ECE471/571 Homework #2 Solution

Note: UG: 100+10, G:100

Data set used in this homework:

X	Y	Class
0.8	1.2	1
0.9	1.4	1
1.2	1.4	1
1.1	1.5	1
0.8	1.1	2
0.6	1	2
0.65	1.1	2
0.75	0.9	2

1) (30/20) Mahalanobis distance vs. Euclidean distance.

   a. (5/5) Show the equations to calculate these two distances.

$$D_{Euclidean} = \sqrt{(x-m)^T (x-m)}$$

$$D_{Mahalanobis} = \sqrt{(x-m)^T \Sigma^{-1} (x-m)}$$

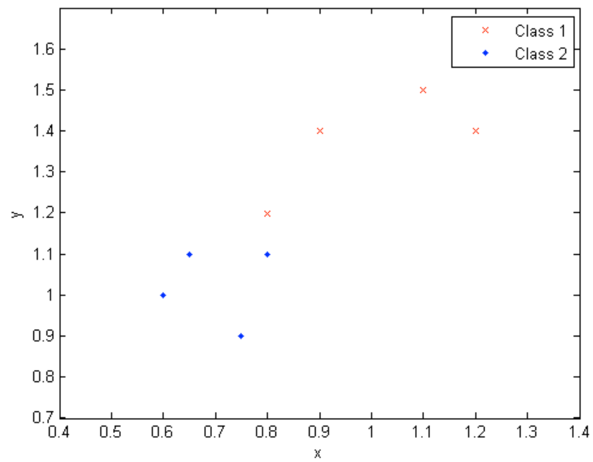
where, $x \in R^n$ is a data to be given for distance measurement. $m \in R^n$ denotes mean value for a group of values. Σ is covariance matrix of the group.

   b. (5/5) Explain intuitively (in no more than three sentences) the differences between the two distances.

Mahalanobis distance considers data distribution by covariance so that it is independent on the scale (**scale-invariant**) of measurement. Euclidean distance provides a kind of point estimation for a group of data by using a mean value as a representative value for the group, so that the distance is simply calculated by finding a shortest distance between two points in Cartesian coordinate in **scale-variant** manner.

c. (20/10) Use the following example to understand the differences these two distances made in classification. Here, the minimum distance classifier is used.

i. Plot the above data set on the same figure?



ii. Given a test sample $x = [\ 0.85 \quad 1.15\ ]^T$, calculate the Euclidean distance to the two classes. Based on the distances, which class should x belong to?

	Class 1	Class 2
Mean	$m_1 = \begin{bmatrix}1 & 1.375\end{bmatrix}^T$	$m_2 = \begin{bmatrix}0.7 & 1.025\end{bmatrix}^T$
Euclidean distance	$D_{Euclidean,1} = 0.2704$	$D_{Euclidean,2} = 0.1953$

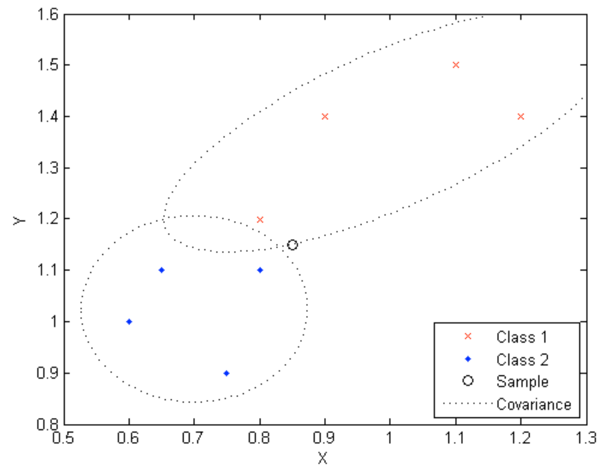
Since $D_{Euclidean,1} > D_{Euclidean,2}$, x belongs to **class 2** based on minimum distance classifier with the criterion, Euclidean distance.

iii. Use the same test sample, calculate the Mahalanobis distance to the two classes. Based on this pair of distances, which class should x belong to?

	Class 1	Class 2
Covariance matrix	$\Sigma_1 = \begin{bmatrix}0.0333 & 0.0167 \\ 0.0167 & 0.0158\end{bmatrix}^T$	$\Sigma_2 = \begin{bmatrix}0.0083 & 0 \\ 0 & 0.0092\end{bmatrix}^T$
Mahalanobis distance	$D_{Mahalanobis,1} = 1.9170$	$D_{Mahalanobis,2} = 2.0987$

Since $D_{Mahalanobis,1} < D_{Mahalanobis,2}$, x also belongs to **class 1** based on minimum distance classifier with the criteria, Mahalanobis distance.

iv. Plot the test sample x on the same figure as the data set. Just by observing the plot, which decision do you think makes more sense?



Class 1 (Look at the figure above intuitively. The dotted circles represent estimated data distribution for class 1 and class 2 data sets)

2) (70/70) Dimensionality reduction with a two-feature two-class data set.

a. (10/10) Preprocessing steps (Need to show step-by-step details):

i. Calculate the mean of each class (m_1, m_2) manually.

$$m_1 = \frac{1}{4}\left(\begin{bmatrix}0.8\\1.2\end{bmatrix}+\begin{bmatrix}0.9\\1.4\end{bmatrix}+\begin{bmatrix}1.2\\1.4\end{bmatrix}+\begin{bmatrix}1.1\\1.5\end{bmatrix}\right)=\begin{bmatrix}1\\1.375\end{bmatrix}$$

$$m_2 = \frac{1}{4}\left(\begin{bmatrix}0.8\\1.1\end{bmatrix}+\begin{bmatrix}0.6\\1\end{bmatrix}+\begin{bmatrix}0.65\\1.1\end{bmatrix}+\begin{bmatrix}0.75\\0.9\end{bmatrix}\right)=\begin{bmatrix}0.7\\1.025\end{bmatrix}$$

ii. Calculate the covariance matrix of each class (Σ_1, Σ_2) manually

Covariance matrix is calculated by,

$$\Sigma_i = \frac{1}{n-1}\sum_{[x\ y]^T \in D_i}\left(\begin{bmatrix}x\\y\end{bmatrix}-m_i\right)\left(\begin{bmatrix}x\\y\end{bmatrix}-m_i\right)^T$$

where n denotes the number of data in i -th class, D_i .

$$
\Sigma_1 = \frac{1}{3}\left\{ \begin{aligned} &\left(\begin{bmatrix}0.8\\1.2\end{bmatrix}-m_1\right)\left(\begin{bmatrix}0.8\\1.2\end{bmatrix}-m_1\right)^T+\left(\begin{bmatrix}0.9\\1.4\end{bmatrix}-m_1\right)\left(\begin{bmatrix}0.9\\1.4\end{bmatrix}-m_1\right)^T \\ &+\left(\begin{bmatrix}1.2\\1.4\end{bmatrix}-m_1\right)\left(\begin{bmatrix}1.2\\1.4\end{bmatrix}-m_1\right)^T+\left(\begin{bmatrix}1.1\\1.5\end{bmatrix}-m_1\right)\left(\begin{bmatrix}1.1\\1.5\end{bmatrix}-m_1\right)^T \end{aligned} \right\}
$$

$$
= \begin{bmatrix}0.0333 & 0.0167\\ 0.0167 & 0.0158\end{bmatrix}
$$

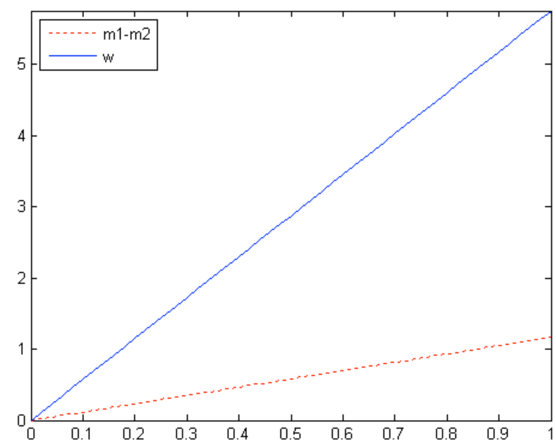
$$
\Sigma_1 = \frac{1}{3}\left\{ \begin{aligned} &\left(\begin{bmatrix}0.8\\1.1\end{bmatrix}-m_2\right)\left(\begin{bmatrix}0.8\\1.1\end{bmatrix}-m_2\right)^T+\left(\begin{bmatrix}0.6\\1\end{bmatrix}-m_2\right)\left(\begin{bmatrix}0.6\\1\end{bmatrix}-m_2\right)^T \\ &+\left(\begin{bmatrix}0.65\\1.1\end{bmatrix}-m_2\right)\left(\begin{bmatrix}0.65\\1.1\end{bmatrix}-m_2\right)^T+\left(\begin{bmatrix}0.75\\0.9\end{bmatrix}-m_2\right)\left(\begin{bmatrix}0.75\\0.9\end{bmatrix}-m_2\right)^T \end{aligned} \right\}
$$

$$
= \begin{bmatrix}0.0083 & 0\\ 0 & 0.0092\end{bmatrix}
$$

b. (20/20) Using Fisher Linear Discriminant (FLD) to find the vector (w) which optimally (in the Fisher sense) separates the projections of these two classes.

$$
\begin{aligned}
w &= S_w^{-1}\left(m_1 - m_2\right) \\
&= (S_1 + S_2)^{-1}(m_1 - m_2) \\
&= \left(3\Sigma_1 + 3\Sigma_2\right)^{-1}(m_1 - m_2) \qquad , \quad \frac{w}{\|w\|} = \begin{bmatrix}0.1713\\0.9852\end{bmatrix} \\
&= \begin{bmatrix}10.9091 & -7.2727\\ -7.2727 & 18.1818\end{bmatrix}\begin{bmatrix}0.3\\0.35\end{bmatrix} \\
&= \begin{bmatrix}0.7273\\4.1818\end{bmatrix}
\end{aligned}
$$

c. (10/10) Is the vector derived from FLD along the same direction as (m1-m2)? Plot m1-m2 and w on the same figure. Also explain it analytically.



Not a same direction since the sum of within scatter matrix is not identity matrix for the given distribution.

d. (20/20) Using principal component analysis (PCA) to reduce the dimension and compare it with the one derived from FLD. (You can use existing functions to find the eigenvectors and eigenvalues.)

For PCA,

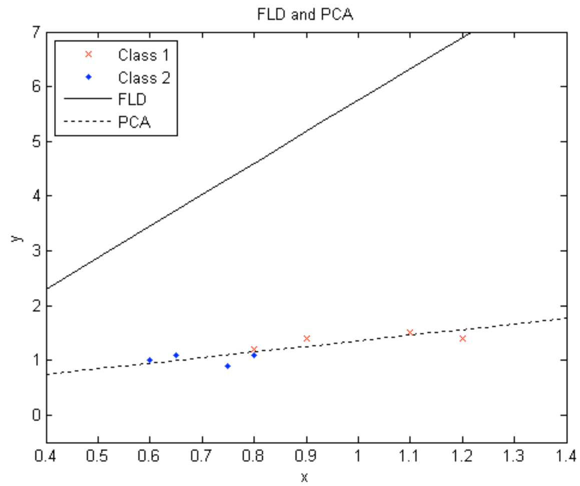
$$S = \sum_{k=1}^{n}(x_k - m)(x_k - m)^T = \begin{bmatrix} 0.305 & 0.26 \\ 0.26 & 0.32 \end{bmatrix} \text{ where } m = \frac{1}{n}\sum_{k=1}^{n} x_k = \begin{bmatrix} 0.85 \\ 1.2 \end{bmatrix}$$

The eigenvector with corresponding largest eigenvalue: $\begin{bmatrix} 0.6968 & 0.7172 \end{bmatrix}^T$

Table: Dimensionality reduction by PCA and FLD

[X	Y]	Class	PCA	FLD
0.8	1.2	1	-0.0348	1.3193
0.9	1.4	1	0.1783	1.5335
1.2	1.4	1	0.3873	1.5849
1.1	1.5	1	0.3894	1.6663
0.8	1.1	2	-0.1066	1.2208
0.6	1	2	-0.3177	1.0880
0.65	1.1	2	-0.2111	1.1951
0.75	0.9	2	-0.2849	1.0152

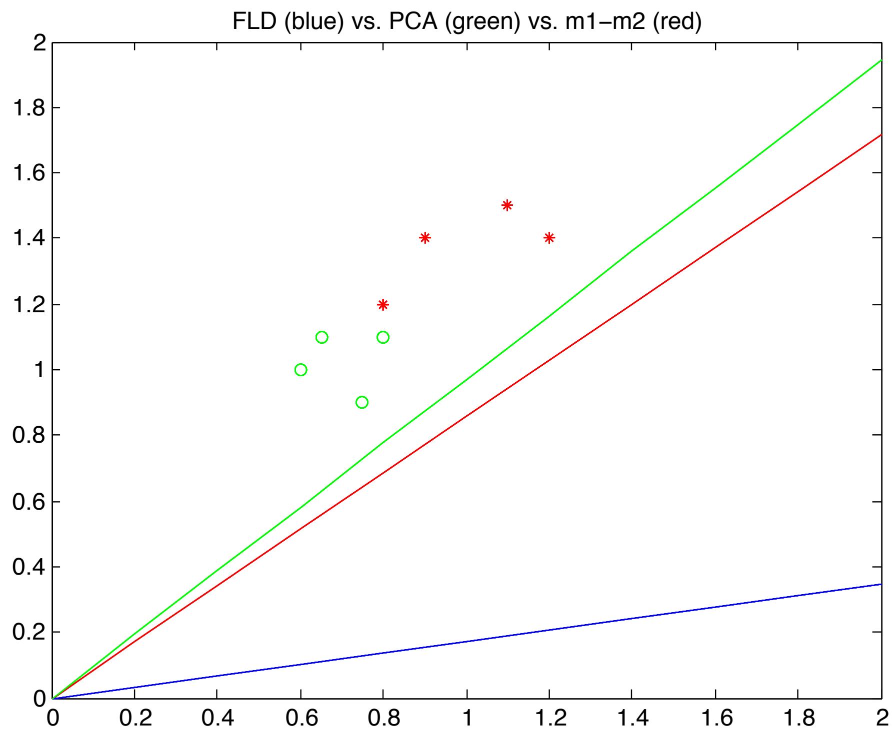
The comparison of PCA and FLD is in the figure below



e. (10/10) Comment on the differences between FLD and PCA.

Both FLD and PCA work well to provide appropriate dimensionality reduction for the given dataset. Because of the difference found in optimality criteria, FLD and PCA show different dimensionality reduction results. In the case of PCA, the 1D vector passes through sample mean compared with FLD crossing the origin. It is not possible to claim either one of these is better from the figure in Q2.d

	PCA	FLD
Utilization of class information	No	Yes
Fundamental approach	Seeks a projection that best represents the data in a least-square sense	Seeks a projection that best separates the data in a least-squares sense



3) (+10/10) Using maximum likelihood method to derive the equation for variance or covariance.

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \mu \\ \Sigma \end{bmatrix}$$

$$p(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right]$$

$$p(x_k \mid \theta) = \frac{1}{(2\pi)^{d/2} |\theta_2|^{1/2}} \exp\left[-\frac{1}{2}(x_k-\theta_1)^T \frac{1}{\theta_2} (x_k-\theta_1) \right]$$

$$l(\theta) = \sum_{k=1}^{n} \left(-\frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln|\theta_2| - \frac{1}{2}(x_k-\theta_1)^T \frac{1}{\theta_2}(x_k-\theta_1) \right)$$

$$\begin{aligned} \frac{\partial l}{\partial \theta_2} &= \sum_{k=1}^{n} \left(-\frac{1}{2}\frac{1}{\theta_2} + \frac{1}{2}\frac{(x_k-\theta_1)^T (x_k-\theta_1)}{\theta_2^2} \right) \\ &= -\frac{n}{\theta_2} + \sum_{k=1}^{n} \frac{(x_k-\theta_1)^2}{\theta_2^2} \\ &= 0 \end{aligned}$$

$$\frac{n}{\theta_2} = \sum_{k=1}^{n} \frac{(x_k - \theta_1)^2}{\theta_2^2},$$

$$\begin{aligned} \theta_2 &= \frac{1}{n}\sum_{k=1}^{n}(x_k - \theta_1)^2 \\ &= \sigma^2 \\ &= \frac{1}{n}\sum_{k=1}^{n}(x_k - \mu)^2 \end{aligned}$$