Στατιστική Μοντελοποίηση

Άσκηση 1

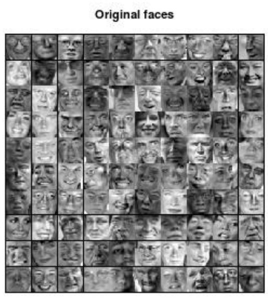
Αναστάσιος Πατερίτσας ΑΜ: 2016030065

Θέμα 1.

Σε αυτή την άσκηση έπρεπε να υπολογίσουμε το Principal Component Analysis. Για να επιτευχτεί αυτό χρειαζόταν να κάνουμε τα εξής βήματα:

- Κάναμε κανονικοποιηση με μέση τιμή το 0 και διασπορά το 1 , που έγινε στον κώδικα featureNormalize.m .
- Υπολογίσαμε τον πίνακα συνιδιασπορας των normalize data συμφώνα με το τύπο $\Sigma = \frac{1}{m} X^T X$ και επιστρεφουμε τα eigenvectors και eigenvalues.
- Χρησιμοποιήσαμε τον κώδικα projectData.m για να μειώσουμε την διάσταση του πρώτου δείγματος. Για να γίνει αυτό χρησιμοποιήσαμε τα top K eigenvectors $z_i = U^T x_i$.
- Χρησιμοποιήσαμε τον κώδικα recoverData.m για να ανακτήσουμε την πληροφορία . Για να γίνει αυτό κάναμε την ανάποδη διαδικασία $x_{rec} = U * z$.

Εφόσον βεβαιωθήκαμε ότι η παραπάνω διαδικασία λειτούργει τότε εφαρμόσαμε το PCA σε δεδομένα 5000 εικόνων.



Η τεχνικη Principal Component Analysis είναι πολύ καλη στο να μειώνει τις διαστάσεις διανύσματος χαρακτηριστικών. Αυτό επιβεβαιώνεται και στο 2ο μέρος με τα πρόσωπα εφόσον δε παρατηρούνται μεγάλες διαφορές μεταξύ των εικόνων με 1024 χαρακτηριστικά και των εικόνων με 100.

Θέμα 2.

Διάνυσμα προβολής:

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

$$Sw = \sum_w \frac{1}{2}(\Sigma_1 + \Sigma_2) = \frac{1}{2}\begin{bmatrix} 11+2 & 9+0 \\ 9+0 & 11+2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 13 & 9 \\ 9 & 13 \end{bmatrix} = \begin{bmatrix} \frac{13}{2} & \frac{9}{2} \\ \frac{9}{2} & \frac{13}{2} \end{bmatrix}$$

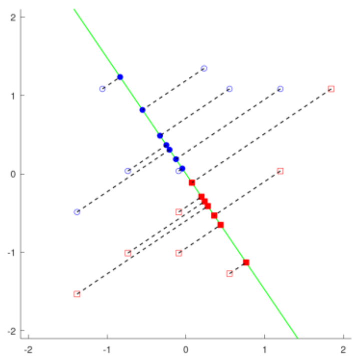
Άρα

$$Sw^{-1} = \sum\nolimits_w^{-1} = \frac{1}{\frac{13^2}{2} - \frac{9^2}{2}} \begin{bmatrix} \frac{13}{2} & -\frac{9}{2} \\ -\frac{9}{2} & \frac{13}{2} \end{bmatrix} = \begin{bmatrix} \frac{13}{44} & -\frac{9}{44} \\ -\frac{9}{44} & \frac{13}{44} \end{bmatrix}$$

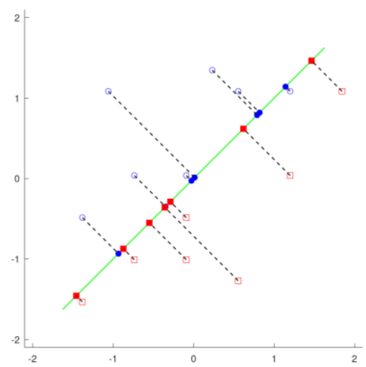
$$w = Sw^{-1}(\mu_1 - \mu_2) = \begin{bmatrix} \frac{13}{44} & -\frac{9}{44} \\ -\frac{9}{44} & \frac{13}{44} \end{bmatrix} \begin{bmatrix} -15 \\ -10 \end{bmatrix} = \begin{bmatrix} -\frac{105}{44} \\ \frac{5}{44} \end{bmatrix}$$

Θέμα 3.

Ο σκοπός της άσκησης είναι να εφαρμοστεί ο LDA για να μειωθεί η διάσταση ενος feature vector για να συγκριθούν τα αποτελέσματα με την μέθοδο PCA. Ο LDA είναι supervised ενώ ο PCA είναι unsupervised και αγνοεί τα class labels.



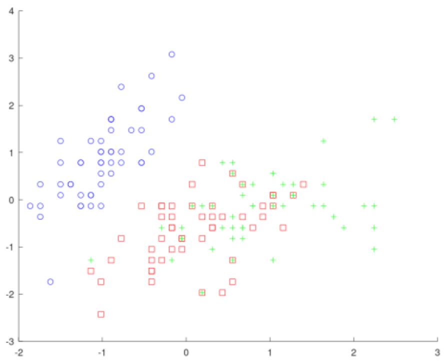
LDA



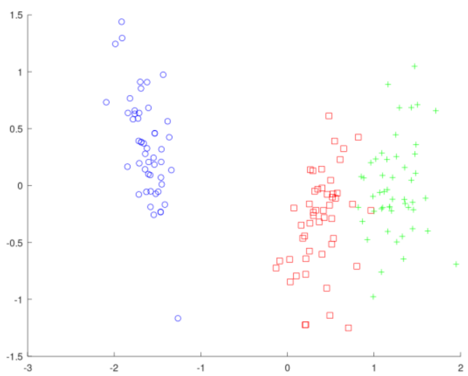
PCA

Στο συγκεκριμένο παράδειγμα βλέπουμε ο LDA να συμπεριφέρεται καλυτέρα διότι χωρίζει τις δυο κλάσεις σωστά , σε αντίθεση με τον PCA να τοποθετεί σε μια διάσταση όπου οι δυο κλάσεις επικαλύπτουν η μια την άλλη.

Για το μέρος δεύτερο χρησιμοποιήσαμε το fisheriris dataset για να εφαρμόσουμε το LDA σε τρεις κλάσεις του dataset και στο τέλος σχεδιάσαμε τα αποτελέσματα του LDA.



Στο παραπάνω παράδειγμα βλέπουμε τρεις διαφορετικές κλάσεις. Τα στοιχειά αυτών είναι τοποθετημένα με βάση τα 2 από τα 6 χαρακτηριστικά τους. Με το myLDA προσπαθούμε να μειώσουμε τις διαστάσεις από 6 σε 2 ψάχνοντας για χαρακτηριστικά τα οποία δε συσχετίζονται πολύ για να μπορέσουν να διαχωρίσουν τις κλάσεις.



Στην παραπάνω εικόνα βλέπουμε την εικόνα μετά την επεξεργασία του LDA όπου μπορεί να διακριθεί ότι το νέο διάνυσμα χαρακτηριστικών μπορεί να ταξινομήσει τις κλάσεις. Αξίζει να σημειωθεί ότι ακόμη και μετά την επεξεργασία δε υπάρχει 100% διαχωρισμός των κλάσεων διότι υπάρχει ακόμη correlation μεταξύ των κλάσεων γεγονός που αναδεικνύει το σφάλμα.

Θέμα 4.

Στο συγκεκριμένο θέμα έπρεπε να υλοποιήσουμε έναν classifier με τον κανόνα του Bayes στο mnist dataset , ενα σετ δεδομένων που περιέχει χειρόγραφες εικόνες ψηφίων. Εμείς ασχοληθήκαμε με τα ψηφία 1,2. Για να κάνουμε την ταξινόμηση χρησιμοποιήθηκε το aspect ratio που είναι ο λόγος του πλάτους προς του ύψος του κάθε ψηφιού, με ελάχιστο το 0.1 και με μέγιστο το 2.22 . Το διάστημα [min max] χρειάστηκε να το χωρίσουμε στα 3 (Low ,Medium, High) για να μπορέσουμε να ταξινομήσουμε τα στοιχειά των δυο κλάσεων.

- Στο Low ανήκουν :6393 από το C1 , 1473 από το C2
- Στο Medium ανήκουν : 349 από το C1 , 4402 από το C2
- Στο High ανήκουν : 0 από το C1 , 83 από το C2

P(C1) = $\frac{C1}{\Omega} = \frac{size\,(C1)}{size\,(C1+C2)} = \frac{6742}{12710} = 0.53$

P(C2) = $\frac{C2}{\Omega} = \frac{size\,(C2)}{size\,(C1+C2)} = \frac{5968}{12710} = 0.47$

Έπειτα υπολογίσαμε τις πιθανότητες των Low ,Medium, High δεδομένου οτι ανήκει στο C1 και αντίστοιχα οτι ανήκει στο C2.

P(L|C1) = $\frac{P(L\cap C1)}{P(C1)} = \frac{size\,(C1\_bin\,(1))}{size\,(C1)} = \frac{6393}{6742} = 0.95$

P(L|C2) = $\frac{P(L\cap C2)}{P(C2)} = \frac{size\,(C2\_bin\,(1))}{size\,(C2)} = \frac{1473}{5958} = 0.25$

P(M|C1) = $\frac{P(M\cap C1)}{P(C1)} = \frac{size\,(C1\_bin\,(2))}{size\,(C1)} = \frac{349}{6742} = 0.05$

P(M|C2) = $\frac{P(M\cap C2)}{P(C2)} = \frac{size\,(C2\_bin(2))}{size\,(C2)} = \frac{4402}{5958} = 0.74$

P(H|C1) = $\frac{P(H \cap C1)}{P(C1)} = \frac{size\ (C1\_bin\ (3))}{size\ (C1)} = \frac{0}{6742} = 0.0$

P(H|C2) = $\frac{P(H \cap C2)}{P(C2)} = \frac{size\ (C2\_bin\ (3))}{size\ (C2)} = \frac{83}{5958} = 0.01$

Ενώ υπολογίσαμε και ολόκληρες τις πιθανότητες P(L),P(M),P(H) .

P(L)=$P(L \cap C1) + P(L \cap C2) = P(L|C1) * P(C1) + P(L|C2) * P(C2) = 0.61937$

P(M)=$P(M \cap C1) + P(M \cap C2) = P(M|C1) * P(C1) + P(M|C2) * P(C2) = 0.37410$

P(H)=$P(H \cap C1) + P(H \cap C2) = P(H|C1) * P(C1) + P(H|C2) * P(C2) = 0.0065354$

Τέλος υπολογίσαμε το P(C1|L) , P(C2|L) όπου προφανές βγαίνει P(C1|L)= P(L|C1)*P(C1)/P(L)= 0.81 και P(C2|L) = P(L|C2)*P(C2)/P(L)= 0.19.

Θέμα 5.

Εφόσον έχουμε ένα πρόβλημα κατηγοριοποίησης με δυο κλάσεις θα ισχύει :

- x=ω1 if P(ω1|x) > P(ω2|x)
- x=ω2 if P(ω2|x) > P(ω1|x)

Το σύνορο απόφασης μπορεί να βρεθεί από την ισότητα P(ω1|x) = P(ω2|x).

$$P(w1|x) = \frac{P(w1 \cap x)}{P(x)} = \frac{P(x|w1)P(w1)}{P(x)}$$

$$P(w2|x) = \frac{P(w2 \cap x)}{P(x)} = \frac{P(x|w2)P(w2)}{P(x)}$$

$$P(x|w1) = \frac{1}{\left(2\pi^{\frac{1}{2}}\right) * |\Sigma|^{\frac{1}{2}}} * \exp(-\frac{1}{2}(x-\mu_1)^T * \sum_i^{-1}(x-\ \mu_1\ ))$$

Ομοίως το $P(x|w2)$, και |Σ1|=|Σ2| = 1.28

P(ω1|x) = P(ω2|x)

$$P(w1) * \exp(-\frac{1}{2} * [x1-3 \quad x2-3] * \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{bmatrix} * \frac{1}{1.28} \begin{bmatrix} x1-3 \\ x2-3 \end{bmatrix})$$

$$=$$

$$P(w2) * \exp(-\frac{1}{2} * [x1-6 \quad x2-6] * \begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix} * \frac{1}{1.28} \begin{bmatrix} x1-6 \\ x2-6 \end{bmatrix})$$

$$\leftrightarrow$$

$$\ln(P(w1)) - \frac{1}{2} * [x1-3 \quad x2-3] * \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{bmatrix} * \frac{1}{1.28} \begin{bmatrix} x1-3 \\ x2-3 \end{bmatrix}$$

$$=$$

$$\ln(P(w2)) - \frac{1}{2} * [x1-6 \quad x2-6] * \begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix} * \frac{1}{1.28} \begin{bmatrix} x1-6 \\ x2-6 \end{bmatrix}$$

$$\leftrightarrow$$

$$[x1-3 \quad x2-3] * \begin{bmatrix} 0.468 & 0.156 \\ 0.156 & 0.468 \end{bmatrix} \begin{bmatrix} x1-3 \\ x2-3 \end{bmatrix} - [x1-6 \quad x2-6] * \begin{bmatrix} 0.468 & -0.156 \\ -0.156 & 0.468 \end{bmatrix} \begin{bmatrix} x1-6 \\ x2-6 \end{bmatrix}$$

$$=$$

$$\ln(\frac{P(w1)}{P(w2)})$$

Για Σ1 = Σ2 = $\begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{bmatrix}$ τότε η σχέση θα διαμορφωθεί

$$(\sum_1^{-1}\mu 1)^T * \chi + \ln(P(w1)) - \frac{1}{2}\mu 1^T * \sum_1^{-1}\mu 1 = (\sum_2^{-1}\mu 2)^T * \chi + \ln(P(w2)) - \frac{1}{2}\mu 2^T * \sum_2^{-1}\mu 2$$

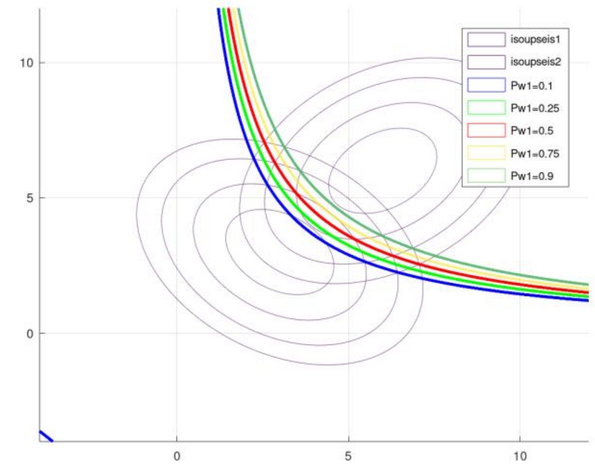
$$\leftrightarrow$$

$$[3 \quad 3] * \begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix} * \frac{1}{1.28} \begin{bmatrix} \chi 1 \\ \chi 2 \end{bmatrix} + \ln(P(w1)) - \frac{1}{2} * [3 \quad 3] * \begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix} * \frac{1}{1.28} \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

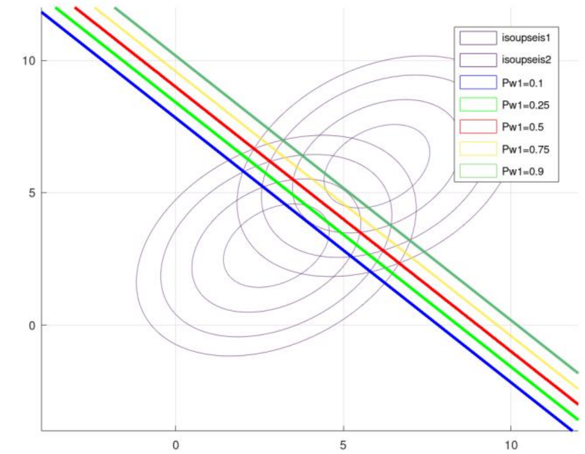
$$=$$

$$[6 \quad 6] * \begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix} * \frac{1}{1.28} \begin{bmatrix} \chi 1 \\ \chi 2 \end{bmatrix} + \ln(P(w2)) - \frac{1}{2} * [6 \quad 6] * \begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix} * \frac{1}{1.28} \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

Για Σ1 = $\begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix}$ και Σ2 = $\begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{bmatrix}$



Για Σ1 = Σ2 = $\begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{bmatrix}$



Είναι εύκολο να παρατηρηθεί ,ότι όσο η P(ω1) αυξάνεται το σύνορο απόφασης πλησιάζει το μέσο της ω2 , μεγαλώνοντας έτσι την περιοχή απόφασης της ω1.

Θέμα 6.

The losses of miss classification are

$$l1 = l11 * P(x|w1) * P(w1) + l21 * P(x|w2) * P(w2) \ (1)$$

$$l2 = l12 * P(x|w1) * P(w1) + l22 * P(x|w2) * P(w2) \ (2)$$

Για να ελαχιστοποιήσουμε το ρίσκο θα πρέπει να λύσουμε το

$$l1 = l2$$

Επειδή $l11 = 0 \ και \ l22 = 0 \ τοτε \ οι \ (1) και \ (2) \ γινονται:$

$l1 = l21 * P(x|w2) * P(w2)$, $l2 = l12 * P(x|w1) * P(w1)$

Άρα

$$l21 * P(x|w2) * P(w2) = l12 * P(x|w1) * P(w1)$$

$$Επειδη \ P(w1) = P(w2)$$

$$l21 * P(x|w2) = l12 * P(x|w1)$$

$$\frac{P(x|w1)}{P(x|w2)} = \frac{l21}{l12}$$

$$\frac{x * e^{-\frac{xo^2}{2}}}{\frac{xo}{4} * e^{-\frac{xo^2}{8}}} = 2$$

$$ln\left(4e^{-\frac{3xo^2}{8}}\right) = ln(2)$$

$$\ln(4) - \frac{3xo^2}{8} = \ln(2)$$

$$xo = 1.35$$