Pattern Recognition

Exercise 1

Report Delivery Date: 20 April 2018

Student: Konidaris Vissarion 2011030123

Software: Octave Code

---

1. In this exercise we had to implement and use the Principal Component Analysis method (PCA) for dimensionality reduction on two given datasets. The steps of this method are the following. First we standardize the data by making each feature have zero mean and one standard deviation. Next we create the covariance matrix of the given dataset and find its eigenvalues and eigenvectors. We sort the eigenvectors according to their corresponding eigenvalues in descending order for the purpose of preserving only the k first ones. After that, we project each data point from $\mathbb{R}^n$ to $\mathbb{R}^k$ simply by taking its inner product with the first k eigenvectors. Finally, we project the data back to $\mathbb{R}^n$. When we project our data down to a k dimensional space, we preserve only the -hopefully- best trends of our data. It is natural then to have some loss of information when we try to map our data again back on the n dimensional space. This becomes more evident on the second dataset, a set of images depicting faces, where we observe that the faces on the images are severely blurred after the recovery of the data.

2. Suppose we have two equiprobable classes ω_1 and ω_2 following a Gaussian distribution. The calculated from the data mean vectors and covariance matrices can be seen below.

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

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

Applying Linear Discriminant Analysis (LDA) on the dataset we can find a vector w, such that when we project our data on to it, the means of the two classes will be as far as possible from each other, and the within class variance will be as small as possible.

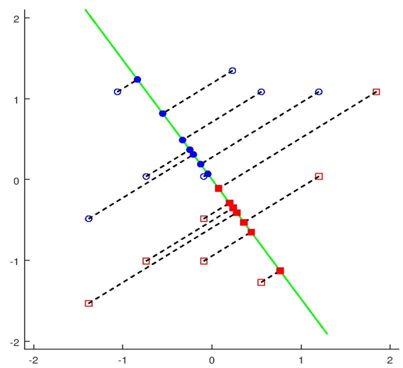
$$w = \Sigma_w^{-1}(\mu_1 - \mu_2)$$

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

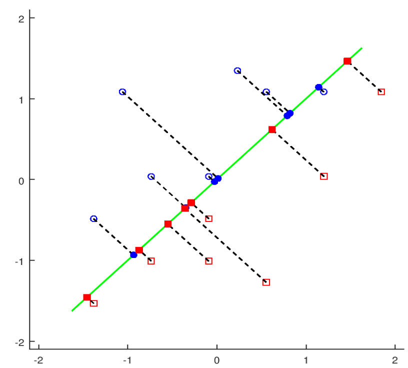
$$\Sigma_w^{-1} = \frac{1}{(\frac{13}{2})^2 - (\frac{9}{2})^2}\begin{bmatrix} \frac{13}{2} & -\frac{9}{2} \\ -\frac{9}{2} & \frac{13}{2} \end{bmatrix} = \begin{bmatrix} \frac{13}{44} & -\frac{9}{44} \\ -\frac{9}{44} & \frac{13}{44} \end{bmatrix}$$

$$w = \begin{bmatrix} \frac{13}{44} & -\frac{9}{44} \\ -\frac{9}{44} & \frac{13}{44} \end{bmatrix}\begin{bmatrix} -5-10 \\ 5-15 \end{bmatrix} = \begin{bmatrix} \frac{13}{44} & -\frac{9}{44} \\ -\frac{9}{44} & \frac{13}{44} \end{bmatrix}\begin{bmatrix} -15 \\ -10 \end{bmatrix} = \begin{bmatrix} -\frac{105}{44} \\ \frac{5}{44} \end{bmatrix}$$

3. In this exercise we had to implement LDA on Matlab/Octave and compare it with the PCA algorithm after running both of them on an example dataset. The results of the two algorithms can be seen on Figure 1. PCA is an unsupervised method for dimensionality reduction, as it doesn't take into consideration the true labels of the data. It tries to find the a set of axis in the direction of greatest variability of the whole dataset. On the other hand, LDA uses the information given by the labels and finds the most discriminant projection by maximizing between-class distance and minimizing within-class distance. On this particular dataset LDA seems to perform better than the PCA as it manages to successfully separate the two classes, while PCA has mapped the dataset on a dimension where the two classes overlap.



(a) LDA method.



(b) PCA method.

Figure 1: Comparison of LDA and PCA methods on the same dataset.

4. In a classification problem we have to classes ω_1 and ω_2 with a-priori probabilities $p(\omega_1)$,$p(\omega_2)$. The likelihood of a data point x with respect to its class is given by the following Gaussian distributions:

$$p(x|\omega_1) = \aleph(\mu_1, \Sigma_1), p(x|\omega_2) = \aleph(\mu_2, \Sigma_2)$$

where

$$\mu_1 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} 6 \\ 6 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{pmatrix}$$

The a-posteriori class probabilities are given by are the following.

$$p(\omega_1|x) = \frac{p(x, \omega_1)}{p(x)} = \frac{p(x|\omega_1)p(\omega_1)}{p(x)}$$

$$p(\omega_2|x) = \frac{p(x, \omega_2)}{p(x)} = \frac{p(x|\omega_2)p(\omega_2)}{p(x)}$$

Given x we classify it by the following rule.

$$If\ p(\omega_1|x) > p(\omega_2|x)\ then\ x \rightarrow \omega_1$$
$$If\ p(\omega_1|x) < p(\omega_2|x)\ then\ x \rightarrow \omega_2$$

The decision region can be found by the equality $p(\omega_1|x) = p(\omega_2|x)$. We know that all class densities $p(x|\omega_i)$ follow the Gaussian distribution

$$p(x|\omega_i) = \frac{1}{(2\pi)^{\frac{l}{2}}|\Sigma|^{\frac{1}{2}}} exp(-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i))$$

$$|\Sigma_1| = 1.2^2 - (-0.4)^2 = 1.2^2 - 0.4^2 = |\Sigma_2| \approx 1.28$$

$$|\Sigma_1|^{\frac{1}{2}} = |\Sigma_2|^{\frac{1}{2}} \approx 1.1314$$

$$\Sigma_1^{-1} = \frac{1}{1.28}\begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{bmatrix} = \begin{bmatrix} 0.9375 & 0.3125 \\ 0.3125 & 0.9375 \end{bmatrix}$$

$$\Sigma_2^{-1} = \frac{1}{1.28}\begin{bmatrix} 1.2 & -0.4 \\ -0.4 & 1.2 \end{bmatrix} = \begin{bmatrix} 0.9375 & -0.3125 \\ -0.3125 & 0.9375 \end{bmatrix}$$

So we have that

$$p(\omega_1|x) = p(\omega_2|x) \Leftrightarrow$$

$$p(\omega_1)\frac{1}{2\pi \cdot 1.1314} exp(-\frac{1}{2}\begin{bmatrix} x_1 - 3 & x_2 - 3 \end{bmatrix}\begin{bmatrix} 0.9375 & 0.3125 \\ 0.3125 & 0.9375 \end{bmatrix}\begin{bmatrix} x_1 - 3 \\ x_2 - 3 \end{bmatrix}) =$$

$$p(\omega_2)\frac{1}{2\pi \cdot 1.1314} exp(-\frac{1}{2}\begin{bmatrix} x_1 - 6 & x_2 - 6 \end{bmatrix}\begin{bmatrix} 0.9375 & -0.3125 \\ -0.3125 & 0.9375 \end{bmatrix}\begin{bmatrix} x_1 - 6 \\ x_2 - 6 \end{bmatrix})$$

and decision region is given by the following quadratic expression:

$$ln(p\omega_1) - \frac{1}{2}\begin{bmatrix} x_1 - 3 & x_2 - 3 \end{bmatrix}\begin{bmatrix} 0.9375 & 0.3125 \\ 0.3125 & 0.9375 \end{bmatrix}\begin{bmatrix} x_1 - 3 \\ x_2 - 3 \end{bmatrix} =$$

$$ln(p\omega_2) - \frac{1}{2}\begin{bmatrix} x_1 - 6 & x_2 - 6 \end{bmatrix}\begin{bmatrix} 0.9375 & -0.3125 \\ -0.3125 & 0.9375 \end{bmatrix}\begin{bmatrix} x_1 - 6 \\ x_2 - 6 \end{bmatrix}$$

On Figure 2 we can see the level sets of the two distributions $p(x|\omega_i) \in \mathbb{R}^2$ along with the decision region for various a-priory values of the two classes. The subplots corresponding to these distributions are the (a), (b), (c), (d) and (e).

If the covariance matrices Σ_1 and Σ_2 are equal, then the decision region will be a hyperplane. Assuming that

$$\Sigma = \Sigma_1 = \Sigma_2 = \begin{bmatrix} 1.2 & 0.4 \\ 0.4 & 1.2 \end{bmatrix},$$

then the decision region is given by the linear expression.

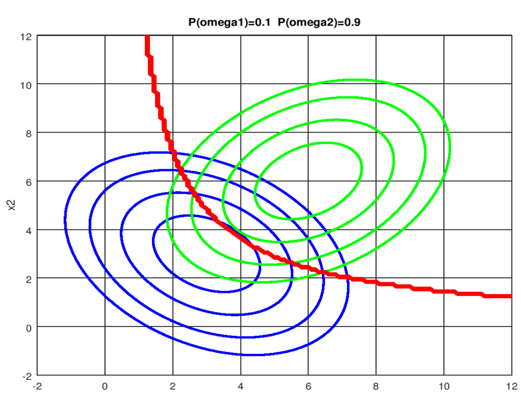
$$(\Sigma^{-1}\mu_1)^T x + ln(p(\omega_1)) - \frac{1}{2}\mu_1^T \Sigma^{-1}\mu_1 = (\Sigma^{-1}\mu_2)^T x + ln(p(\omega_2)) - \frac{1}{2}\mu_2^T \Sigma^{-1}\mu_2$$

$$\Leftrightarrow \begin{bmatrix} 3 & 3 \end{bmatrix} \begin{bmatrix} 0.9375 & -0.3125 \\ -0.3125 & 0.9375 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + ln(p(\omega_1)) - \frac{1}{2} \begin{bmatrix} 3 & 3 \end{bmatrix} \begin{bmatrix} 0.9375 & -0.3125 \\ -0.3125 & 0.9375 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

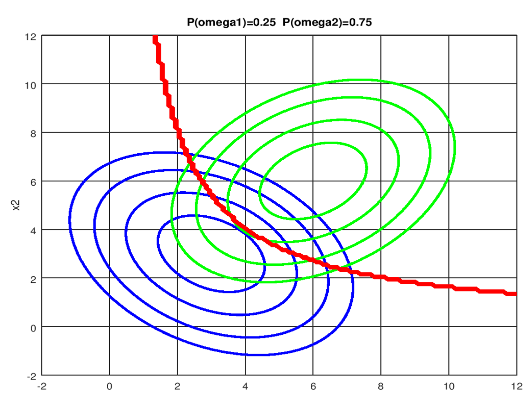
$$= \begin{bmatrix} 6 & 6 \end{bmatrix} \begin{bmatrix} 0.9375 & -0.3125 \\ -0.3125 & 0.9375 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + ln(p(\omega_2)) - \frac{1}{2} \begin{bmatrix} 6 & 6 \end{bmatrix} \begin{bmatrix} 0.9375 & -0.3125 \\ -0.3125 & 0.9375 \end{bmatrix} \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

The subplots corresponding to the level sets and decision boundaries of those classes on Figure 2 are the (f), (g), (h), (i) and (j).

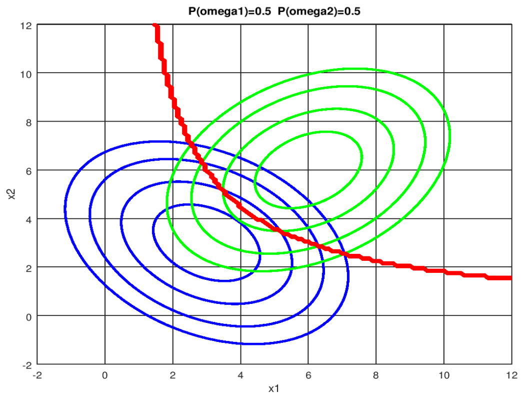
It is evident from the plots that the decision boundary is changing along with the a-priori probabilities of the two classes. As $p(\omega_1)$ increases, the decision boundary moves closer to the mean of the $\omega 2$, making the decision region of ω_1 bigger.



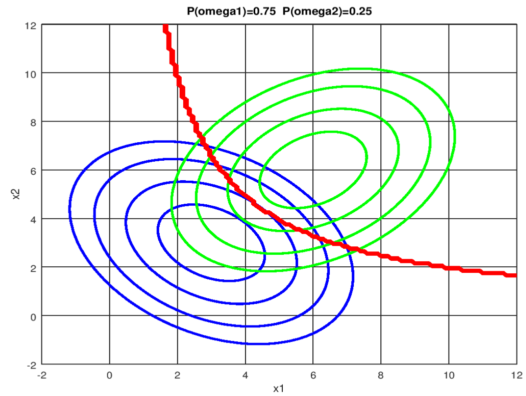
(a) Different covariance matrices, $p(\omega_1) = 0.1$.



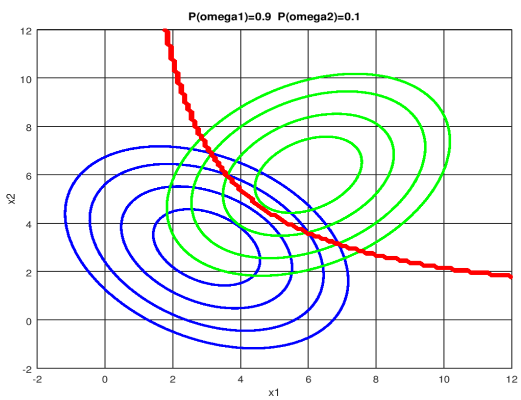
(b) Different covariance matrices, $p(\omega_1) = 0.25$.



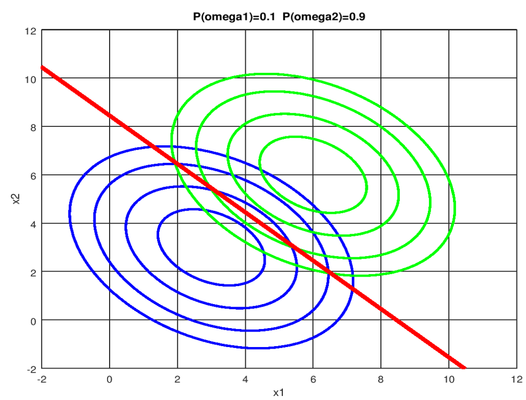
(c) Different covariance matrices, $p(\omega_1) = 0.5$.



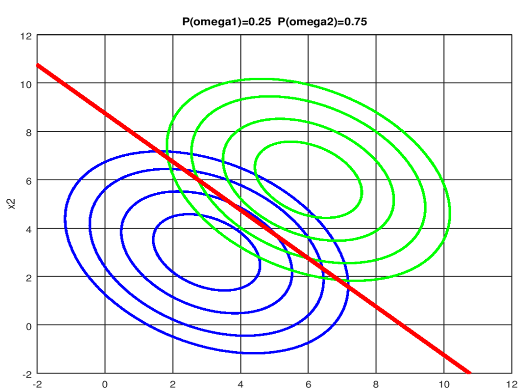
(d) Different covariance matrices, $p(\omega_1) = 0.75$.



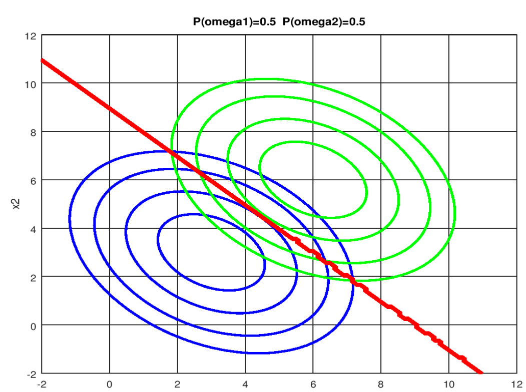
(e) Different covariance matrices, $p(\omega_1) = 0.9$.



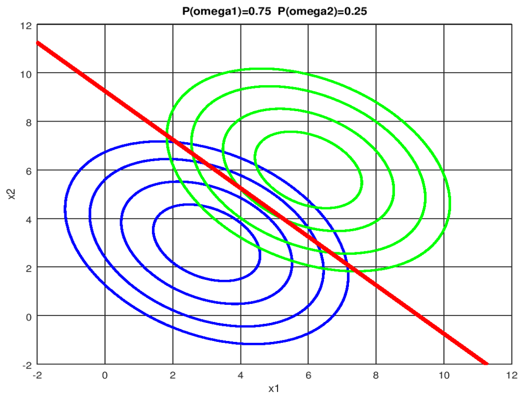
(f) Same covariance matrices, $p(\omega_1) = 0.1$.



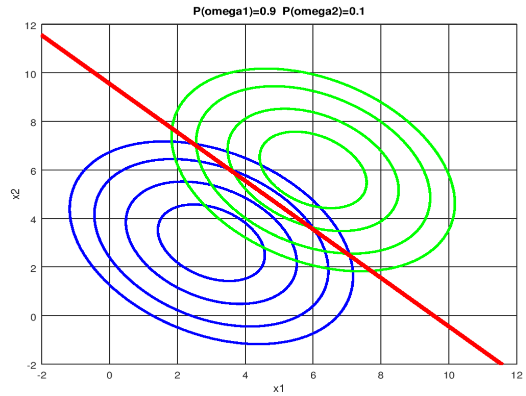
(g) Same covariance matrices, $p(\omega_1) = 0.25$.



(h) Same covariance matrices, $p(\omega_1) = 0.5$.



(i) Same covariance matrices, $p(\omega_1) = 0.75$.



(j) Same covariance matrices, $p(\omega_1) = 0.9$.

Figure 2: Level sets and decision boundaries of the two classes.

5. For this exercise we have a classification problem with two classes ω_1 and ω_2 with equal a-priori probabilities. The data points x are one dimensional and their likelihood w.r.t ω_i follow a Rayleigh distribution.

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

where $\sigma_1 = 1$ and $\sigma_2 = 2$. The risk matrix is

$$L = \begin{pmatrix} 0 & 0.5 \\ 1.0 & 0 \end{pmatrix}.$$

The losses of miss classification are:

$$l1 = l_{11} p(x|\omega_1) p(\omega_1) + l_{21} p(x|\omega_2) p(\omega_2)$$

$$l2 = l_{12} p(x|\omega_1) p(\omega_1) + l_{22} p(x|\omega_2) p(\omega_2)$$

To find the x_0 that minimizes the risk we must solve for $l1 = l2$.

$$l_{21} p(x|\omega_2) p(\omega_2) = l_{12} p(x|\omega_1) p(\omega_1)$$

$$\Leftrightarrow l_{21} p(x|\omega_2) = l_{12} p(x|\omega_1)$$

$$\Leftrightarrow \frac{p(x|\omega_1)}{p(x|\omega_2)} = \frac{l_{21}}{l_{12}} \Leftrightarrow \frac{xe^{-\frac{x_0^2}{2}}}{\frac{x_0}{4}e^{-\frac{x_0^2}{8}}} = 2$$

$$ln(4e^{-\frac{3x_0^2}{8}}) = ln2 \Leftrightarrow ln4 - \frac{x_0^2}{8} = ln2 \Leftrightarrow x_0^2 = \frac{8ln2}{3} \approx 1.8484$$

$$x_0 \approx 1.3596$$