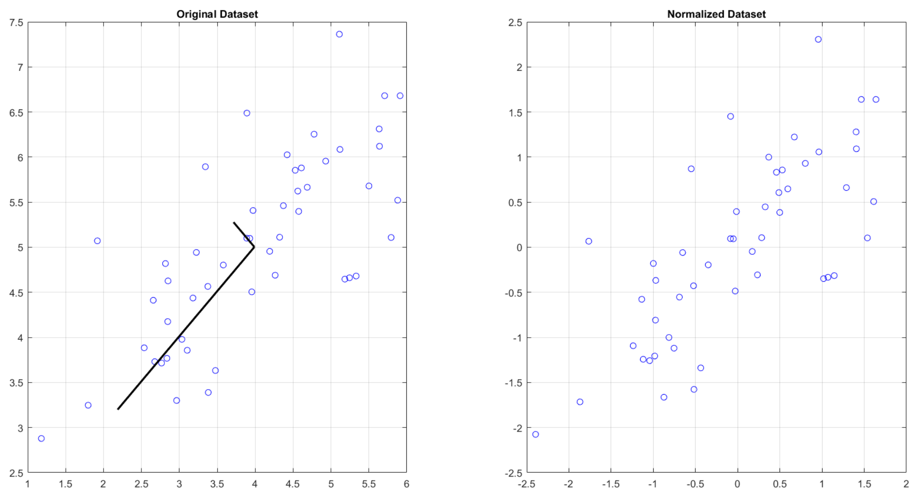
1η Σειρά Ασκήσεων

Ονομα: Μαυρογιώργης Δημήτρης
ΑΜ: 2016030016
ΤΗΛ 311 - Στατιστική Μοντελοποίηση και Αναγνώριση Προτύπων
Πολυτεχνειο Κρητης

May 22, 2021

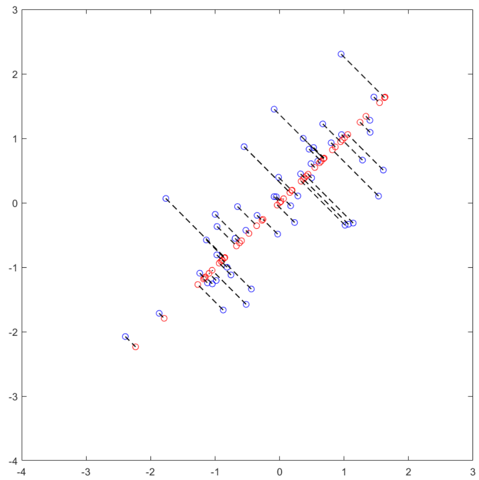
Άσκηση 1: Principal Component Analysis (PCA)

Στο πρώτο μέρος της άσκησης κλήθήκαμε να υλοποιήσουμε τη μέθοδο PCA, για να μειώσουμε ις διαστάσεις των δεδομένων. Αρχικά, αφού διαβάστηκαν τα δεδομένα από το αρχείο 'ex1_1_data1.mat', έγινε μία κανονικοποίηση έτσι, ώστε να έχουν μέση τιμή 0 και συνδιασπορά 1. Για το λόγο αυτό συμπληρώθηκε ο κατάλληλος κώδικας στο αρχείο 'featureNormalize.m', όπου τα καινούρια δεδομένα υπολογίζονται με βάση τον τύπο $x_{new} = \frac{x_{old} - \mu}{\sigma}$. Τα αποτελέσματα του normalization παρουσιάζονται παρακάτω.



Στην συνέχεια, αφού γίνει η κανονικοποίηση των δεδομένων, πρέπει να υπολογίσουμε τις κύριες συνιστώσες του αλγορίθμου PCA (απεικονίζονται στο ίδιο subplot με το original dataset). Έτσι, στο αρχείο 'myPCA.m' υπολογίζουμε τον πίνακα συνδιασποράς με βάση τον τύπο $\sum \frac{1}{m} X^T X$ όπου $X \in \mathbb{R}^{m \times n}$ είναι ένας πίνακας όπου κάθε γραμμή αντιστοιχεί σε ένα παράδειγμα με n χαρακτηριστικά. Τέλος, υπολογίζουμε τα ιδιοδιανύσματα και τις ιδιοτιμές του covariance matrix με τη χρήση της συνάρτησης svd() της Matlab.
Κατόπιν, υπολογίζουμε την συνεισφορά που έχει κάθε κύρια συνιστώσα στην συνολική διακύμανση και έπειτα να εφαρμόζει τον αλγόριθμο PCA στα αρχικά δείγματα για να μειωθεί η διάστασή τους από 2D σε 1D. Γι' το λόγο αυτό στο αρχείο 'projectData.m' συμπληρώθηκε ο κατάλληλος κώδικας,

με τον οποίο προβάλλουμε κάθε δείγμα x_i στις K κύριες συνιστώσες με την εφαρμογή του γραμμικού μετασχηματισμού $z_i = U^T x_i$.
Μετά την προβολή των δεδομένων μπορούμε να ανακτήσουμε τα προσεγγιστικά δεδομένα επαναφέροντας τη διάστασή τους ξανά σε 2D. Έτσι, στο αρχείο 'recoverData.m' προβάλουμε τα δεδομένα πάνω σε όλες τις κύριες συνιστώσες σύμφωνα με το μετασχηματισμό $x_{i,rec} = Uz_i$. Τα αποτελέσματα που προέκυψαν είναι τα εξής μετά από την παραπάνω διαδιακασία:

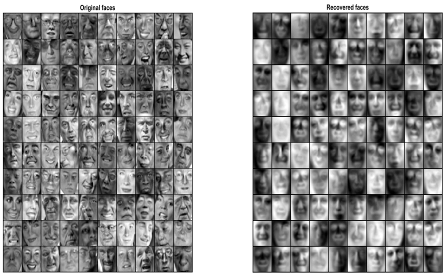


Στο δεύτερο μέρος της άσκησης κληθήκαμε να εφαρμόσουμε τη μέθοδο PCA σε ένα πραγματικό dataset από εικόνες διάφορων προσώπων. Αρχικά, αφού διαβάσουμε τις 100 πρώτες εικόνες και τις εμφανίσουμε με τη χρήση της συνάρτησης displayData() που η υλοποίησή της μας δίνεται έτοιμη, κάνουμε normalization των δειγμάτων, υπολογίζουμε τις κύριες συνιστώσες με τη συνάρτηση myPCA() και εμφανίζουμε τις 36 πρώτες κύριες συνιστώσες. Αυτό που παρατηρούμε είναι ότι απεικονίζονται τα 36 πρόσωπα των ιδιοδιανυσμάτων που υπολογίστηκαν από τη συνάρτηση myPCA().

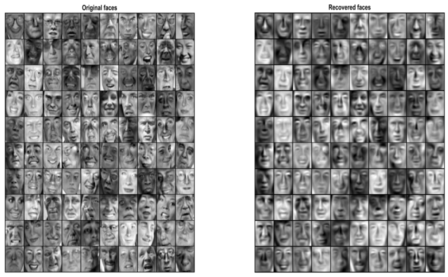
Κατόπιν, μειώνουμε τη διάσταση των δειγμάτων χρησιμοποιώντας τις 100 πρώτες συνιστώσες. Τέλος, εμφανίζουμε τα δείγματα μειωμένης διάστασης αφού πρώτα προβάλουμε στον αρχικό δειγματικό χώρο. Τα αποτελέσματα που προκύπτουν για K=100 είναι τα παρακάτω:

Figure 1: K=100 κύριες συνιστώσες

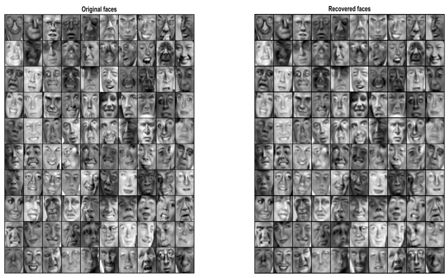
Αυτό που παρατηρούμε είναι ότι οι ανακτημένες εικόνες είναι πιο θολες σε σχέση με τις αρχικές, δηλαδή υπάρχει κάποια μικρή απώλεια πληροφορίας. Τέλος, επαναλαμβάνουμε την παραπάνω διαδικασία για διάφορες τιμές των κύριων συνιστωσών. Παρακάτω βλέπουμε τις εικόνες που ανακτούμε για K=10, 50 και 200 κύριες συνιστώσες.



(a) K=10 κύριες συνιστώσες



(b) K=50 κύριες συνιστώσες



(c) K=200 κύριες συνιστώσες

Στην περίπτωση που δικιμάσουμε για K=10, 50 και 200 παρατηρούμε ότι όσο μειώνονται οι κύριες συνιστωσες οι εικόνες που ανακτούμε γίνονται ακόμη πιο θολές. Αντίθετα, αν αυξήσουμε τις κύριες συνιστώσες βλέπουμε ότι βελτιώνονται τα αποτελέσματα, δηλαδή έχουμε λιγότερη απώλεια πληροφορίας στις εικόνες των προσώπων.

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

Έστω δύο ισοπίθανες κλάσεις ω_1, ω_2 , οι κατανομές των οποίων είναι Γκαουσιανές. Οι πίνακες συνδιασποράς και οι μέσες τιμές έχουν εκτιμηθεί ως εξής:

$$\mu_1 = \begin{bmatrix} -5 \\ 5 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} 10 \\ 15 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 11 & 9 \\ 9 & 11 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Το διάνυσμα προβολής w υπολογίζεται με βάση τον τύπο

$$w = S_w^{-1} \cdot (\mu_1 - \mu_2), \quad \text{όπου} \quad S_w = \frac{1}{2}\sum_{i=1}^{n} \Sigma_i$$

$$S_w = \frac{1}{2}\begin{bmatrix} 11 & 9 \\ 9 & 11 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 13 & 9 \\ 9 & 13 \end{bmatrix} = \begin{bmatrix} 13/2 & 9/2 \\ 9/2 & 13/2 \end{bmatrix}$$

$$S_w^{-1} = \frac{1}{\frac{13}{2}\frac{13}{2} - \frac{9}{2}\frac{9}{2}}\begin{bmatrix} 13/2 & -9/2 \\ -9/2 & 13/2 \end{bmatrix} == \frac{1}{22}\begin{bmatrix} 13/2 & -9/2 \\ -9/2 & 13/2 \end{bmatrix} = \begin{bmatrix} 13/44 & -9/44 \\ -9/44 & 13/44 \end{bmatrix}$$

$$\mu = \mu_1 - \mu_2 = \begin{bmatrix} -5 \\ 5 \end{bmatrix} - \begin{bmatrix} 10 \\ 15 \end{bmatrix} = \begin{bmatrix} 10 \\ 15 \end{bmatrix} = \begin{bmatrix} -13 \\ -10 \end{bmatrix}$$

Συνεπώς, καταλήγουμε στο διάνυσμα προβολής

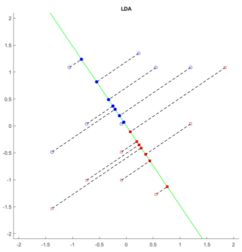
$$w = S_w^{-1} \cdot \mu = \begin{bmatrix} 13/44 & -9/44 \\ -9/44 & 13/44 \end{bmatrix} \cdot \begin{bmatrix} -13 \\ -10 \end{bmatrix} = \begin{bmatrix} (-13 \cdot 13 + 9 \cdot 10)/44 \\ (9 \cdot 15 - 13 \cdot 10)/44 \end{bmatrix} = \begin{bmatrix} -105/44 \\ 5/44 \end{bmatrix}$$

Άσκηση 3: Linear Discriminant Analysis (LDA) vs PCA

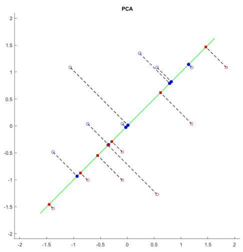
Ο σκοπός του πρώτου μέρους της άσκησης είναι εφαρμόσουμε Linear Discriminant Analysis έτσι, ώστε να μειώσουμε τη διάσταση ενός feature vector και να συγκρίνουμε τα αποτελέσματα αυτά με τη μέθοδο PCA.
Αρχικά, διαβάζουμε τα δεδομένα και εφαρμόζουμε μία κανονικοποίηση με μέση τιμή 0 και διασπορά 1. Για το normalization χρησιμοποιήθηκε η συνάρτηση featureNormalize όπως έγινε και στην άσκηση 1.

Στη συνέχεια, για την υλοποίηση του fisherLinearDiscriminant υπολογίζουμε τις μέσες τιμές και τις διασπορές των δύο κλάσεων και στη συνέχεια με βάση τους τύπους που χρησιμοποιήθηκαν στην άσκηση 2 υπολογίζουμε το within class S_w, το διάνυσμα προβολής w και το κανονικοποιημένο διάνυσμα προβολής w.

Για να μειώσουμε τη διάσταση των δεδομένων σε 1D προστέθηκε ο κατάλληλος κώδικας στο αρχείο 'projectLDA.m' όπου στο συγκεκριμένο απλώς υπολογίζουμε τον πίνακα δεδομένων $Z = X \cdot v$. Τέλος, στο αρχείο 'recoverLDA.m' κάνουμε ανακατασκευή των δειγμάτων μειωμένης διάστασης στον δισδιάστατο χώρο προβάλλοντας τα πάνω στην κατεύθυνση του διανύσματος προβολής LDA. Οπότε προκύπτουν τα recover δείγματα με βάση τον τύπο $x_{rec} = Z \cdot v^T$. Στα ίδια δείγματα εφαρμόστηκε ο αλγόριθμος PCA που υλοποιήθηκε στην άσκηση 1 οπότε έχουμε τα εξής αποτελέσματα έπειτα από την εφαρμογή του, καθώς και τα αποτελέσματα του αλγορίθμου LDA φαίνονται παρακάτω



(a) LDA algorithm

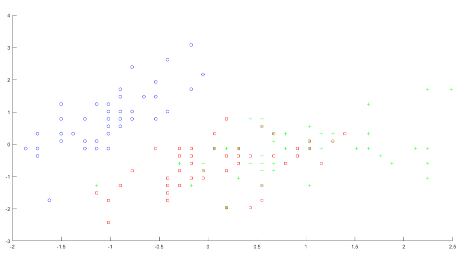


(b) PCA algorithm

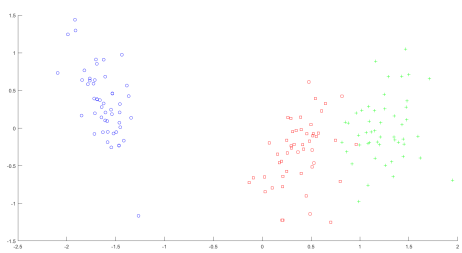
Αυτό που παρατηρούμε από τα παραπάνω αποτελέσματα των δύο αργορίθμων είναι ότι ο αλγόριθμος LDA διαχωρίζει καλύτερα τα δεδομένα σε 2 κλάσεις χωρίς να υπάρχει επικάλυψη της μίας κλάσης με την άλλη. Αντίθετα, βλέπουμε ότι στον αλγόριθμο PCA, τα δείγματα της μίας κλάσης επικαλύπτονται με της άλλης και, επομένως, δεν υπάρχει ξεκάθαρος διαχωρισμός των δειγμάτων σε δύο διαφορετικές κλάσεις.

Στο δεύτερο μέρος της άσκησης εφαρμόζουμε τον αλγόριθμο LDA στη βάση δεδομένων Iris, η οποία περιέχει δείγματα από 3 διαφορετικά είδη της οικογεένειας λουλουδιών Iris.

Όπως και στο πρώτο μέρος εφαρμόστηκε πριν τον αλγόριθμο μια κανονικοποίηση των δειγμάτων. Στη συνέχεια, στη συνάρτηση myLDA() υπολογίζουμε τις prior πιθανότητες κάθε κλάσης, τις μέσες τιμές, τον ολικό μέσο, τους πίνακες within-class και between-class και τον πίνακα $S_w^{-1} \cdot S_b$ του γενικευμένου συστήματος ιδιοτιμών, στον οποίο εφαρμόζετε eigendecomposition. Τέλος, εφαρμόζουμε τα διανύσματα προβολής πάνω στα αρχικά δείγματα με τη συνάρτηση projectDataLDA έτσι ώστε να μειώσουμε τη διάσταση του σε 2D. Τα αποτελέσματα που προέκυψαν είναι τα εξής:



(a) Dataset



(b) LDA algorithm

Αυτό που προκύπτει ως συμπέρασμα από την εφαρμογή του αλγορίθμου LDA στο dataset είναι ότι υπάρχει διαχωρισμός των δειγμάτων σε 3 διαφορετικές κλάσεις. Ωστόσο, μεταξύ της κόκκινης κλάσης και της πράσινης δεν υπάρχει πλήρης διαχωρισμός, καθώς βλέπουμε ότι κάποια ελάχιστα δείγματα της κόκκινης κλάσης εμπεριέχονται ή είναι αρκετά κοντά στην πράσινη κλάση.

Άσκηση 4: Bayes

Έστω δύο κλάσεις ω_1 και ω_2 των οποίων οι εκ των προτέρων πιθανότητες είναι $\mathrm{P}(\omega_1)$ και $\mathrm{P}(\omega_2)$ αντίστοιχα. Τα δείγματα x που πρέπει να κατηγοριοποιηθούν είναι δισδιάστατα και οι κλάσεις περιγράφονται από τις ακόλουθες κανονικές κατανομές:

$$P(x|\omega_1) = N(\mu_1, \Sigma_1), \qquad P(x|\omega_2) = N(\mu_2, \Sigma_2)$$

$$\mu_1 = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$
$$\Sigma_1 = \begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{bmatrix}$$

Με βάση τον κανόνα απόφασης έχουμε τα εξής:

- $\mathrm{x} = \omega_1$ αν $\mathrm{P}(\omega_1 \mid \mathrm{x}) > \mathrm{P}(\ \omega_2 \mid \mathrm{x})$
- $\mathrm{x} = \omega_2$ αν $\mathrm{P}(\omega_1 \mid \mathrm{x}) < \mathrm{P}(\ \omega_2 \mid \mathrm{x})$

Για να βρούμε το όριο απόφασης, αρκεί να λύσουμε την παρακάτω εξίσωση

$$\begin{gathered} P(\omega_1|x) = P(\omega_2|x) \Rightarrow \\ \frac{P(x|\omega_1)\cdot P(\omega_1)}{P(x)} = \frac{P(x|\omega_2)\cdot P(\omega_2)}{P(x)} \Rightarrow \\ P(x|\omega_1)\cdot P(\omega_1) = P(x|\omega_2)\cdot P(\omega_2) \end{gathered}$$

Επιπλέον, γνωρίζουμε ότι

$$P(x|\omega_i) = \frac{1}{\sqrt{2\pi}|\Sigma_i|} e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)}$$

Όσον αφορά τις ορίζουσες των πινάκων συνδιασποράς έχουμε ότι
$|\Sigma_1| = 1.2^2 - (-0.4)^2 = 1.28 \quad$ και $\quad |\Sigma_2| = 1.2^2 - 0.4^2 = 1.28$

$$\begin{gathered} P(x|\omega_1)\cdot P(\omega_1) = P(x|\omega_2)\cdot P(\omega_2) \\ \frac{1}{\sqrt{2\pi}|\Sigma_1|} e^{-\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)} \cdot P(\omega_1) = \frac{1}{\sqrt{2\pi}|\Sigma_2|} e^{-\frac{1}{2}(x-\mu_2)^T \Sigma_i^{-1}(x-\mu_2)} \cdot P(\omega_2) \\ e^{-\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)} \cdot P(\omega_1) = e^{-\frac{1}{2}(x-\mu_2)^T \Sigma_i^{-1}(x-\mu_2)} \cdot P(\omega_2) \\ -\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) + \ln(P(\omega_1)) = -\frac{1}{2}(x-\mu_2)^T \Sigma_i^{-1}(x-\mu_2) + \ln(P(\omega_2)) \end{gathered}$$

$$\ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right) = \frac{1}{2}\begin{bmatrix} x_1 - 3 & x_2 - 3 \end{bmatrix} \frac{1}{1.28} \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{bmatrix} \begin{bmatrix} x_1 - 3 \\ x_2 - 3 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} x_1 - 6 & x_2 - 6 \end{bmatrix} \frac{1}{1.28} \begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix} \begin{bmatrix} x_1 - 6 \\ x_2 - 6 \end{bmatrix}$$
$$\ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right) = \begin{bmatrix} x_1 - 3 & x_2 - 3 \end{bmatrix} \begin{bmatrix} 0.46875 & 0.15625 \\ 0.15625 & 0.46875 \end{bmatrix} \begin{bmatrix} x_1 - 3 \\ x_2 - 3 \end{bmatrix} - \begin{bmatrix} x_1 - 6 & x_2 - 6 \end{bmatrix} \begin{bmatrix} 0.46875 & -0.15625 \\ -0.15625 & 0.46875 \end{bmatrix} \begin{bmatrix} x_1 - 6 \\ x_2 - 6 \end{bmatrix}$$

$$\ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right) = (x1-3)\cdot(0.46875\cdot x1 + 0.15625\cdot x2 - 1.875) + (x2-3)\cdot(0.15625\cdot x1 + 0.46875\cdot x2 - 1.875) - (x1-6)\cdot(0.46875\cdot x1 - 0.15625\cdot x2 - 1.875) - (x2-6)\cdot(0.46875\cdot x2 - 0.15625\cdot x1 - 1.875)$$

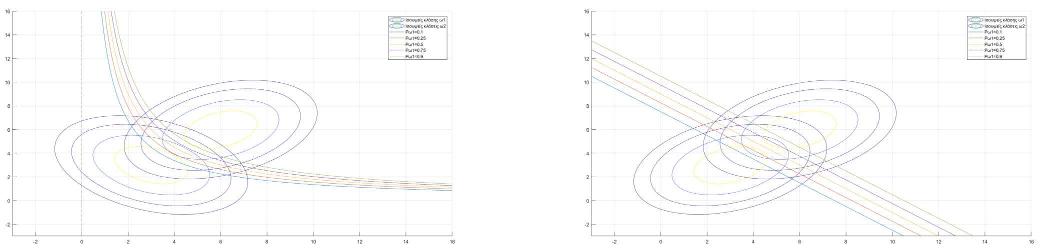
Στην περίπτωση που έχουμε ίδιους πίνακες συνδιασποράς προκύπτουν τα εξής:

$$\ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right) = \frac{1}{2}\begin{bmatrix} x_1-3 & x_2-3 \end{bmatrix}\frac{1}{1.28}\begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix}\begin{bmatrix} x_1-3 \\ x_2-3 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} x_1-6 & x_2-6 \end{bmatrix}\frac{1}{1.28}\begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix}\begin{bmatrix} x_1-6 \\ x_2-6 \end{bmatrix}$$

$$\ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right) = \begin{bmatrix} x_1-3 & x_2-3 \end{bmatrix}\begin{bmatrix} 0.46875 & -0.15625 \\ -0.15625 & 0.46875 \end{bmatrix}\begin{bmatrix} x_1-3 \\ x_2-3 \end{bmatrix} - \begin{bmatrix} x_1-6 & x_2-6 \end{bmatrix}\begin{bmatrix} 0.46875 & -0.15625 \\ -0.15625 & 0.46875 \end{bmatrix}\begin{bmatrix} x_1-6 \\ x_2-6 \end{bmatrix}$$

$$\ln\left(\frac{P(\omega_1)}{P(\omega_2)}\right) = (x1-3)\cdot(0.46875\cdot x1 - 0.15625\cdot x2 - 0.9375) + (x2-3)\cdot(0.46875\cdot x2 - 0.15625\cdot x1 - 0.9375) - (x1-6)\cdot(0.46875\cdot x1 - 0.15625\cdot x2 - 1.875) - (x2-6)\cdot(0.46875\cdot x2 - 0.15625\cdot x1 - 1.875)$$

Παρακάτω φαίνονται οι ισοϋψείς καμπύλες, καθώς και τα όρια απόφασης για δεδομένες πιθανότητες $P(\omega_1) = 0.1,\ 0.25,\ 0.5,\ 0.75,\ 0.9$.



(a) Για διαφορετικού πίνακες συνδιασποράς

(b) Για ίδιους πίνακες συνδιασποράς

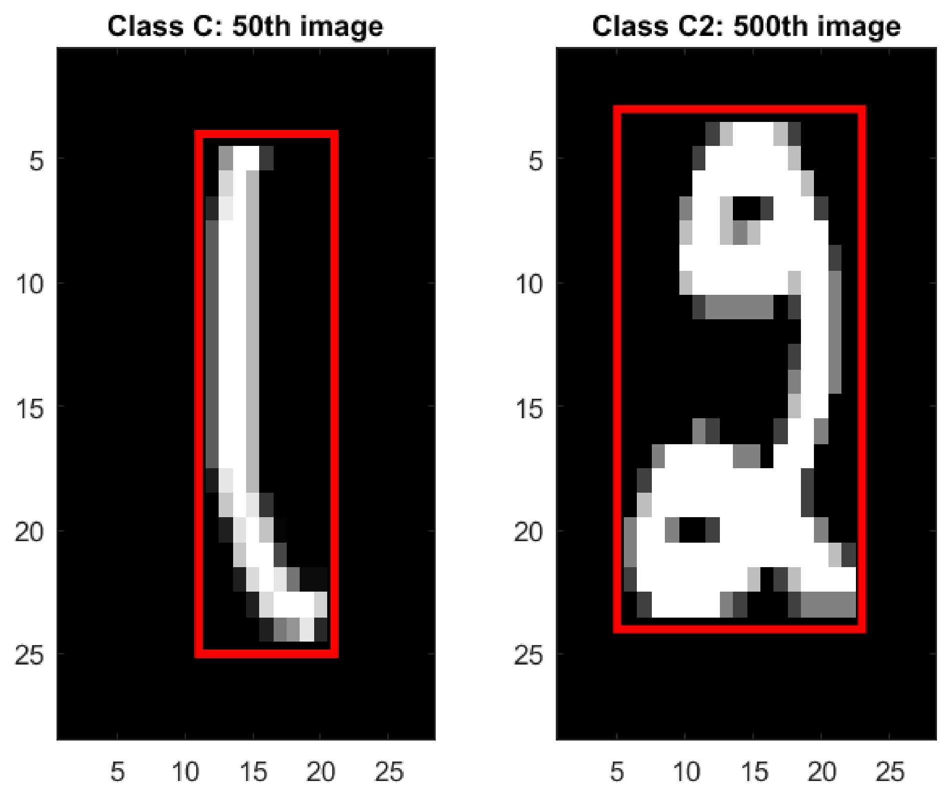
Από τις παραπάνω εικόνες παρατηρούμε ότι τα σύνορα απόφασης για διαφορετικούς πίνακες συνδιασποράς έχουν τη μορφή υπερβολής και όσο αυξάνεται η πιθανότητα $\mathrm{P}(\omega_1)$ τόσο μεγαλώνει η περιοχή απόφασης της κλάσης ω_1 και τα σύνορα απόφασης πλησιάζουν προς το κέντρο της Γκαουσιανής της κλάσης ω_2. Στην περίπτωση που έχουμε ίδιους πίνακες συνδιασποράς τότε η μορφή του ορίου απόφασης γίνεται γραμμική, ενώ πάλι με την αύξηση της $\mathrm{P}(\omega_1)$ παρατηρούμε το ίδιο με την περίπτωση όπου είχαμε διαφορετικούς πίνακες συνδιασποράς για τις δύο κλάσεις.

Άσκηση 5: Εξαγωγή χαρακτηριστικών και Bayes Classification

Σκοπός αυτής της άσκησης είναι να δημιουργήσουμε έναν απλό ταξινομητή Bayes ο οποίος θα ταξινομεί δείγματα σε δύο κλάσεις και συγκεκριμένα στις κλάσεις C_1 και C_2 των ψηφίων 1 και 2 αντίστοιχα που βρίσκονται στο αρχείο 'minst.mat'.

Αρχικά, δημιουργήθηκε μία συνάρτηση computeAspectRatio() με την οποία υπολογίζουμε το ελάχιστο παραλληλόγραμμο που περικλύει το κάθε ψηφίο που απεικονίζεται στις εικόνες. Ειδικότερα, με αυτό που υπολογίζουμε είναι το height και το width των παραλληλογράμμων και επιστρέφουμε το λόγο $\frac{width}{height}$ της κάθε εικόνας.

Στη συνέχεια. υπολογίζουμε για όλες τις εικόνες της κλάσης C_1 και C_2 το aspect ratio με τη συνάρτηση που υλοποιήσαμε. Το ελάχιστο aspect ratio βρέθηκε ότι είναι 0.0526, ενώ το μέγιστο 2.375. Επιπλέον, παρακάτω παρουσιάζονται η εικόνα 50 από την πρώτη κλάση και η εικόνα 500 από τη 2η κλάση μαζί με τα αντίστοιχα παραλληλόγραμμα που περικλύουν τα ψηφία



Κατόπιν, υπολογίσαμε τις prior πιθανότητες της κάθε κλάσης ως το πηλίκο του αριθμού των εικόνων που έχει η κλάση C_i με τον συνολικό αριθμό εικόνων που έχουν οι δύο κλάσεις μαζί. Οι πιθανότητες της κάθε κλάσης αντίστοιχα είναι:

- $P(C_1) = 0.5309$

- $P(C_2) = 0.4691$

Επιπλέον, η κατανομή του aspect ratio χαρακτηριστικού σε κάθε κλάση ακολουθεί κανονική κατανομή με μέση τιμή και συνδιασπορά

$$\bar{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$$
$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{\mu})^2}$$

Για το classification της κάθε εικόνας στα δείγματα test υπολογίζουμε το aspect ratio με τη συνάρτηση computeAspectRatio() και στη συνέχεια υπολογίζουμε τις δύο likelihood πιθανότητες με βάση τον τύπο $p(x|C_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\frac{(x-\bar{\mu_i})^2}{\sigma_i^2}}$. Αφού υπολογίσουμε αυτές τις δύο πιθανότητες, υπολογίζουμε τις πιθανότητες a-post eriori πολλαπλασιάζοντας με τις prior της αντίστοιχης κλάσης και, τέλος, βλέπουμε ποια είναι μεγαλύτερη. Στην περίπτωση που ισχύει ότι $p(x|C_1) \cdot P(C1) > p(x|C_2) \cdot P(C2)$ αποφασίζουμε ότι το δείγμα x ανήκει στην κλάση C_1, αλλιώς ανήκει στην C_2. Αφού επαναλάβουμε αυτή τη διαδικασία και για τα δύο σύνολα εικόνων test, υπολογίζουμε το συνολικό ποσοστό σφάλματος του ταξινομητή μας ως το πηλίκο του αριθμού των λανθασμένων αποφάσεων που πήρε ο ταξινομητής με το συνολικό αριθμό εικόνων που κάναμε classify. Το συνολικό σφάλμα υπολογίστηκε ότι είναι περίπου 10.94%

Άσκηση 6: Minimum risk

Έστω ένα πρόβλημα κατηγοριοποίησης σε δύο κλάσεις ω_1 και ω_2 όπου οι εκ των προτέρων πιθανότητες είναι ίσες $P(\omega_1) = P(\omega_2)$. Τα δείγματα x που πρέπει να κατηγοριοποιηθούν είναι μονοδιάστατα και ακολουθούν κατανομή Rayleigh με συνάρτηση πυκνότητας πιθανότητας η οποία είναι:

$$p(x|\omega_i) = \begin{cases} \frac{x}{\sigma_i^2} \cdot e^{-\frac{x^2}{2\cdot\sigma_i^2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Επιπλέον, $\sigma_1 = 1$ και $\sigma_2 = 2$ ενώ ο πίνακας ρίσκου είναι

$$L = \begin{bmatrix} 0 & 0.5 \\ 1 & 0 \end{bmatrix}$$

Τότε

$$l_1 = l_{11} \cdot p(\omega_1|x) + l_{21} \cdot p(\omega_2|x) = 0 \cdot p(\omega_1|x) + 1 \cdot p(\omega_2|x) = p(\omega_2|x)$$
$$l_2 = l_{12} \cdot p(\omega_1|x) + l_{22} \cdot p(\omega_2|x) = 0.5 \cdot p(\omega_1|x) + 0 \cdot p(\omega_2|x) = 0.5 \cdot p(\omega_1|x)$$

Για να βρούμε το όριο απόφασης θα πρέπει να λύσουμε την εξίσωση

$$l_1 = l_2$$
$$p(\omega_2|x) = 0.5 \cdot p(\omega_1|x)$$
$$p(\omega_1|x) = 2 \cdot p(\omega_2|x)$$
$$\frac{x}{\sigma_1^2} \cdot e^{-\frac{x^2}{2\cdot\sigma_1^2}} = 2 \cdot \frac{x}{\sigma_2^2} \cdot e^{-\frac{x^2}{2\cdot\sigma_2^2}}$$
$$\frac{x}{1^2} \cdot e^{-\frac{x^2}{2\cdot 1^2}} = 2 \cdot \frac{x}{2^2} \cdot e^{-\frac{x^2}{2\cdot 2^2}}$$

$$
\begin{aligned}
e^{-\frac{x^2}{2}} &= \frac{1}{2} \cdot e^{-\frac{x^2}{8}} \\
2 \cdot e^{-\frac{x^2}{2}} &= e^{-\frac{x^2}{8}} \\
\ln(2) - \frac{x^2}{2} &= -\frac{x^2}{8} \\
\frac{3 \cdot x^2}{8} &= \ln(2) \\
x &= \sqrt{\frac{8 \cdot \ln(2)}{3}} \approx 1.36
\end{aligned}
$$

Άσκηση 7: Singular Value Decomposition (SVD)

Έστω ο παρακάτω πίνακας

$$
X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 3 \end{bmatrix}
$$

Τότε ο πίνακας $X^T X$ είναι ο εξής

$$
\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 7 \\ 7 & 14 \end{bmatrix}
$$

$$
\begin{aligned}
\det(X^T X - \lambda \cdot \mathbb{I}) &= \det\left(\begin{bmatrix} 6 & 7 \\ 7 & 14 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} 6-\lambda & 7 \\ 7 & 14-\lambda \end{bmatrix}\right) \\
&= (6-\lambda) \cdot (14-\lambda) - 49 \\
&= \lambda^2 - 20 \cdot \lambda + 35
\end{aligned}
$$

Η διακρίνουσα του διωνύμου είναι $\Delta = 20^2 - 4 \cdot 1 \cdot 35 = 260$. Οι λύσεις του παραπάνω δυωνύμου είναι:

$$
\lambda_{1,2} = \frac{-(-20) \pm \sqrt{260}}{2 \cdot 1} = \frac{20 \pm 2 \cdot \sqrt{65}}{2} = 10 \pm \sqrt{65}
$$

Συνεπώς, οι ιδιοτιμές του πίνακα $X^T X$ είναι οι παρακάτω

$$
\lambda_1 = 10 + \sqrt{65} = 18.062 \qquad \lambda_2 = 10 - \sqrt{65} = 1.938
$$

Για να βρούμε τα ιδιοδιανύσματα θα πρέπει να λύσουμε την εξίσωση $(X^T X - \lambda \cdot \mathbb{I}) \cdot u_i = 0$.
Για $\lambda = 18.062$ έχουμε

$$
(X^T X - \lambda \cdot \mathbb{I}) \cdot u_1 = 0 \Rightarrow \left(\begin{bmatrix} 6 & 7 \\ 7 & 14 \end{bmatrix} - \begin{bmatrix} 18.062 & 0 \\ 0 & 18.062 \end{bmatrix}\right) \cdot \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} -12.062 & 7 \\ 7 & -4.062 \end{bmatrix} \cdot \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} = 0 \Rightarrow
$$

$$\begin{bmatrix} -12.062 & 7 \\ 7 & -4.062 \end{bmatrix} \cdot \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix} = 0 \Rightarrow \begin{cases} -12.062 \cdot u_{11} + 7 \cdot u_{12} = 0 \\ 7 \cdot u_{11} - 4.062 \cdot u_{12} = 0 \end{cases} \Rightarrow \begin{cases} u_{12} = 1.723 \cdot u_{11} \\ u_{12} = 1.723 \cdot u_{11} \end{cases}$$

Συνεπώς, τα ιδιοδυανίσματα για $\lambda = 18.062$ είναι της μορφής

$$u_1 = k_1 \cdot \begin{bmatrix} 1 \\ 1.723 \end{bmatrix}$$

Επιπλέον, για να έχουν μέτρο μονάδα θα πρέπει να ισχύει

$$k_1^2 + (1.723 \cdot k_1)^2 = 1 \Rightarrow k_1^2 + 2.969 \cdot k_1^2 = 1 \Rightarrow 3.969 \cdot k_1^2 = 1 \Rightarrow k_1 = \sqrt{\frac{1}{3.969}} \approx 0.5019$$

Άρα, καταλλήγουμε στο ιδιοδιάνυσμα

$$u_1 = k_1 \cdot \begin{bmatrix} 1 \\ 1.723 \end{bmatrix} = 0.5019 \cdot \begin{bmatrix} 1 \\ 1.723 \end{bmatrix} = \begin{bmatrix} 0.5019 \\ 0.8648 \end{bmatrix}$$

Για $\lambda = 1.938$ έχουμε

$$(X^T X - \lambda \cdot \mathbb{I}) \cdot u_2 = 0 \Rightarrow \left(\begin{bmatrix} 6 & 7 \\ 7 & 14 \end{bmatrix} - \begin{bmatrix} 1.938 & 0 \\ 0 & 1.938 \end{bmatrix} \right) \cdot \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} 4.062 & 7 \\ 7 & 12.062 \end{bmatrix} \cdot \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} = 0 \Rightarrow$$

$$\begin{bmatrix} 4.062 & 7 \\ 7 & 12.062 \end{bmatrix} \cdot \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} = 0 \Rightarrow \begin{cases} 4.062 \cdot u_{21} + 7 \cdot u_{22} = 0 \\ 7 \cdot u_{21} + 12.062 \cdot u_{22} = 0 \end{cases} \Rightarrow \begin{cases} u_{21} = -1.723 \cdot u_{22} \\ u_{21} = -1.723 \cdot u_{22} \end{cases}$$

Συνεπώς, τα ιδιοδυανίσματα για $\lambda = 1.938$ είναι της μορφής

$$u_2 = k_2 \cdot \begin{bmatrix} -1.723 \\ 1 \end{bmatrix}$$

Επιπλέον, για να έχουν μέτρο μονάδα θα πρέπει να ισχύει

$$(-1.723 \cdot k_2)^2 + k_2^2 = 1 \Rightarrow 2.969 \cdot k_2^2 + k_2^2 = 1 \Rightarrow 3.969 \cdot k_2^2 = 1 \Rightarrow k_2 = \sqrt{\frac{1}{3.969}} \approx 0.5019$$

Άρα, καταλλήγουμε στο ιδιοδιάνυσμα

$$u_2 = k_2 \cdot \begin{bmatrix} 1 \\ 1.723 \end{bmatrix} = 0.5019 \cdot \begin{bmatrix} -1.723 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.8648 \\ 0.5019 \end{bmatrix}$$

Για τον υπολογισμό των singular values έχουμε ότι $\sigma_i = \sqrt{\lambda_i}$

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{18.062} \approx 4.25$$
$$\sigma_2 = \sqrt{\lambda_2} = \sqrt{1.938} \approx 1.392$$

Όσον αφορά τον πίνακα XX^T έχουμε

$$\begin{bmatrix}1 & 2\\ 2 & 1\\ 1 & 3\end{bmatrix} \cdot \begin{bmatrix}1 & 2 & 1\\ 2 & 1 & 3\end{bmatrix} = \begin{bmatrix}5 & 4 & 7\\ 4 & 5 & 5\\ 7 & 5 & 10\end{bmatrix}$$

$$\begin{aligned}\det(X^TX - \lambda \cdot \mathbb{I}) &= \det\left(\begin{bmatrix}5 & 4 & 7\\ 4 & 5 & 5\\ 7 & 5 & 10\end{bmatrix} - \begin{bmatrix}\lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda\end{bmatrix}\right) = \det\left(\begin{bmatrix}5-\lambda & 4 & 7\\ 4 & 5-\lambda & 5\\ 7 & 5 & 10-\lambda\end{bmatrix}\right)\\ &= (5-\lambda)\cdot\det\left(\begin{bmatrix}5-\lambda & 5\\ 5 & 10-\lambda\end{bmatrix}\right) - 4\cdot\det\left(\begin{bmatrix}4 & 5\\ 7 & 10-\lambda\end{bmatrix}\right) + 7\cdot\det\left(\begin{bmatrix}4 & 5-\lambda\\ 7 & 5\end{bmatrix}\right)\\ &= (-\lambda^3 + 20\cdot\lambda^2 - 100\cdot\lambda + 125) + (16\cdot\lambda - 20) + (49\cdot\lambda - 105)\\ &= -\lambda^3 + 20\cdot\lambda^2 - 35\cdot\lambda\\ &= -\lambda\cdot(\lambda^2 - 20\cdot\lambda + 35)\end{aligned}$$

Άρα, βλέπουμε ότι ο πίνακας XX^T έχει τις ίδιες ιδιοτιμές με τον πίνακα X^TX και μία επιπλέον ιδιοτιμή την $\lambda = 0$.

Συνεπώς, οι ιδιοτιμές του πίνακα XX^T είναι οι παρακάτω

$$\lambda_1 = 18.062 \qquad \lambda_2 = 1.938 \qquad \lambda_3 = 0$$

Για $\lambda = 18.062$ έχουμε

$$\begin{aligned}(X^TX - \lambda\cdot\mathbb{I})\cdot v_1 = 0 \Rightarrow \left(\begin{bmatrix}5 & 4 & 7\\ 4 & 5 & 5\\ 7 & 5 & 10\end{bmatrix} - \begin{bmatrix}18.062 & 0 & 0\\ 0 & 18.062 & 0\\ 0 & 0 & 18.062\end{bmatrix}\right)\cdot\begin{bmatrix}v_{11}\\ v_{12}\\ v_{13}\end{bmatrix} &= 0 \Rightarrow\\ \begin{bmatrix}-13.062 & 4 & 7\\ 4 & -13.062 & 5\\ 7 & 5 & -8.062\end{bmatrix}\cdot\begin{bmatrix}v_{11}\\ v_{12}\\ v_{13}\end{bmatrix} &= 0 \Rightarrow\end{aligned}$$

$$\begin{cases}-13.062\cdot v_{11} + 4\cdot v_{12} + 7\cdot v_{13} = 0\\ 4\cdot v_{11} - 13.062\cdot v_{12} + 5\cdot v_{13} = 0\\ 7\cdot v_{11} + 5\cdot v_{12} - 8.062\cdot v_{13} = 0\end{cases} \Rightarrow \begin{cases}v_{11} = 0.306\cdot v_{12} + 0.536\cdot v_{13}\\ v_{11} = 3.2655\cdot v_{12} - 1.25\cdot v_{13}\\ v_{11} = -0.714\cdot v_{12} + 1.152\cdot v_{13}\end{cases}$$

Από τις δύο πρώτες σχέσεις έχουμε ότι

$$\begin{aligned}0.306\cdot v_{12} + 0.536\cdot v_{13} = 3.2655\cdot v_{12} - 1.25\cdot v_{13} \Rightarrow\\ 3.2655\cdot v_{12} - 0.306\cdot v_{12} = 0.536\cdot v_{13} + 1.25\cdot v_{13} \Rightarrow\\ 2.9595\cdot v_{12} = 1.786\cdot v_{13} \Rightarrow\\ v_{12} = 0.6035\cdot v_{13}\end{aligned}$$

Αντικαθιστώντας στην τελευταία σχέση προκύπτει ότι

$$v_{11} = -0.714\cdot 0.6035\cdot v_{13} + 1.152\cdot v_{13} = 0.721\cdot v_{13}$$

Συνεπώς, τα ιδιοδυανίσματα για $\lambda = 18.062$ είναι της μορφής

$$v_1 = k_1 \cdot \begin{bmatrix} 0.721 \\ 0.6035 \\ 1 \end{bmatrix}$$

Επιπλέον, για να έχουν μέτρο μονάδα θα πρέπει να ισχύει

$$(0.721 \cdot k_1)^2 + (0.6035 \cdot k_1)^2 + k_1^2 = 1 \Rightarrow 1.884 \cdot k_1^2 = 1 \Rightarrow k_1 = \sqrt{\frac{1}{1.884}} \approx 0.7285$$

Άρα, καταλλήγουμε στο ιδιοδιάνυσμα

$$v_1 = k_1 \cdot \begin{bmatrix} 0.721 \\ 0.6035 \\ 1 \end{bmatrix} = 0.7285 \cdot \begin{bmatrix} 0.721 \\ 0.6035 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5253 \\ 0.4397 \\ 0.7285 \end{bmatrix}$$

Για $\lambda = 1.938$ έχουμε

$$\begin{aligned}(X^T X - \lambda \cdot \mathbb{I}) \cdot v_2 = 0 \Rightarrow \left(\begin{bmatrix} 5 & 4 & 7 \\ 4 & 5 & 5 \\ 7 & 5 & 10 \end{bmatrix} - \begin{bmatrix} 1.938 & 0 & 0 \\ 0 & 1.938 & 0 \\ 0 & 0 & 1.938 \end{bmatrix} \right) \cdot \begin{bmatrix} v_{21} \\ v_{22} \\ v_{23} \end{bmatrix} &= 0 \Rightarrow \\ \begin{bmatrix} 3.062 & 4 & 7 \\ 4 & 3.062 & 5 \\ 7 & 5 & 8.062 \end{bmatrix} \cdot \begin{bmatrix} v_{21} \\ v_{22} \\ v_{23} \end{bmatrix} &= 0 \Rightarrow\end{aligned}$$

$$\begin{cases} 3.062 \cdot v_{21} + 4 \cdot v_{22} + 7 \cdot v_{23} = 0 \\ 4 \cdot v_{21} + 3.062 \cdot v_{22} + 5 \cdot v_{23} = 0 \\ 7 \cdot v_{21} + 5 \cdot v_{22} + 8.062 \cdot v_{23} = 0 \end{cases} \Rightarrow \begin{cases} v_{21} = -1.306 \cdot v_{22} - 2.286 \cdot v_{23} \\ v_{21} = -0.765 \cdot v_{22} - 1.25 \cdot v_{23} \\ v_{21} = -0.714 \cdot v_{22} - 1.152 \cdot v_{23} \end{cases}$$

Από τις δύο πρώτες σχέσεις έχουμε ότι

$$\begin{aligned} -1.306 \cdot v_{22} - 2.286 \cdot v_{23} &= -0.765 \cdot v_{22} - 1.25 \cdot v_{23} \Rightarrow \\ 1.306 \cdot v_{22} - 0.765 \cdot v_{22} &= -2.286 \cdot v_{23} + 1.25 \cdot v_{23} \Rightarrow \\ 0.541 \cdot v_{22} &= -1.036 \cdot v_{23} \Rightarrow \\ v_{22} &= -1.915 \cdot v_{23} \end{aligned}$$

Αντικαθιστώντας στην τελευταία σχέση προκύπτει ότι

$$v_{21} = -0.714 \cdot (-1.915 \cdot v_{23}) - 1.152 \cdot v_{23} = 0.215 \cdot v_{13}$$

Συνεπώς, τα ιδιοδυανίσματα για $\lambda = 1.938$ είναι της μορφής

$$v_2 = k_2 \cdot \begin{bmatrix} 0.215 \\ -1.915 \\ 1 \end{bmatrix}$$

Επιπλέον, για να έχουν μέτρο μονάδα θα πρέπει να ισχύει

$$(0.215 \cdot k_2)^2 + (-1.915 \cdot k_2)^2 + k_2^2 = 1 \Rightarrow 4.713 \cdot k_2^2 = 1 \Rightarrow k_2 = \sqrt{\frac{1}{4.713}} \approx 0.4606$$

Άρα, καταλλήγουμε στο ιδιοδιάνυσμα

$$v_2 = k_2 \cdot \begin{bmatrix} 0.215 \\ -1.915 \\ 1 \end{bmatrix} = 0.4606 \cdot \begin{bmatrix} 0.215 \\ -1.915 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.099 \\ 0.8821 \\ 0.4606 \end{bmatrix}$$

Αν λάβουμε υπόψη μας μόνο τη μεγαλύτερη ιδιοτιμή $\lambda = 18.062$, τότε το καλύτερο rank-1 approximation προκύπτει ως εξής

$$\begin{aligned} \tilde{X} &= \sigma_1 \cdot u_1 \cdot v_1^T \\ &= 4.25 \cdot \begin{bmatrix} 0.5019 \\ 0.8648 \end{bmatrix} \begin{bmatrix} 0.5253 & 0.4397 & 0.7285 \end{bmatrix} \\ &= \begin{bmatrix} 1.12 & 1.93 \\ 0.93 & 1.616 \\ 1.554 & 2.68 \end{bmatrix} \end{aligned}$$