταξύ τους.

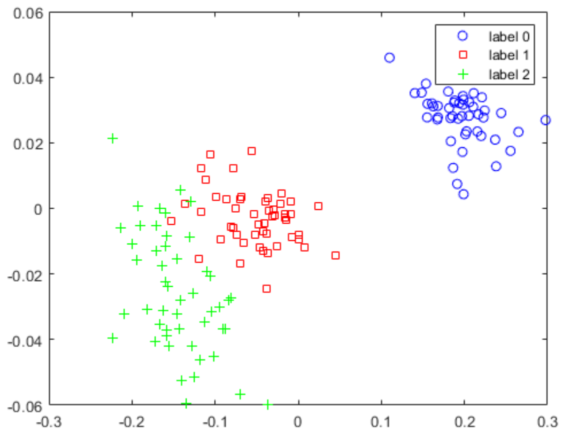
Για την εφαρμογή των δεδομένων σε δεδομένα παραπάνω κλάσεων, ακολουθήθηκε η ίδια λογική αλλά για περισότερες διαστάσεις.

Ποιο συγκεκριμένα για τα παρακάτω δείγματα,



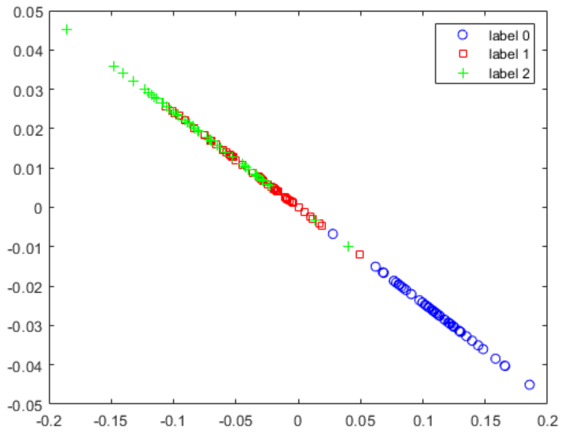
Σχήμα 9: Samples of data

Εφαρμόζοντας τον αλγόριθμο, παίρνουμε το βέλτιστο διάνυσμα. Με την προβολή των δεδομένων, στις 2 διαστάσεις σε αυτό, έχουμε



Σχήμα 10: Data Clasification after LDA at 2d

Αντοίστυχη διαχωριστικότητα θα έχουμε και στην μία διάσταση με αποτελέσματα:



Σχήμα 11: Data Clasification after LDA at 2d

Σε αυτό το σημείο να αναφερθεί ότι τα αρχικά δεδομένα διέφεραν από αυτά που διεθετε η matlab μιας και απαιτούσε την χρήση βιβλιοθηκών που δεν είχα.

4^{η} Bayes

Για τις κατηγορίες:

$$\omega_1 = \mathcal{N}(\mu_1, \Sigma_1) \;,\quad \mu_1 = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \;,\quad \Sigma_1 = \begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix}$$
$$\omega_2 = \mathcal{N}(\mu_2, \Sigma_2) \;,\quad \mu_2 = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \;,\quad \Sigma_2 = \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{bmatrix}.$$

Ανάλογα με την παρατήρηση θα επιλεχθεί η αντίστοιχη κατηγορία.

- Χρησιμοποιώντας τον κανόνα Bayes:

$$P(x)P(\omega_i \mid x) = P(x \mid \omega_i)P(\omega_i) \Rightarrow P(\omega_i \mid x) = \frac{P(x \mid \omega_i)P(\omega_i)}{P(x)}, \quad \omega \in [1, 2]$$

Το σύνορο απόφασης θα βρίσκεται στο σημείο όπου

$$P(\omega_1 \mid x) = P(\omega_2 \mid x) \Rightarrow \frac{P(x \mid \omega_1)P(\omega_1)}{\cancel{P(x)}} = \frac{P(x \mid \omega_2)P(\omega_2)}{\cancel{P(x)}} \Rightarrow$$
$$P(x \mid \omega_1)P(\omega_1) = P(x \mid \omega_2)P(\omega_2) \Rightarrow$$
$$\frac{P(\omega_1)}{P(\omega_2)} = \frac{P(x \mid \omega_2)}{P(x \mid \omega_1)}$$

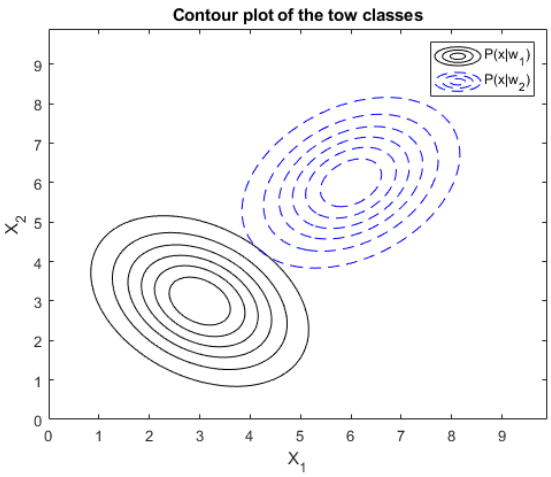
Γνωρίζουμε όμως, τις κατανομές της κάθε κλάσης:

$$P(x \mid \omega_i) = \frac{\exp\left\{-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right\}}{|\Sigma_i|^{\frac{1}{2}} 2\pi}$$

Άρα:

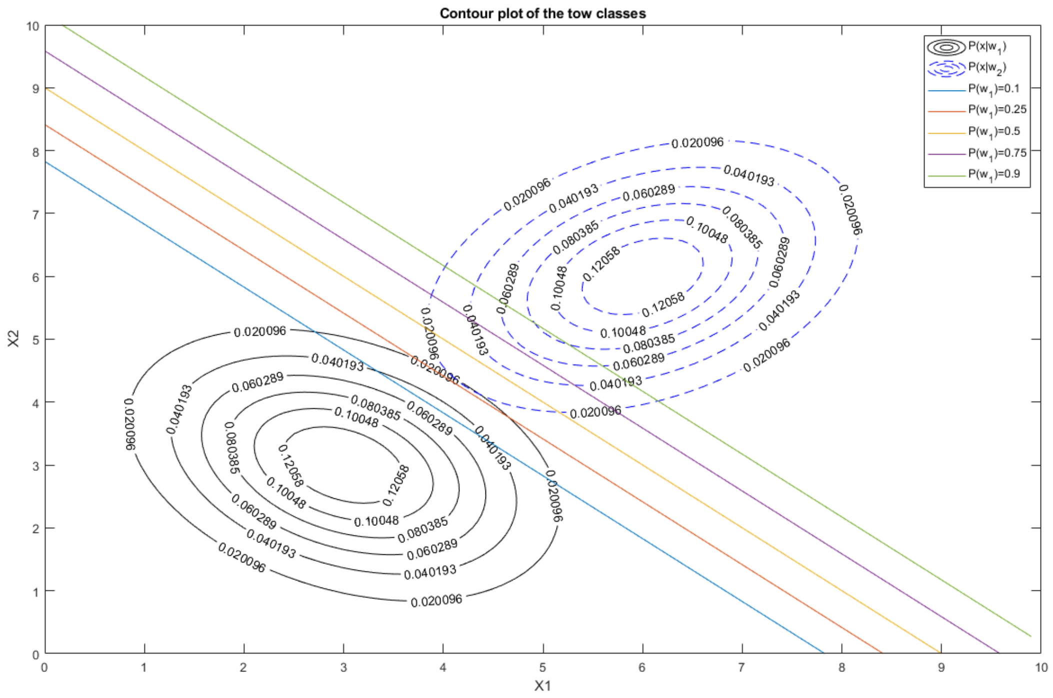
$$\frac{P(\omega_1)}{P(\omega_2)} = \frac{P(x \mid \omega_2)}{P(x \mid \omega_1)} \Rightarrow$$
$$ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right) = ln\left(\frac{P(x \mid \omega_2)}{P(x \mid \omega_1)}\right) \Rightarrow$$
$$ln(P(\omega_1)) - ln(P(\omega_2)) = ln(P(x \mid \omega_2) - ln(P(x \mid \omega_1)) =$$
$$ln\left(\frac{\exp\left\{-\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right\}}{|\Sigma_2|^{\frac{1}{2}}2\pi}\right) - ln\left(\frac{\exp\left\{-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)\right\}}{|\Sigma_1|^{\frac{1}{2}}2\pi}\right) =$$
$$\left(-\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right) - \cancel{ln\left(|\Sigma_2|^{\frac{1}{2}}2\pi\right)} - \left(-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)\right) + \cancel{ln\left(|\Sigma_1|^{\frac{1}{2}}2\pi\right)} =$$
$$\frac{1}{2}\left((x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) - (x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right) =$$
$$\frac{1}{2}\left(\begin{bmatrix}x_1-3 & x_2-3\end{bmatrix} \cdot \begin{bmatrix}0.9375 & 0.3125\\ 0.3125 & 0.9375\end{bmatrix} \cdot \begin{bmatrix}x_1-3\\ x_2-3\end{bmatrix} - \cdots\right) =$$
$$\cancel{0.9375x_1^2} + 0.625x_1x_2 - \cancel{7.5x_1} + \cancel{0.9375x_2^2} + 22.5 - \cancel{7.5x_2} -$$
$$\cancel{0.9375x_1^2} + 0.625x_1x_2 + \cancel{7.5x_1} - \cancel{0.9375x_2^2} - 45 + \cancel{7.5x_2} =$$
$$1.25x_1x_2 - 22.5 \Rightarrow$$
$$\frac{5\left(ln(P(\omega_1)) - ln(P(\omega_2))\right)}{4} + 18 = x_1x_2 \Rightarrow$$
$$x_1 = \frac{\frac{5(ln(P(\omega_1))-ln(P(\omega_2)))}{4} + 18}{x_2}$$

- Οι ισοϋψείς καμπύλες των δεσμευμένων πιθανοτήτων:



Σχήμα 12: Samples of data

- Οι ισοϋψείς καμπύλες των δεσμευμένων πιθανοτήτων με τα σύνορα απόφασης : Παρατηρο-



Σχήμα 13: Samples of data

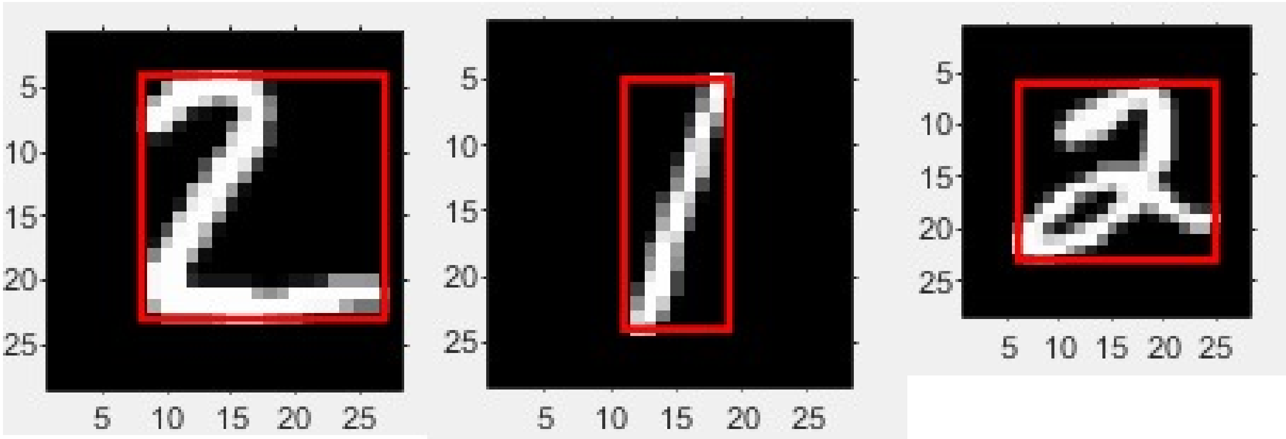
ύμε πως τα όρια είναι ανάλογα με την κάθε κατανομή.Για μικρό όριο της κλάσης ω_1 είναι αντοίστοιχα και μικρή η περιοχή κατονομής σε σχέση με την δεύτερη κλάση.

Στην περίπτωση που οι πίνακες συνδιακύμανσης είναι ίσοι, παύουν να μας ενδιαφέρουν οι τετραγωνικοί όροι καθώς δεν τους λαμβάνουμε υπ όψιν στις συκρίσεις. Οι συναρτήσεις διάκρισης είναι γραμμικρές και η περιοχή απόφασης μετακινείται προς την περιοχή με την μικρότερη πιθανότητα:

5^η Εξαγωγή χαρακτηριστικών και Bayes Classification

Αρχικά υπολογίστηκε το aspec ratio για κάθε εικόνα, κάθε κλάσης, (test και test) και αποθυκεύτηκε.

Ενδεικτικά κάποια από τα δεδομένα κατά τον υπολογισμό:



Σχήμα 14: Samples of data

Στην συνέχεια, έχοντας υπολογίσει τις prior πιθανότητες προσπαθοούμε να φριάξουμε τις κατανομές τις κάθε κλάσης του ταξινομητή.

Υπολογίζουμε τις μέσες τιμές τις κάθε κλάσης ($\mu_{C1} = 0.468$ και $\mu_{C2} = 0.961$) και την διακύμανσή τους ($\sigma_{C1} = 0.199$ και $\sigma_{C1} = 0.1883$).

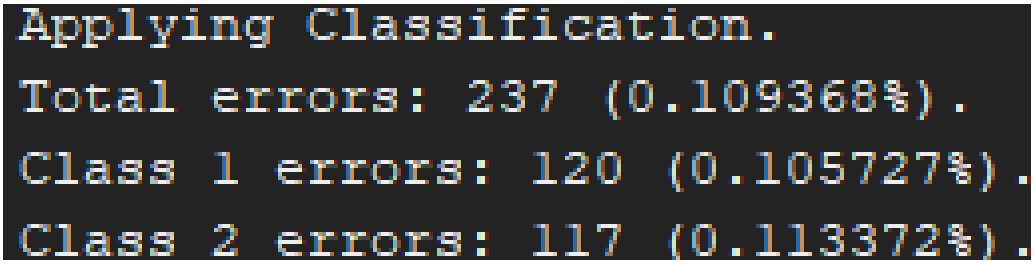
Έπειτα κατασκευάζουμε τις κατανομές τους ώς εξής

$$P(\omega_i \mid x) = \frac{1}{\sigma_i\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)$$

Γνωρίζοντας τις κατανομές κάθε κλάσης και άρα τον ταξινομητή, "περνάμε" εικόνα, από το test set από αυτόν και λαμβάνουμε τις πιθανότητες $P(x \mid \omega_1)$ (δεν είναι κανονικοποιημένες).

Τέλος, για κάθε εικόνα αποθηκεύουμε την κλάση που εμφανίζει την μέγιστη πιθανότητα και ελέγχουμε την εσφαλμένη ταξινόμηση, για καθε μια από αυτή.

Για σαν τελικά αποτελέσματα έχουμε τα παρακάτω



Σχήμα 15: Classification Errors

Παρατειρούμε πως η ταξινόμηση με βάση το ελάχιστο ορθογώνιο που περικλείει το κάθε ψηφίο είναι εμφανίζει σφάλμα της τάξεως του 10%

6^η Minimum risk

Για τον πίνακα ρίσκου $L = \begin{bmatrix} 0 & 0.5 \\ 1.0 & 0 \end{bmatrix}$ το κόστος πρόβλεψης ω_1 δεδομένου της ω_2 είναι $L_{12} = 0.5$ και αντίστοιχα δεδομένου ω_1, το κόστος για ω_2 είναι $L_{21} = 1$.

Για να υπολογιστεί το όριο απόφασης με το μικρότερο ρίσκο θα πρέπει να υπολογιστεί το ρίσκο για κάθε κλάση.

$$R(\omega_1 \mid x) = \cancelto{0}{L_{11}P(\omega_1 \mid x)} + L_{12}P(\omega_2 \mid x)$$
$$R(\omega_2 \mid x) = \cancelto{0}{L_{22}P(\omega_1 \mid x)} + L_{21}P(\omega_1 \mid x)$$

Για το όριο:

$$\begin{aligned}
L_{12}P(\omega_2 \mid x) = L_{21}P(\omega_1 \mid x) \Rightarrow \\
L_{12}P(x \mid \omega_2)P(\omega_2) = L_{21}P(x \mid \omega_1)P(\omega_1) \Rightarrow \\
\frac{P(x \mid \omega_2)}{P(x \mid \omega_1)} = \frac{P(\omega_1)L_{21}}{L_{12}P(\omega_2)} = 2 \Rightarrow \\
\frac{x}{\sigma_2^2} exp\{-\frac{x^2}{2\sigma_2}\} = 2\frac{x}{\sigma_1^2} exp\{-\frac{x^2}{2\sigma_1}\} \Rightarrow \\
\frac{x}{4} exp\{-\frac{x^2}{8}\} = 2\frac{x}{1} exp\{-\frac{x^2}{2}\} \Rightarrow \\
\frac{1}{8} exp\{-\frac{x^2}{8}\} = exp\{-\frac{x^2}{2}\} \overset{ln()}{\Rightarrow} \\
ln(\frac{1}{8}) - \frac{x^2}{8} = -\frac{x^2}{2} \Rightarrow \\
ln(1) - ln(2^3) = \frac{x^2}{8} - \frac{x^2}{2} \Rightarrow \\
-2.08 = \frac{-3x^2}{8} \Rightarrow x \simeq 2.35
\end{aligned}$$