Office of Sustainability
Amirkabir University of Technology
(Tehran Polytechnic)

COMPUTER ENGINEERING && IT DEPARTMENT

AMIRKABIR UNIVERSITY OF TECHNOLOGY

Statistical Pattern Recognition

Submitted To:
Mohammad Rahmati
Asst. Professor
Computer Engineering
Department

Submitted By :
Ahmad Asadi
94131091
Group-G1
Fall-95

Contents

1   Problem 1

Two normal distribution are characterized by:

$$P_1 = P_2 = 0.5$$

$$M_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$
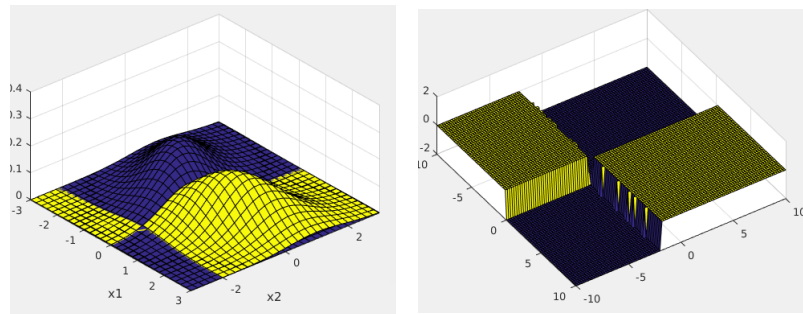
$$M_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$

1. Draw the Bayes decision boundary to minimize the probability of error.

   The bayesian decision function is:

$$P(X|w_1) \cdot P(w_1) \underset{<w_2}{\overset{>w_1}{\gtrless}} P(X|w_2) \cdot P(w_2)$$

$$\frac{P(X|w_1) \cdot P(w_1)}{P(X|w_2) \cdot P(w_2)} \underset{<w_2}{\overset{>w_1}{\cdot\gtrless}} 1 \rightarrow \frac{P(X|w_1)}{P(X|w_2)} \underset{<w_2}{\overset{>w_1}{\cdot\gtrless}} 1$$

   The figure 1 displays the data distributions and decision boundaries of bayesian decision function without considering costs.

(a) Distribution of data. The distribution of class 1 is displayed in blue color and the distribution of class 2 is displayed in yellow.

(b) Decision Boundary. The decision boundary of class 1 is displayed in blue color and that of class 2 is displayed in yellow.
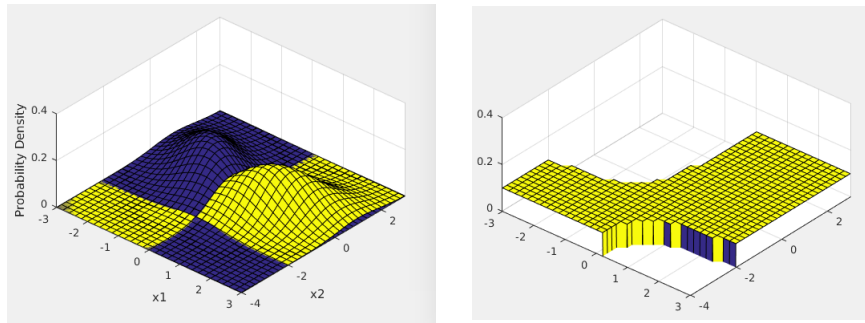
Figure 1: Bayesian Decision Boundary without considering decision costs

2. Draw the Bayes decision boundary to minimize the cost with $c_{11} = c_{22} = 0$ and $c_{12} = 2c_{21}$

The Bayesian decision boundary when considering costs of decisions turns to the form bellow:

$$\frac{p_1(X)}{p_2(X)} \overset{>^{w_1}}{\underset{<_{w_2}}{.}} \frac{(c_{12} - c_{22})P_2}{(c_{21} - c_{11})P_1} \tag{1}$$

Figure **??** displays the decision boundary in this case.

(a) Distribution of data. The distribution of class 1 is displayed in yellow color and the distribution of class 2 is displayed in blue.

(b) Decision Boundary. The decision boundary of class 1 is displayed in yellow color and the that of class 2 is displayed in blue.

Figure 2: Bayesian Decision Boundary considering decision costs $c_{11} = c_{22} = 0$ and $c_12 = 2c_{21}$

---

3. Assume that $c_{11} = c_{22} = 0$ and $c_{12} = c_{21}$:

   (a) Plot the operating characteristics.

   As equation (1) yields, with $c_{11} = c_{22} = 0$ and $c_{12} = c_{21}$ the decision boundary is not different from that of section 1. So the figure 1 is displaying the required decision boundary

   ---

   (b) Find the total error when Neyman-Pearson test is performed with $\epsilon_1 = 0.05$

   We should estimate μ such that the following function r be minimized.

   $$r = \mu(\epsilon_1 - 0.05) + \epsilon_2 \tag{2}$$

3

After finding appropriate μ, the decision boundary is derivable from:

$$\frac{p_1(X)}{p_2(X)} \cdot {\genfrac{}{}{0pt}{}{>^{w_1}}{<_{w_2}}} \mu \tag{3}$$

To find such μ an script will iterate on following equation to reach the best fitting value.

$$\epsilon_1 = \int_{\mu}^{\infty} p_h(h|W-1)dh = 0.05 \tag{4}$$

The $\mu = 1.0795e - 78$ is the best value estimated with $\epsilon_1 = 0.051$ error rate. Figure 3 illustrates the resulted decision boundary.
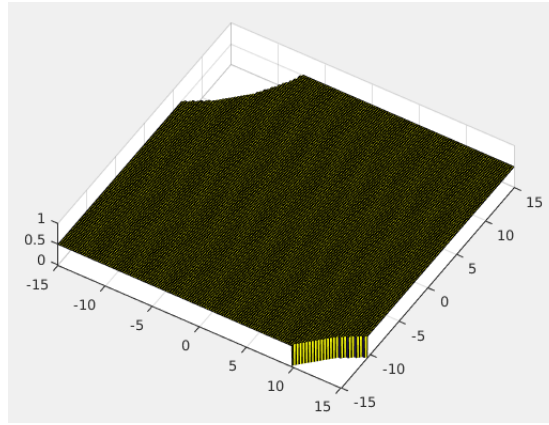

Figure 3: Decision boundary resulted from Neyman-Pearson test with $\epsilon_1 = 0.05$

(c) Find the threshold value and total error for the minimax test.

(d) Plot the error-reject curve.

2   Problem 2

Two normal distribution are characterized by:

$$P_1 = P_2 = 0.5$$

$$M_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}$$

1. Find the linear discriminant function which maximize the Fisher criterion and minimize the error by adjusting the threshold.

$$f = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \rightarrow \frac{\partial f}{\partial \sigma_1^2} = \frac{\partial f}{\partial \sigma_2^2} = \frac{-(\mu_1 - \mu_2)^2}{(\sigma_1^2 + \sigma_2^2)^2}$$

$$\rightarrow s = \frac{\dfrac{\partial f}{\partial \sigma_1^2}}{\dfrac{\partial f}{\partial \sigma_1^2} + \dfrac{\partial f}{\partial \sigma_2^2}} = 0.5$$

$$\rightarrow V = [0.5\Sigma_1 + 0.5\Sigma_2]^- 1(\mu_1 - \mu_2) = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}^- 1 \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$\rightarrow V = \begin{bmatrix} -0.5 \\ 0 \end{bmatrix} \tag{5}$$

Equation (5) illustrates the best mapping vector V to project distribution over, therefore the LDA will be in the form of $V^t x \geq \alpha$, in which α should be find from equation (6).

$$\frac{\partial f}{\partial \mu_1} + \frac{\partial f}{\partial \mu_2} = 0 \tag{6}$$

which results:

$$\frac{2\mu_1 - 2\mu_2}{\sigma_1^2 + \sigma_2^2} + \frac{-2\mu_1 + 2\mu_2}{\sigma_1^2 + \sigma_2^2} = 0 \rightarrow 0 = 0$$

Therefore, the best α is not derivable analytical. In order to estimate the best α a MATLAB script using bootstrapping has been written by me, which is appended to the homework submission zip file. The bootstrapping process, iterates 1000 number of times in each iteration generating 1000 multivariate random variables regarding given means and variances for each class. We

attempt to estimate best α in each iteration using grid search in the range of $[-4, 4]$ with step length 0.01. To measure the accuracy of chosen α we use the ratio $\frac{tp+tn-fp-fn}{N}$ as the measurement function. After all iterations, the mean of all best estimated αs is reported as the best α. Experiments resulted the $\alpha = 0.23$ as the best value separating two classes from each other. So the LDA function is being reformed as bellow:

$$\begin{bmatrix} -0.5 \\ 0 \end{bmatrix}^T X \geq^{w_1} 0.23 \tag{7}$$

2. Find the optimum linear discriminant function which minimize the probability of error.

The iterative process has been coded in MATLAB and its code is attached to the submitted homework zip file.

The parameter s has been estimated using a grid search in rage $[0, 1]$ with step length 0.01. the error probability *epsilon* in each iteration with respect to selected s has been calculated. The figure 4 illustrates the behaviour of error probability w.r.t. selected s. The best value of s is in 0 or 1.
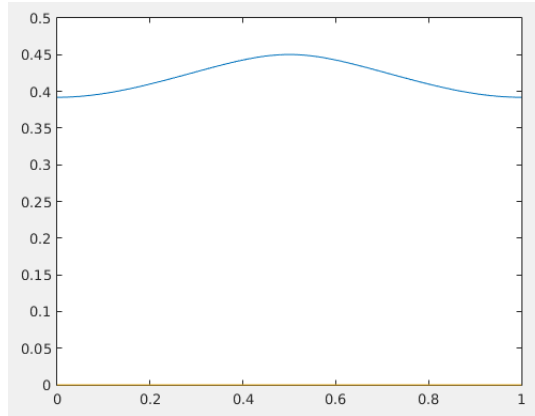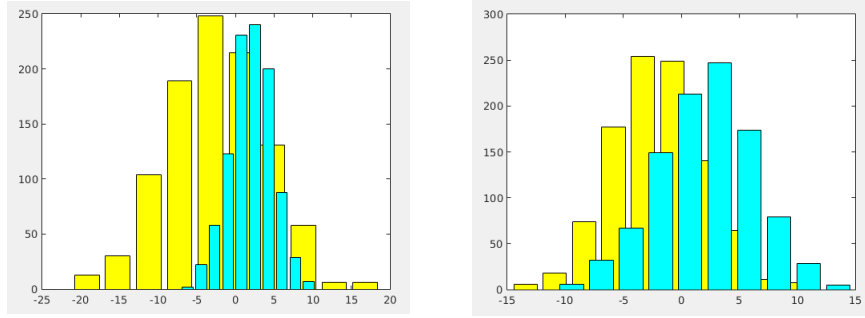

Figure 4: The behaviour of error probability ϵ w.r.t. chosen s

To through a light over reality, the projection of data in two cases of $s = 0.5$ and $s = 1$ has been displayed in the figure 7, respectively the worst- and the best-case. As its obviously clear, the separation of classes is more suitable in case $s = 1$ rather than $s = 0.5$, hence the lower error probability will be gained when $s = 1$.

(a) Projected distribution of data. The projected distribution of class 1 w.r.t. mapping vector V is displayed in blue color and the that of class 2 is displayed in yellow when $s = 1$ (BEST CASE).

(b) Projected distribution of data. The projected distribution of class 1 w.r.t. mapping vector V is displayed in blue color and the that of class 2 is displayed in yellow when $s = 0.5$ (WORST CASE).

Figure 5: A comparison between projected distributions over matrices V w.r.t. selected s

3   Problem 3

We consider a classification problem in dimension $d = 2$, with $k = 3$ classes where: $p(x|w_i) \propto N(\mu_i, \Sigma_i), i = 1, 2, 3$, with:

$$\mu_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \mu_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \mu_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \Sigma_i = \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$$

1. Calculate the discriminant function $g_i(x)$ for each class.
   Generating a discriminant function between all pairs of classes can be a good solution. Assuming $P(W_1) = P(W_2) = P(W_3)$, the Bayes classifier is a good choice:

$$d_{ij} : \frac{P(X|W_i)}{P(X|W_j)} \cdot \overset{>^{W_i}}{\underset{<_{W_j}}{}} 1$$

Defining such classifiers, the feature function $g_i(X)$ can be defined as :

$$g_1(X) = \Pi_{j=2,3} sgn(\frac{P(X|W_1)}{P(X|W_j)} - 1)$$

$$g_2(X) = \Pi_{j=2,3} sgn(\frac{P(X|W_2)}{P(X|W_j)} - 1)$$

$$g_3(X) = \Pi_{j=2,3} sgn(\frac{P(X|W_3)}{P(X|W_j)} - 1)$$

In which the positive value of each function $g_i(X)$ indicates that X belongs to W_i and vice versa.

_____

2. Express your discriminant functions in the form of linear discriminant functions.
   The d_{ij}s could be expressed as linear functions as perpendicular bisectors of linking line between the center points of each class:

$$d_{12} = X_2 + 3X_1 + 3$$

$$d_{13} = X_2 + \frac{1}{2}X_1 - \frac{3}{4}$$

$$d_{23} = X_2 + 2X_1 - \frac{9}{2}$$

Defining $d_{ij}(X) = -d_{ji}(X)$, the vector X belongs to class W_i iff $\forall j \neq i, d_{ij}(X) > 0$

_____

3. Determine and plot the decision boundaries. Figure 6 displays the distribution of data given mean vectors and covariance matrices.

   Figure ?? displays the data distribution boundary and deriven boundaries using two-class discriminant functions defined in section 2 of this problem.
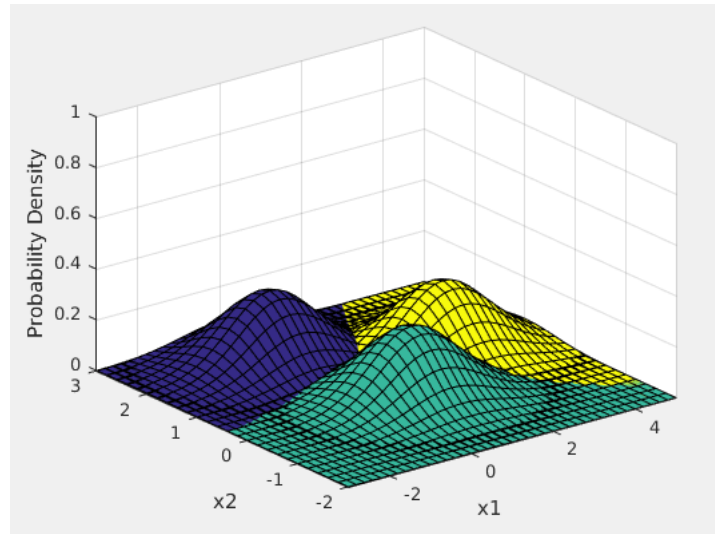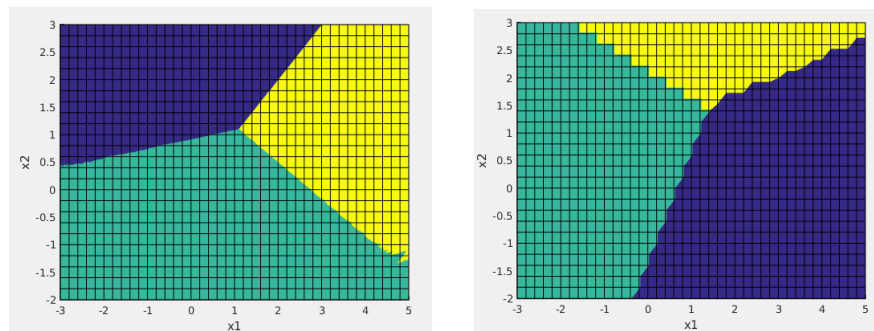
_____

8

Figure 6: Data distribution w.r.t. means and covariance matrices

(a) Projected distribution of data on $X1 - X2$ coordinates

(b) Boundaries generated by discriminant functions

Figure 7: A comparison between projected distributions over matrices V w.r.t. selected s

9