
Πολυτεχνείο Κρήτης
Σχολή ΗΜΜΥ
Στατιστική Μοντελοποίηση και Αναγνώριση Προτύπων
1η Σειρά Ασκήσεων
Ομάδα Χρηστών 90

Επώνυμο	Όνομα	A.M.
Νικολός	Κωνσταντίνος	2019030096
Μπέκος	Κωνσταντίνος	2019030082

Θέμα 1: Principal Component Analysis (PCA)

Μέρος 1

- Μετά από την ανάγνωση του αρχείου με τα δεδομένα, επιλέγονται τέσσερα (4) ζεύγη χαρακτηριστικών και απεικονίζονται στον 2D χώρο. Στα αντίστοιχα γραφήματα, παρατηρούμε, ότι, κατά κύριο λόγο, τα χαρακτηριστικά που ανήκουν στην κλάση με ετικέτα μηδέν (0, καλοήθες) έχουν μεγαλύτερες τιμές από τα αντίστοιχα της άλλης κατηγορίας, αλλά δεν είναι δυνατόν να διαχωριστούν με απόλυτη επιτυχία. [Figure 1]

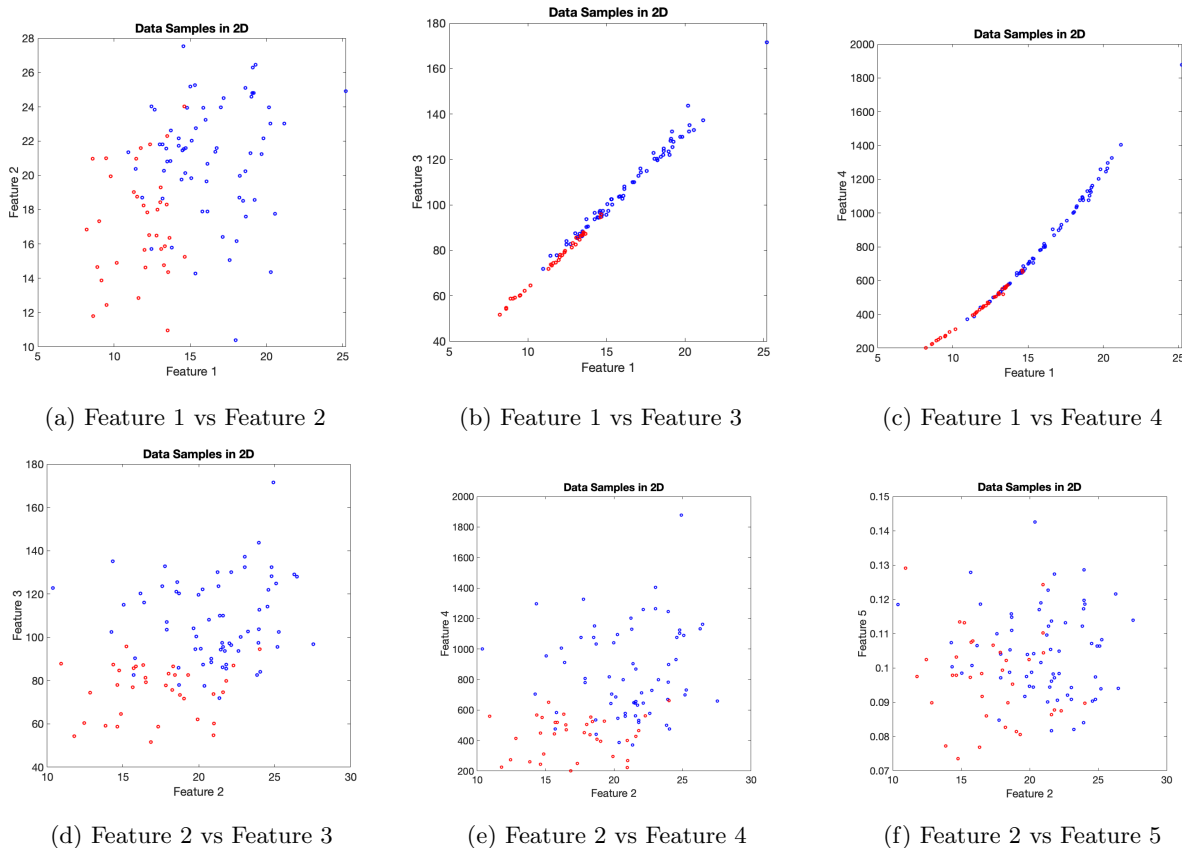
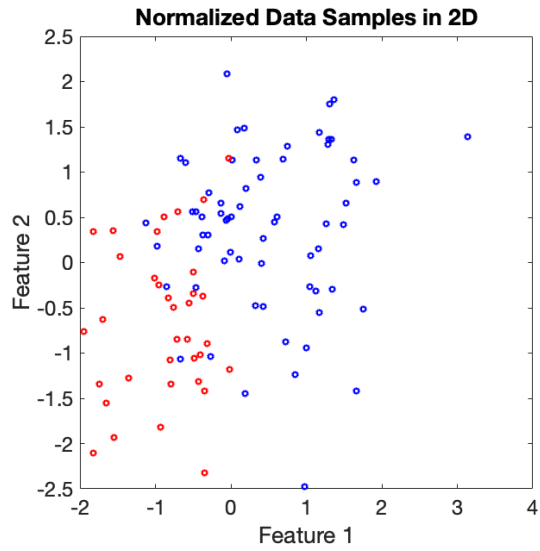
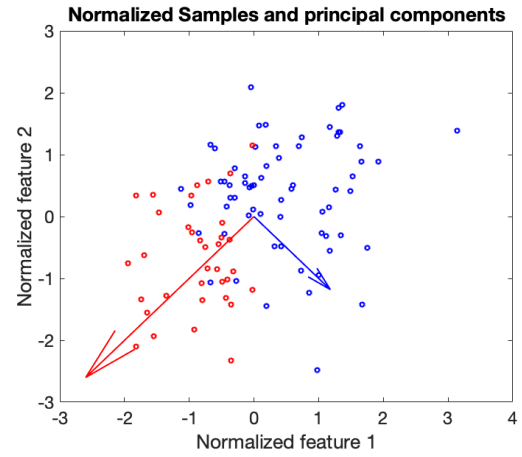


Figure 1: Graphical Representation of Feature Pairs on 2D plane

- Στη συνέχεια, επιλέγονται τα πρώτα δύο (2) χαρακτηριστικά (Feature 1, Feature 2) και εφαρμόζεται κανονικοποίηση (standardization) στα αρχικά δείγματα. Παρατηρούμε, ότι τα κανονικοποιημένα δείγματα εμφανίζονται συγκεντρωμένα γύρω από την αρχή των αξόνων, ενώ τα αρχικά είναι περισσότερο κατανεμημένα και μακριά από την αρχή των αξόνων. [Figure 2a]
- Στο επόμενο βήμα, εφαρμόζεται ο αλγόριθμος PCA και υπολογίζονται οι κύριες συνιστώσες των χαρακτηριστικών, οι οποίες σχεδιάζονται μαζί με τα δύο (2) πρώτα χαρακτηριστικά (Feature 1 vs Feature 2) των κανονικοποιημένων δειγμάτων. Οι κύριες συνιστώσες είναι κάθετες μεταξύ τους και, επειδή τα δείγματα έχουν δύο διαστάσεις (χαρακτηριστικά) εξαρχής, αυτό είναι εμφανές και στο διδιάστατο γράφημα. [Figure 2b]
- Όταν τα δεδομένα προβληθούν στον χώρο δύο (2) διαστάσεων, ο οποίος ορίζεται από τις δύο (2) κύριες συνιστώσες που υπολογίζονται από τον αλγόριθμο PCA, το αποτέλεσμα είναι απλώς μία περιστροφή τους, διότι βρίσκονται ήδη στον ίδιο διδιάστατο χώρο. [Figure 3a]
- Αν προβληθούν σε μονοδιάστατο χώρο (1D) και ανακτηθούν (με προβολή στον αρχικό χώρο δύο διαστάσεων), εμφανίζονται ως οι προβολές των αρχικών σημείων, στην ευθεία που ορίζεται από την πρώτη κύρια συνιστώσα. [Figure 3b]
- Ως τελευταίο βήμα, η ίδια ανάλυση PCA επαναλαμβάνεται για όλα τα χαρακτηριστικά (30) των δεδομένων. Μετά την κανονικοποίηση και τον υπολογισμό των κυρίων συνιστωσών των δειγμάτων, σχεδιάζονται οι

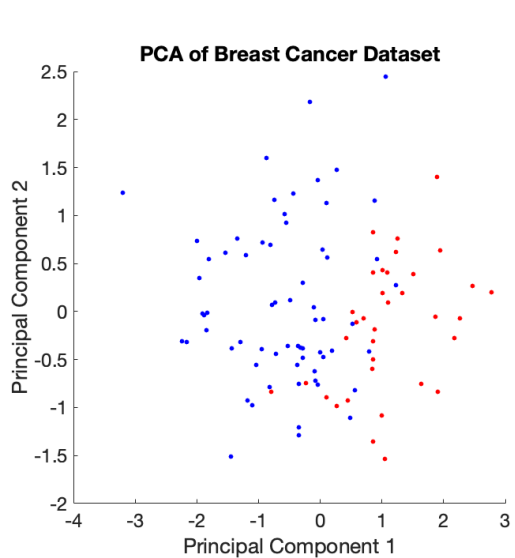


(a) Feature 1 vs Feature 2 Normalized

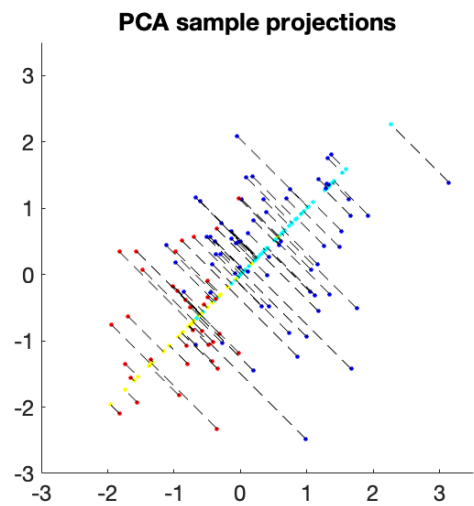


(b) PCA Principal Components and Feature 1 vs Feature 2 Normalized

Figure 2: Feature Normalization



(a) Projected Samples on 2D space



(b) Projection of Samples on 1D space and Recovering

Figure 3: PCA Sample Projection

προβολές των δύο (2) πρώτων κυρίων συνιστωσών και των δειγμάτων, στο επίπεδο που ορίζεται από τους άξονες των δύο (2) διαστάσεων (χαρακτηριστικών) που εμφανίζουν τη μεγαλύτερη διακύμανση. Τα διανύσματα που εμφανίζονται δεν είναι (απαραίτητα) κάθετα μεταξύ τους, επειδή αποτελούν τις προβολές δύο κάθετων διανυσμάτων, από χώρο τριάντα διαστάσεων (30D), σε δισδιάστατο (2D). [Figure 4a]

- Τέλος, αν προβληθούν τα δείγματα πάνω στις δύο (2) πρώτες κύριες συνιστώσες, διατηρείται περίπου το 63% της αρχικής συνολικής διασποράς τους, χωρίς όμως να διαχωρίζονται πλήρως οι δύο κλάσεις, αφού δεν είναι αυτός ο σκοπός της ανάλυσης και του αλγορίθμου. [Figure 4b]

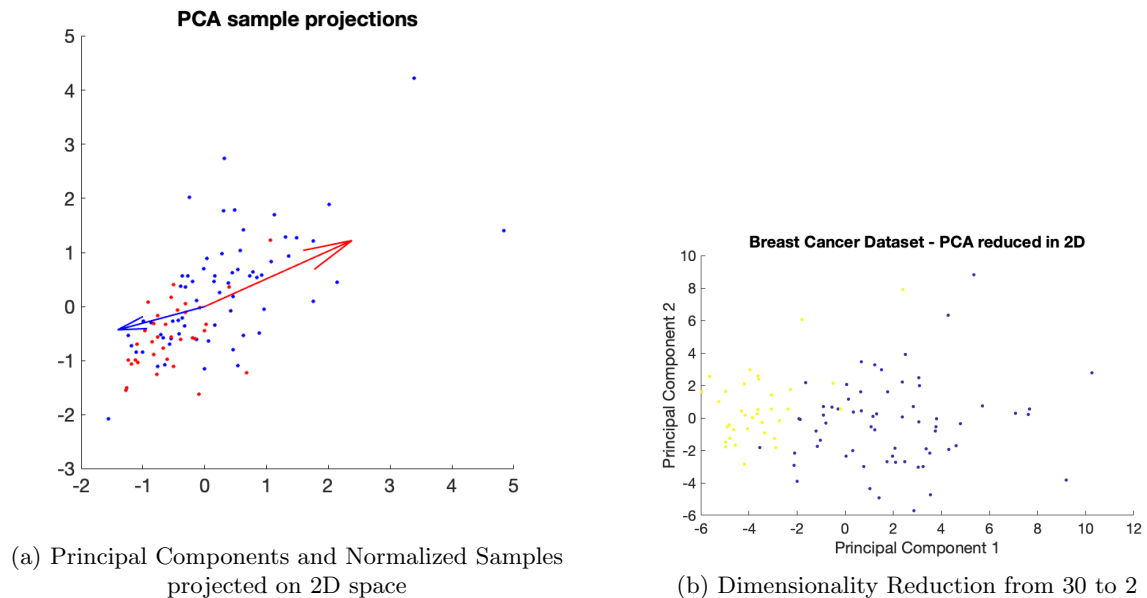


Figure 4: PCA Feature Extraction of 30D data

Μέρος 2

- Μετά τη φόρτωση των δεδομένων των εικόνων, σχεδιάζονται εκατό (100) πρόσωπα. [Figure 5]



Figure 5: Printing 100 images using original data

- Έπειτα, εφαρμόζεται κανονικοποίηση στα δεδομένα και υπολογίζονται οι κύριες συνιστώσες τους, μέσω του αλγορίθμου *PCA*. Αφού επιλέξουμε τις πρώτες τριάντα έξι (36) από αυτές, τις μεταχειριζόμαστε ως ισάριθμες εικόνες και τις τυπώνουμε, θεωρώντας τα στοιχεία τους ως τιμές pixel. Οι τριάντα έξι (36) εικόνες που σχηματίζονται θυμίζουν θολά σχέδια προσώπων. [Figure 6]



Figure 6: Printing 36 Principal Components

- Μειώνοντας τη διάσταση των δειγμάτων σε εκατό (100) και προβάλλοντάς τα στον αρχικό χώρο, παρατηρούμε, ότι η ανάκτησή τους είναι αρκετά ικανοποιητική, με την έννοια, ότι διατηρούνται σε μεγάλο βαθμό οι μορφές των προσώπων, με εξαίρεση κάποιες λεπτομέρειες. [Figure 7]

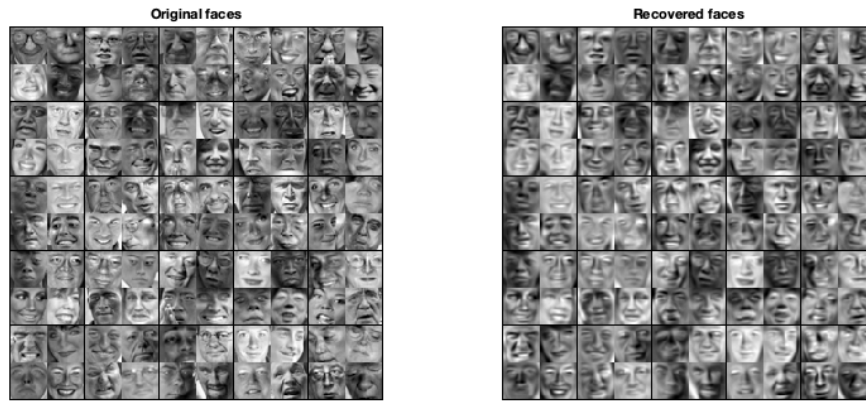


Figure 7: Original Data and Recovered Data after Dimensionality Reduction to 100

- Το προηγούμενο βήμα επαναλαμβάνεται με διαφορετικούς ρυθμούς κυρίων συνιστωσών. Πιο συγκεκριμένα, η διάσταση των δειγμάτων μειώνεται σε 10, 50 και 200. Προφανώς, όσο μικρότερη είναι η διάσταση, τόσο λιγότερο επιτυχής είναι η ανάκτηση των δεδομένων. [Figure 8]

Θέμα 2: Σχεδιάστε ένα ταξινομητή LDA (Linear Discriminant Analysis)

1. Το διάνυσμα προβολής w που ζητείται είναι:

$$w = S_w^{-1}(\mu_A - \mu_B)$$

Ο πίνακας σκέδασης είναι:

$$S_w = P(\omega_A) \cdot \Sigma_A + P(\omega_B) \cdot \Sigma_B = 0.4 \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} + 0.6 \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 2.4 & 1 \\ 1 & 2.6 \end{bmatrix}$$

Ο αντίστροφος του πίνακα σκέδασης είναι:

$$S_w^{-1} = \frac{1}{2.4 \cdot 2.6 - 1} \begin{bmatrix} 2.6 & -1 \\ -1 & 2.4 \end{bmatrix} = \frac{1}{6.24} \begin{bmatrix} 2.6 & -1 \\ -1 & 2.4 \end{bmatrix}$$



(a) 10 Principal Components



(b) 50 Principal Components



(c) 200 Principal Components

Figure 8: Original Data and Recovered Data after Dimensionality Reduction

Οπότε,

$$w = \frac{1}{6.24} \begin{bmatrix} 2.6 & -1 \\ -1 & 2.4 \end{bmatrix} \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \frac{1}{6.24} \begin{bmatrix} -7.2 \\ 6.8 \end{bmatrix} = \frac{1}{1.56} \begin{bmatrix} -1.8 \\ 1.7 \end{bmatrix}$$

Το κανονικοποιημένο διάνυσμα (μέτρου 1) στην κατεύθυνση του w είναι:

$$w_1 = \frac{w}{\|w\|} = \frac{1}{\sqrt{6.13}} \begin{bmatrix} -1.8 \\ 1.7 \end{bmatrix}$$

2. Η προβολή του διανύσματος x_A , στον άξονα που ορίζεται από το διάνυσμα w , είναι:

$$(w_1^T x_A)w_1 = \frac{1}{6.13} \left(\begin{bmatrix} -1.8 & 1.7 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right) \begin{bmatrix} -1.8 \\ 1.7 \end{bmatrix} = \frac{3.3}{6.13} \begin{bmatrix} -1.8 \\ 1.7 \end{bmatrix} \approx \begin{bmatrix} -0.97 \\ 0.92 \end{bmatrix}$$

Η προβολή του διανύσματος x_B , στον άξονα που ορίζεται από το διάνυσμα w , είναι:

$$(w_1^T x_B)w_1 = \frac{1}{6.13} \left(\begin{bmatrix} -1.8 & 1.7 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right) \begin{bmatrix} -1.8 \\ 1.7 \end{bmatrix} = \frac{-7.1}{6.13} \begin{bmatrix} -1.8 \\ 1.7 \end{bmatrix} \approx \begin{bmatrix} 2.08 \\ -1.97 \end{bmatrix}$$

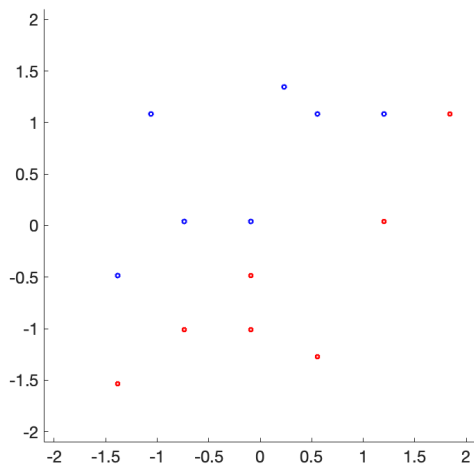
Θέμα 3: Linear Discriminant Analysis (LDA) vs PCA

Μέρος 1

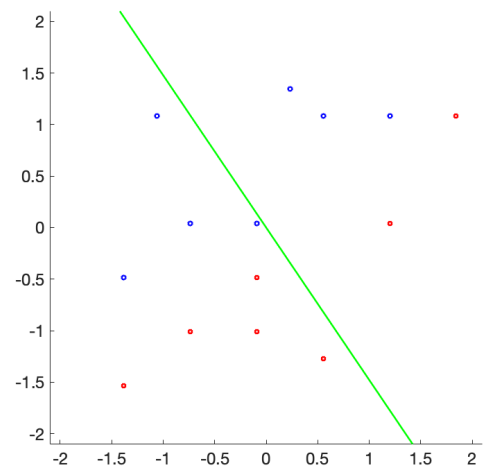
- Με τη μέθοδο *PCA*, τα δείγματα δεν μπορούν να διαχωριστούν σε κλάσεις, μετά την προβολή τους, ενώ με τη μέθοδο *LDA* μπορούν. [Figure 9]

Μέρος 2

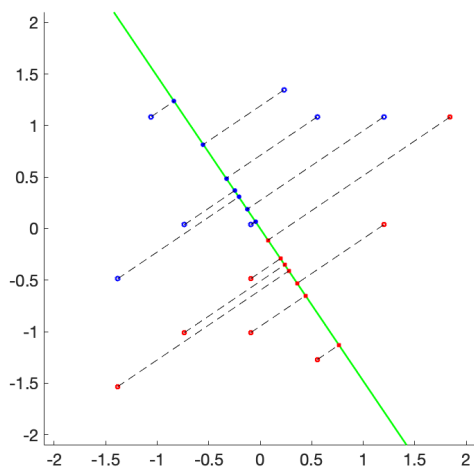
- Μετά από την εφαρμογή του αλγορίθμου *LDA* στη βάση δεδομένων *Iris*, τα δείγματα των τριών (3) κλάσεων είναι σχεδόν πλήρως διαχωρίσιμα, με μόνο κάποιο μικρό σφάλμα, μεταξύ της πράσινης και της κόκκινης κλάσης. [Figure 10]



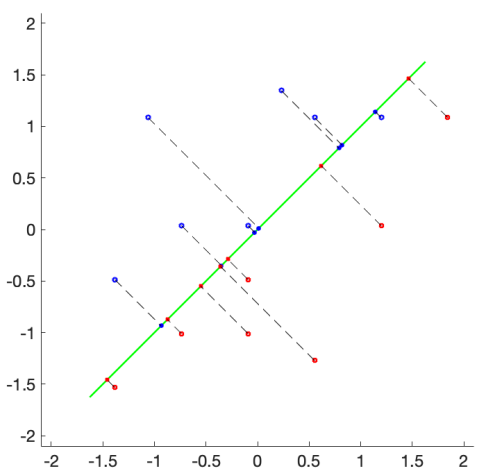
(a) Normalized Data



(b) Optimal Direction for Maximum Class Separation (*LDA*)

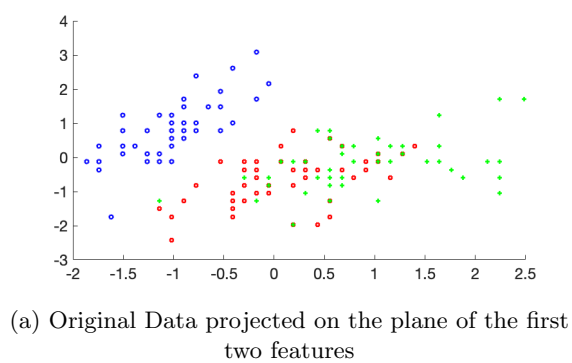


(c) Recovering the Projected Data (*LDA*)

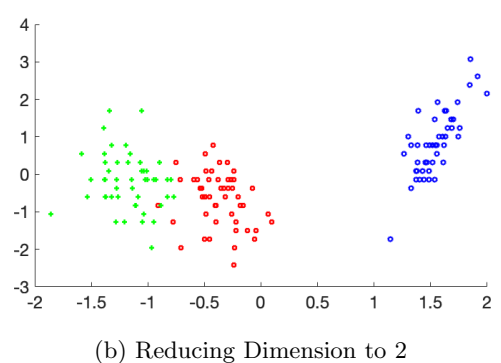


(d) Applying *PCA* algorithm

Figure 9: *LDA* vs *PCA* for a 2-class problem



(a) Original Data projected on the plane of the first two features



(b) Reducing Dimension to 2

Figure 10: Applying *LDA* algorithm on *Iris* dataset

Θέμα 4: Bayes

1. Η εξίσωση του συνόρου απόφασης δίνεται από τη σχέση:

$$\begin{aligned}
 P(x|\omega_1)P(\omega_1) &= P(x|\omega_2)P(\omega_2) \\
 \Rightarrow \frac{P(\omega_1)}{2\pi\sqrt{|\Sigma_1|}}e^{-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)} &= \frac{P(\omega_2)}{2\pi\sqrt{|\Sigma_2|}}e^{-\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)} \\
 \Rightarrow \frac{P(\omega_1)}{\sqrt{|\Sigma_1|}}e^{-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)} &= \frac{P(\omega_2)}{\sqrt{|\Sigma_2|}}e^{-\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)} \\
 \Rightarrow \ln \frac{P(\omega_1)}{\sqrt{|\Sigma_1|}} - \frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) &= \ln \frac{P(\omega_2)}{\sqrt{|\Sigma_2|}} - \frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2) \\
 \Rightarrow 2\ln \frac{P(\omega_1)}{P(\omega_2)} + \ln \frac{|\Sigma_2|}{|\Sigma_1|} &= (x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) - (x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2) \\
 \Rightarrow 2\ln \frac{P(\omega_1)}{P(\omega_2)} + \ln \frac{1.11}{1.75} \\
 &= (x^T - [2 \ 3])\frac{1}{1.75} \begin{bmatrix} 1 & -0.5 \\ -0.5 & 2 \end{bmatrix} \left(x - \begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) - (x^T - [4 \ 4])\frac{1}{1.11} \begin{bmatrix} 0.8 & 0.3 \\ 0.3 & 1.5 \end{bmatrix} \left(x - \begin{bmatrix} 4 \\ 4 \end{bmatrix}\right) \\
 &= \frac{1}{1.75} [(x_1 - 2) - 0.5(x_2 - 3) \quad (-0.5)(x_1 - 2) + 2(x_2 - 3)] \begin{bmatrix} x_1 - 2 \\ x_2 - 3 \end{bmatrix} \\
 &\quad - \frac{1}{1.11} [(x_1 - 4)0.8 + (x_2 - 4)0.3 \quad (x_1 - 4)0.3 + (x_2 - 4)1.5] \begin{bmatrix} x_1 - 4 \\ x_2 - 4 \end{bmatrix} \\
 &= \frac{1}{1.75} [(x_1 - 2)^2 - 0.5(x_1 - 2)(x_2 - 3) - 0.5(x_1 - 2)(x_2 - 3) + 2(x_2 - 3)^2] \\
 &\quad - \frac{1}{1.11} [(x_1 - 4)^2 0.8 + (x_1 - 4)(x_2 - 4)0.3 + (x_1 - 4)(x_2 - 4)0.3 + (x_2 - 4)^2 1.5] \\
 &\Rightarrow 2\ln \frac{P(\omega_1)}{P(\omega_2)} + \ln \frac{1.11}{1.75} \\
 &= \frac{1}{1.75} [(x_1 - 2)^2 - (x_1 - 2)(x_2 - 3) + 2(x_2 - 3)^2] - \frac{1}{1.11} [(x_1 - 4)^2 0.8 + 0.6(x_1 - 4)(x_2 - 4) + (x_2 - 4)^2 1.5]
 \end{aligned}$$

2. Οι συσφείς καμπύλες των δύο κατανομών φαίνονται παρακάτω (ω_1 : μπλε, ω_2 : κόκκινο). [Figure 11]

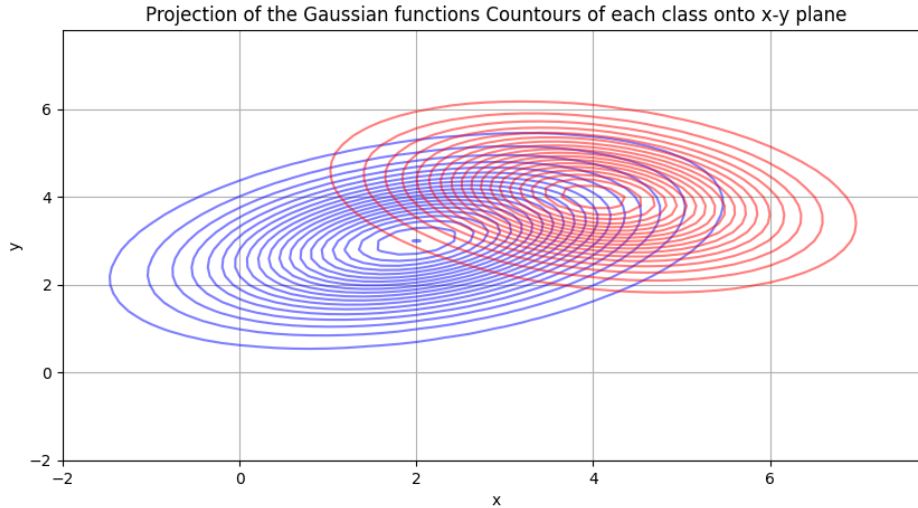


Figure 11: Level Sets of 2-dimensional Normal PDFs

3. Αρχικά παρατηρείται, ότι, για διαφορετικές τιμές πινάκων Σ , προκύπτει, ως όριο απόφασης, μια καμπύλη, λόγω της διαφορετικότητας των δύο καμπανών. Εφόσον στο παράδειγμα μας δίνεται, ότι οι πιθανότητες των δύο Gaussian συναρτήσεων είναι αντίθετες, όσο αυξάνεται η τιμή του $P(\omega_1)$, τόσο το σύνορο

απόφασης μετακινείται προς την κλάση ω_1 , δίνοντας περισσότερη σημασία σε αυτή (κόκκινη γραμμή: $P(\omega_1) = 0.1$, μαύρη γραμμή: $P(\omega_1) = 0.9$). [Figure 12]

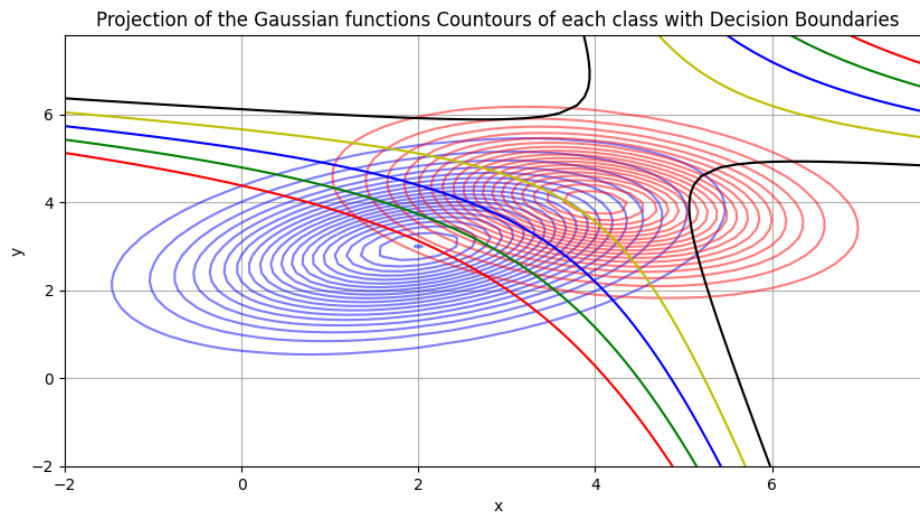


Figure 12: Decision Boundaries for different covariance matrices

4. Στην περίπτωση που οι τιμές των πινάκων Σ είναι ίδιες, προκύπτει, ότι το όριο απόφασης είναι μία ευθεία γραμμή καθώς οι δύο καμπάνες είναι ίδιες και διαφέρουν μόνο ως προς το κέντρο τους. [Figure 13]

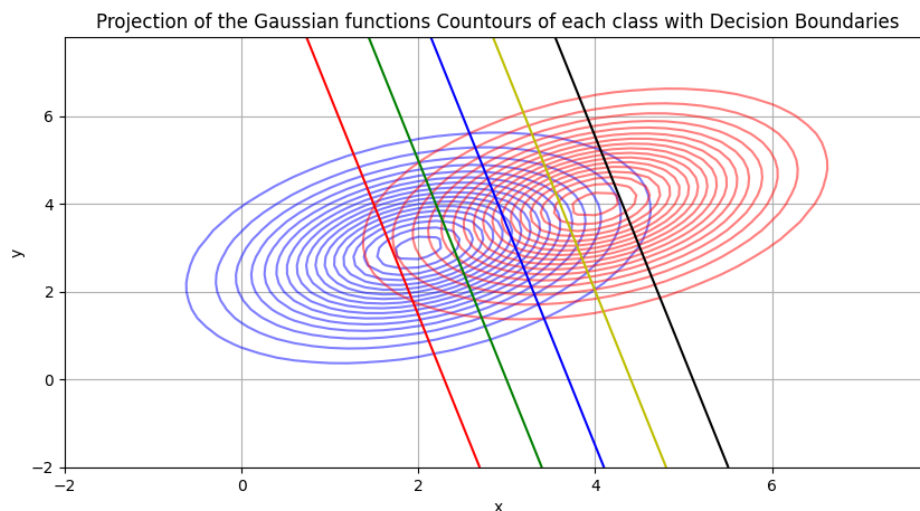
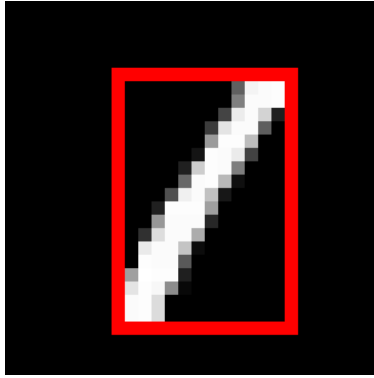


Figure 13: Decision Boundaries for identical covariance matrices

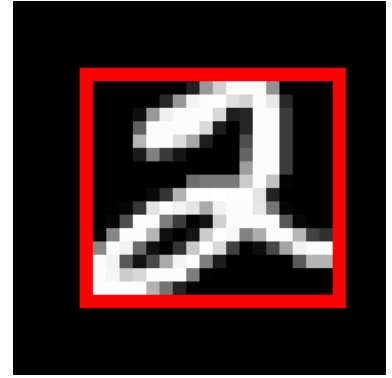
Θέμα 5: Εξαγωγή χαρακτηριστικών και Bayes Classification

Μέρος Α

- Παρατηρείται, από το αποτέλεσμα, ότι ο υπολογισμός του aspect ratio είναι σωστός, αφού, στην εικόνα που τυπώνεται, ο αριθμός περικλείεται σωστά από το bounding box και το bounding box υπολογίζεται με τα ίδια δεδομένα με τα οποία υπολογίζεται και το aspect ratio. [Figure 14]
- Αφού εκπαιδευτεί ο Bayes Classifier με τα χαρακτηριστικά των training data που δημιουργήθηκαν (aspect-ratio), χρησιμοποιούνται τα υπόλοιπα δεδομένα ως test ακρίβειας του classifier, δίνοντας ένα ικανοποιητικό score:0,954545.



(a) digit 1



(b) digit 2

Figure 14: Digits with bounding boxes

- Προστίθενται δύο νέα χαρακτηριστικά για την εκπαίδευση του Bayesian Classifier δίνοντας ακόμη καλύτερα αποτελέσματα με score:0,96694 που σημαίνει ότι και τα δύο νέα χαρακτηριστικά λειτουργούν σωστά αυξάνοντας την ακρίβεια του ταξινομητή.

Θέμα 6: Minimum risk

Το όριο απόφασης δίνεται από τη λύση της εξίσωσης:

$$\begin{aligned}
 \frac{P(x_0|\omega_1)}{P(x_0|\omega_2)} &= \frac{P(\omega_1)}{P(\omega_2)} \cdot \frac{\lambda_{21}}{\lambda_{12}} \\
 \Rightarrow \frac{\frac{x_0}{\sigma_1^2} e^{-\frac{x_0^2}{2\sigma_1^2}}}{\frac{x_0}{\sigma_2^2} e^{-\frac{x_0^2}{2\sigma_2^2}}} &= \frac{\lambda_{21}}{\lambda_{12}} \\
 \Rightarrow \frac{4e^{-\frac{x_0^2}{2}}}{e^{-\frac{x_0^2}{8}}} &= \frac{1}{0.5} \\
 \Rightarrow 2e^{\frac{x_0^2}{8} - \frac{x_0^2}{2}} &= 1 \\
 \Rightarrow e^{\frac{-3x_0^2}{8}} &= \frac{1}{2} \\
 \Rightarrow \frac{-3x_0^2}{8} &= -\ln 2 \\
 \Rightarrow x_0^2 &= \frac{8}{3} \ln 2 \\
 \Rightarrow x_0 &= \sqrt{\frac{8}{3} \ln 2}
 \end{aligned}$$