Statistical Pattern Recognition
Assignment 2

Ali Gholami

Department of Computer Engineering & Information Technology
Amirkabir University of Technology

http:// ceit.aut.ac.ir/~aligholamee
aligholamee@aut.ac.ir

Abstract

In this assignment, we'll be focusing on the *Bayes Classifier*. We'll work with *Bayesian Discriminators* and *Bayes Error*. The *Bhattacharyya* error bound is also analyzed as an upper bound for the *Bayes Classifier* error. The detailed computations of *Bayesian Discriminators* are also given in an exact definition. Finally, we'll be going through a more practical example of a linear discriminator by classifying the flowers in the *Iris* dataset.

Keywords. *Linear Discriminator, Quadratic Discriminator, Bayes Classification, Bayes Error, Optimal Classification, Bhattacharyya Distance, Bhattacharyya Upper Bound, Iris Dataset, Iris Classification.*

1 Quadratic & Linear Discriminant Analysis

We consider a classification problem in dimension $d = 2$, with $k = 3$ classes where:

$$p(x \mid w_i) \sim N(\mu_i, \Sigma_i), \quad i = 1, 2, 3$$

and

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

(a) Calculate the discriminant function $g_i(x)$ for each class.

(b) Express your discriminant functions in the form of linear discriminant functions.

(c) Determine and plot the decision boundaries.

Solution

(a) The general form of a Bayesian discriminator is given below.

$$g_i(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_i)^T \Sigma_i^{-1} (\underline{x} - \underline{\mu}_i) - \frac{1}{2} \log |\Sigma_i| + \log P(\omega_i) \tag{1.1}$$

In the problem case, the classes have the same covariance matrix, but the features have different variances. Since the Σ_i is diagonal, we'll have

$$g_i(\underline{x}) = -\frac{1}{2}(\underline{x}-\underline{\mu}_i)^T \begin{bmatrix} \sigma_1^{-2} & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^{-2} & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^{-2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma_N^{-2} \end{bmatrix} (\underline{x}-\underline{\mu}_i) - \frac{1}{2}\log \begin{vmatrix} \sigma_1^{-2} & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^{-2} & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^{-2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma_N^{-2} \end{vmatrix} + \log(P(\omega_i))$$

Since we have the following criteria:

$$(\underline{x} - \underline{\mu}_i)^T = \begin{bmatrix} x[1] - \mu_i[1] \\ x[2] - \mu_i[2] \\ x[3] - \mu_i[3] \\ x[4] - \mu_i[4] \\ \vdots \\ x[N] - \mu_i[N] \end{bmatrix}$$

where μ_{iN} denotes the N'th feature of class i. Removing the constant term for different classes, which is $x[k]^2$, we'll have the following results after the matrix multiplication and determinant computation:

$$g_i(\underline{x}) = -\frac{1}{2}\sum_{k=1}^{N} \frac{2x[k]\mu_i[k] + \mu_i[k]^2}{\sigma_k^2} - \frac{1}{2}\log\prod_{k=1}^{N}\sigma_k^2 + \log(P(\omega_i)) \tag{1.2}$$

One can simply find each discriminator, $g_i(\underline{x})$, by replacing the given information in the problem description in the formula given above. Thus we'll have the following results for the section (a).

$$g_1(\underline{x}) = -\frac{1}{2}(\frac{2x[1]*0+2}{1} + \frac{2x[2]*2+4}{\frac{1}{9}}) - \frac{1}{2}\log(1*\frac{1}{9}) + ?$$

$$g_2(\underline{x}) = -\frac{1}{2}(\frac{2x[1]*3+3}{1} + \frac{2x[2]*1+1}{\frac{1}{9}}) - \frac{1}{2}\log(1*\frac{1}{9}) + ?$$

$$g_3(\underline{x}) = -\frac{1}{2}(\frac{2x[1]*1+1}{1} + \frac{2x[2]*0+0}{\frac{1}{9}}) - \frac{1}{2}\log(1*\frac{1}{9}) + ?$$

The simplified results are

$$g_1(\underline{x}) = -18x[2] - \frac{1}{2}\log\frac{1}{9} - 19$$

$$g_2(\underline{x}) = -3x[1] + 9x[2] - \frac{1}{2}\log\frac{1}{9} - 6$$

$$g_3(\underline{x}) = -x[1] - \frac{1}{2}\log\frac{1}{9} - \frac{1}{2}$$

(b) The final results given above where in the format of a linear discriminant already. In order to lighten everything up, just assume the linear discriminant function as:

$$g_i(\underline{x}) = W_2 x[2] + W_1 x[1] + W_0$$

where the value of W_i is different for each of the discriminators.

$$g_1(\underline{x}) \quad W_2 = -18 \;\; W_1 = 0 \;\; W_0 = -\frac{1}{2}\log\frac{1}{9} - 19$$

$$g_2(\underline{x}) \quad W_2 = 9 \;\; W_1 = -3 \;\; W_0 = -\frac{1}{2}\log\frac{1}{9} - 6$$

$$g_1(\underline{x}) \quad W_2 = 0 \;\; W_1 = -1 \;\; W_0 = -\frac{1}{2}\log\frac{1}{9} - \frac{1}{2}$$

Each of the $g_i(\underline{x})$ represent a discriminator plane in the *3D* space.

(c) Here are the plots of distributions and discriminators below. These are coded in Python using *PyLab*.

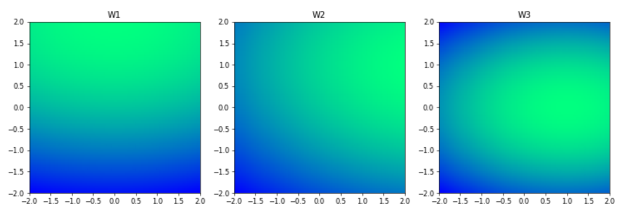

Figure 1.1: Distributions of three classes described in the problem description.

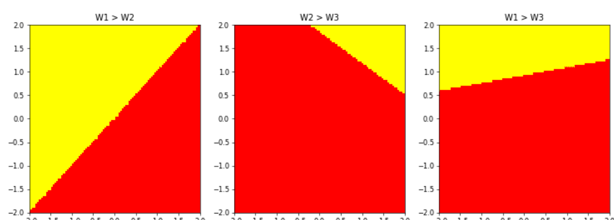

Figure 1.2: Linear discriminators of Figure 1.1 distributions.

2 Bayes Decision Rule & Bayes Error Boundaries

Consider the following 2-class classification problem involving a single feature x. Assume equal class priors and $0-1$ loss function.

$$p(x \mid w_1) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & otherwise \end{cases} \qquad p(x \mid w_2) = \begin{cases} 2-2x & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

(a) Sketch the two densities.

(b) State the Bayes decision rule and show the decision boundary.

(c) What is the Bayes classification error?

(d) How will the decision boundary change if the prior for class w1 is increased to 0.7?

Solution

(a) Figure 2.1, illustrates the density functions of these two classes. I've used the *Seaborn* library to generate these density functions.

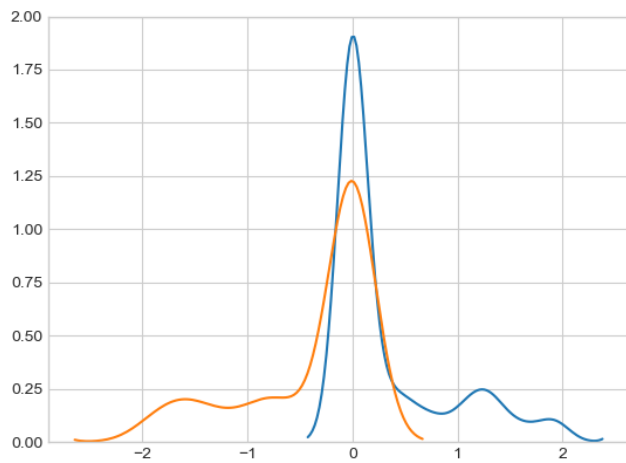

Figure 2.1: Illustration of density functions of w_1(Blue) and w_2(Orange).

(b) We drive the Bayes decision rule for these two classes below. $g_1(x)$ and $g_2(x)$ represent the decision function for the classes 1 and 2 respectively.

$$g_i(x) \underset{\omega_j}{\overset{\omega_i}{\gtrless}} g_j(x)$$

which is our decision baseline for the Bayes classifier. Since $g_i(x) = p(\omega_1 \mid x)$, expanding the equation according to the Bayes rule and we get:

$$g_i(x) = \frac{p(\underline{x} \mid \omega_i)P(\omega_i)}{p(\underline{x})}$$

Replacing the $g_i(x)$ in the decision baseline and we'll have the following results.

$$\frac{p(\underline{x} \mid \omega_i)P(\omega_i)}{p(\underline{x})} \underset{<}{\overset{\omega_i}{>}}_{\omega_j} \frac{p(\underline{x} \mid \omega_j)P(\omega_j)}{p(\underline{x})}$$

Omitting the constant parts from both sides and replacing the equations from the problem description will result in the following decision function.

$$g(x) = 4x - 2 \underset{<}{\overset{\omega_i}{>}}_{\omega_j} 0 \tag{2.1}$$

Thus, the linear discriminant function can be displayed as so:

$$g(x) = 4x - 2$$

in which the point $x = \frac{1}{2}$ is the separation point of two classes. The values greater than $\frac{1}{2}$ are assigned a label from class i. The values less the $\frac{1}{2}$ are assigned a label of class j.

(c) Here is the Bayes classification error given in (2.2).

$$\varepsilon = \varepsilon_1 P(\omega_1) + \varepsilon_2 P(\omega_2) \tag{2.2}$$

in which the ε_1 and ε_2 represent the probability of class 1 error by integrating the class 1 density over the region of class 2 and the probability of class 2 error by integrating the class 2 density over the region of class 1 respectively.

$$\varepsilon_1 = \int_{R_2} p(\underline{x} \mid \omega_1) d\underline{x}$$

$$\varepsilon_2 = \int_{R_1} p(\underline{x} \mid \omega_2) d\underline{x}$$

According the section (b), the discriminating point is $x = 0.5$. Correspondingly, the regions R_1 and R_2 can be easily driven like so:

$$R_1 = [0\ 0.5] \quad R_2 = [0.5\ 1]$$

By integrating the given equation (2.2) over the boundaries of these two regions, we'll have the following:

$$\varepsilon_1 = \int_0^{0.5} (2x) dx = \frac{1}{4}$$

$$\varepsilon_2 = \int_{0.5}^{1} (2 - 2x) dx = \frac{1}{4}$$

The final value for the Bayes error will be:

$$\varepsilon = \frac{1}{4} * \frac{1}{2} + \frac{1}{4} * \frac{1}{2} = \frac{1}{4}$$

(d) Changing the prior probabilities for classes ω_1 and ω_2, the bias will be changed. We'll have the following biases as the prior probabilities.

$$P(\omega_1) = 0.7$$

$$P(\omega_2) = 0.3$$

Rewriting the likelihood ratio for these two classes, we'll have the following results:

$$\frac{p(\underline{x} \mid \omega_1)}{p(\underline{x} \mid \omega_2)} \overset{\omega_1}{\underset{\omega_2}{\gtrless}} \frac{P(\omega_2)}{P(\omega_1)}$$

$$\frac{2x}{2-2x} \overset{\omega_1}{\underset{\omega_2}{\gtrless}} \frac{3}{7}$$

which changes the final discriminant function, $g(x)$ to

$$g'(x) = 10x - 3 \overset{\omega_1}{\underset{\omega_2}{\gtrless}} 0$$

3 Bayes Decision Boundary & Bhattacharyya Error Bound

Consider a two-category classification problem in two dimensions with

$$p(x \mid w_1) \sim N(0, I), \quad p(x \mid w_2) \sim N(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, I)$$

and

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}$$

(a) Calculate the Bayes Decision Boundary.

(b) Calculate the Bhattacharyya error bound.

(c) Repeat the above for the same probabilities, but

$$p(x \mid w_1) \sim N(0, \begin{bmatrix} 2 & 0.5 \\ 0.5 & 2 \end{bmatrix}), \quad p(x \mid w_2) \sim N(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix})$$

Solution

(a) The genera for of a Bayesian discriminator was discussed in (1.1). Comparing this to the linear classifier which can be displayed as below:

$$g_i(\underline{x}) = W_{i1}^T \underline{x} + W_{i0}$$

where the weights are also given below:

$$W_{i1} = \frac{\mu_i}{\sigma^2}$$
$$W_{i0} = \frac{-1}{2\sigma^2}\mu_i^T \mu_i + \log P(\omega_i)$$

Since $\Sigma_i = \sigma^2 I$ with $\sigma^2 = 1$(according to problem description) with equal prior probabilities($P(\omega_1) = P(\omega_2) = \frac{1}{2}$, the discriminator can be derived using the *Euclidean Distance* of x and μ.

$$g_i(\underline{x}) = (\underline{x} - \underline{\mu}_i)^T (\underline{x} - \underline{\mu}_i)$$

we can simply find the $g_1(\underline{x})$ and $g_2(\underline{x})$ by replacing the parameters in the problem description. We'll get:

$$g_1(\underline{x}) = (\underline{x} - 0)^T (\underline{x} - 0) = (\underline{x})^2$$
$$g_2(\underline{x}) = (\underline{x} - \begin{bmatrix}1\\1\end{bmatrix})^T (\underline{x} - \begin{bmatrix}1\\1\end{bmatrix}) = (\underline{x} - 1)^2$$

In order to find the decision boundary, we'll coincide the two discriminators:

$$\underline{x} = \frac{1}{2}$$

Meaning that the decision boundary is the plane $\underline{x} = \frac{1}{2}$.

(b) The *Bhattacharyya* error bound is an specific condition of *Chernoff* error bound. This condition happens when $s = \frac{1}{2}$ in the *Chernoff* bound formula. The *Bhattacharyya* formula is given below.

$$\varepsilon_{n-B} = \sqrt{P(\omega_1)P(\omega_2)} \int \sqrt{p(\underline{x} \mid \omega_1)p(\underline{x} \mid \omega_2)}d\underline{x} = e^{-\mu(s=\frac{1}{2})} \tag{3.1}$$

If we have access to the parameters of two distributions we can derive the *Bhattacharyya* error bound by computing the $\mu(s = \frac{1}{2})$ and replacing the result in the $e^{-\mu(s=\frac{1}{2})}$.

$$\mu(\frac{1}{2}) = \frac{1}{8}(\underline{m}_2 - \underline{m}_1)^T (\frac{\Sigma_1 + \Sigma_2}{2})^{-1} (\underline{m}_2 - \underline{m}_1) + \frac{1}{2} \ln \frac{|\frac{\Sigma_1+\Sigma_2}{2}|}{\sqrt{|\Sigma_1||\Sigma_2|}}$$

after some minor matrix multiplication, we'll get the following results:

$$\mu(\frac{1}{2}) = (\frac{1}{8})(4) = \frac{1}{2}$$

The *Bhattacharyya* error bound will be $\varepsilon_{n-B} = e^{-\frac{1}{2}}$.

(c) $g_1(\underset{\sim}{x})$ and $g_2(\underset{\sim}{x})$ can be easily computed using the *general form* of the Bayes classifier.

$$g_1(\underset{\sim}{x}) = -\frac{1}{2}(\underset{\sim}{x} - 0)^T \begin{bmatrix} \frac{8}{15} & \frac{2}{15} \\ \frac{2}{15} & \frac{8}{15} \end{bmatrix} (\underset{\sim}{x} - 0) - \frac{1}{2}\log\frac{15}{4} + \log\frac{1}{2}$$

$$g_2(\underset{\sim}{x}) = -\frac{1}{2}(\underset{\sim}{x} - \begin{bmatrix} 1 \\ 1 \end{bmatrix})^T \begin{bmatrix} \frac{5}{9} & \frac{4}{9} \\ \frac{4}{9} & \frac{5}{9} \end{bmatrix} (\underset{\sim}{x} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}) - \frac{1}{2}\log 9 + \log\frac{1}{2}$$

Coinciding the two decision functions and we get the following decision boundary:

$$\underset{\sim}{x}^T \Sigma_1^{-1} \underset{\sim}{x} - (\underset{\sim}{x} - 1)^T \Sigma_2^{-1} (\underset{\sim}{x} - 1) + \log\frac{5}{12} = 0$$

which appears to be a *Hyper-ellipsoid*. Furthermore, the *Bhattacharyya* error bound can be calculated as following.

$$\mu(\frac{1}{2}) = \frac{1}{8}(\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix})^T (\frac{\begin{bmatrix} 7 & 4.5 \\ 4.5 & 7 \end{bmatrix}}{2})^{-1} (\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix}) + \frac{1}{2}\ln\frac{\frac{\begin{vmatrix} 7 & 4.5 \\ 4.5 & 7 \end{vmatrix}}{2}}{\sqrt{\begin{vmatrix} 2 & 0.5 \\ 0.5 & 2 \end{vmatrix} \begin{vmatrix} 5 & 4 \\ 4 & 5 \end{vmatrix}}}$$

which results in $\mu(\frac{1}{2}) = -0.48$. Finally, the *Bhattacharyya* error bound result is given below.

$$\varepsilon_{n-B} = e^{0.48}$$

4 Bayes Decision Boundary & Dataset Samples

Consider the two-dimensional data points from two classes ω_1 and ω_2 below. Each of them are coming from a Gaussian distribution $p(x \mid \omega_k) \sim N(\mu_k, \Sigma_k)$.

ω_1	ω_2
(0, 0)	(6, 9)
(0, 1)	(8, 9)
(2, 2)	(9, 8)
(3, 1)	(9, 9)
(3, 2)	(9, 10)
(3, 3)	(8, 11)

Table 4.1: Data points from class ω_1 and ω_2.

(a) What is the prior probability for each class?

(b) Calculate the mean and covariance matrix for each class.

(c) Derive the equation for the decision boundary that separates these two classes, and plot the boundary.

(d) Think of the case that the penalties for misclassification are different for the two classes (i.e. not zero-one loss), will it affect the decision boundary, and how?

Solution

(a) The prior probability can be estimated using the following formula.

$$P(\omega_i) = \frac{|S| \quad \epsilon \ \omega_i}{|S|} \tag{4.1}$$

this value is the same for both of these classes and its equal to $P(\omega_1) = P(\omega_2) = \frac{1}{2}$.

(b) For the covariance matrix, diagonal elements are computed as below.

$$\sigma_{11}^2 = \frac{\sum_{i=1}^{N} x_1[i]}{N} - \mu_{11} \quad \sigma_{22}^2 = \frac{\sum_{i=1}^{N} x_2[i]}{N} - \mu_{12}$$

the non-diagonal elements are computed as below.

$$\sigma_{12}^2 = \frac{1}{n}\sum_{i=1}^{N}(x_i - E[x])(y_i - E[x])$$

and same for the other element. The results would be:

$$\underline{\mu}_1 = \begin{bmatrix} 1.83 \\ 1.5 \end{bmatrix} \quad \underline{\mu}_2 = \begin{bmatrix} 8.16 \\ 9.33 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 2.16 & 1.1 \\ 1.1 & 1.1 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 1.36 & -0.06 \\ -0.06 & 1.06 \end{bmatrix}$$

(c) The general case for the Bayes decision is given below. formerly, we have been working with the exact equation given in (1.2).

$$g_i(\underline{x}) = \underline{x}^T W_i \underline{x} + w_i^T \underline{x} + w_{i0} \tag{4.2}$$

The weights are computed as following.

$$W_i = -\frac{1}{2}\Sigma_i^{-1}$$

$$w_i = \Sigma_i^{-1}\underline{\mu}_i$$

$$w_{i0} = -\frac{1}{2}\underline{\mu}_i^T\Sigma_i^{-1}\underline{\mu}_i - \frac{1}{2}\log(|\Sigma_i|) + \log P(\omega_i)$$

Saving some time, we'll be devoting the burden of this computation to Python!. The result will be:

$$W_1 = \begin{bmatrix} -0.46 & 0.46 \\ 0.46 & -0.92 \end{bmatrix}$$

$$w_1 = \begin{bmatrix} 0.31 & 1.05 \end{bmatrix}$$

$$w_{10} = -1.84$$

The corresponding result will be as following for the second discriminator.

$$W_1 = \begin{bmatrix} -0.36 & -0.02 \\ -0.02 & -0.47 \end{bmatrix}$$

$$w_1 = \begin{bmatrix} 6.42 & 9.15 \end{bmatrix}$$

$$w_{10} = -69.80$$

Simplifying the weights given above in the equation $g_1(x) = g_2(x)$, we'll get the following results.

$$-0.1x^2 - 0.45y^2 - 6.11x - 8.1y + 0.96xy + 67.96 = 0 \tag{4.3}$$

Here is the plotted results.

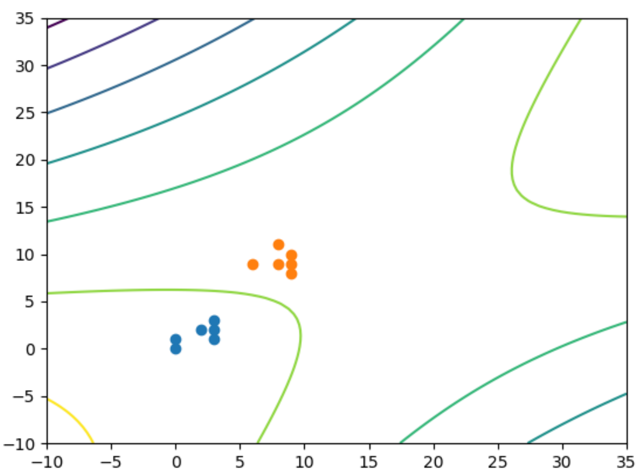

Figure 4.1: Contour lines for the Bayesian discriminator (4.3)

(d) If the classification is based on the misclassification penalty; for example when the penalty for the class 1 being misclassified as class 2 is 0.08 and the other penalty is 0.02, the classifier tends to assign most of the samples to the class 2 because it will cost less. So, changing the misclassification penalty would definitely exerts influence on the classification task.

5 Decision Boundaries of Exponential & Uniform Distributions

Consider a classification problem with 2 classes and a single real-valued feature vector X. for class 1, $p(x \mid c_1)$ is uniform $U(a, b)$ with $a = 2$ and $b = 4$. For class 2, $p(x \mid c_2)$ is exponential with density $\lambda \exp(-\lambda x)$ where $\lambda = 1$. Let $P(c_1) = P(c_2) = 0.5$.

(a) Determine the location of optimal decision regions.

(b) Draw a sketch of the two class densities multiplied by $P(c_1)$ and $P(c_2)$ respectively, as a function of x, clearly showing the optimal decision boundary.

(c) Compute the Bayes error rate for this problem within 3 decimal places of accuracy.

(d) Answer the questions above with $a = 2$ and $b = 22$.

Solution

We'll derive the *log likelihood* for these two classes.

$$\frac{p(\underset{\sim}{x} \mid c_1)}{p(\underset{\sim}{x} \mid c_2)} \underset{c_2}{\overset{c_1}{\gtrless}} \frac{P(c_2)}{P(c_1)}$$

The density function for the uniform distribution on an interval $[a, b]$ is equal to $f(x) = \frac{1}{b-a}$. In this case we'll have $p(\underset{\sim}{x} \mid c_1) = \frac{1}{2}$. Replacing the densities in the *likelihood ratio*:

$$\frac{\frac{1}{2}}{\exp(-x)} \underset{c_2}{\overset{c_1}{\gtrless}} \frac{\frac{1}{2}}{\frac{1}{2}}$$

Thus, the *log likelihood* will be:

$$\ln\frac{1}{2} - \ln\exp(-x) \underset{c_2}{\overset{c_1}{\gtrless}} 0$$

$$d(x) = x - 0.693 \underset{c_2}{\overset{c_1}{\gtrless}} 0$$

in which $d(x)$ is the decision boundary for the given classes.

(b) Here is the result for after plotting these two distributions. The green line illustrates the decision boundary.

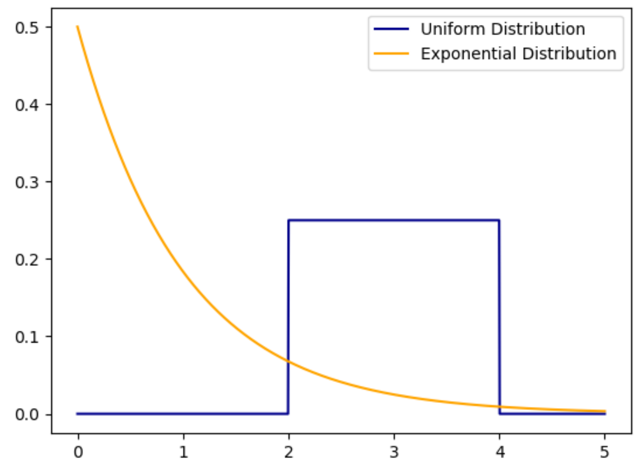

Figure 5.1: Class 1 and Class 2 Distribution Plots.

(c) According to the figure (5.2), the Bayes error can be calculated by integrating the exponential density over the period [2, 4].

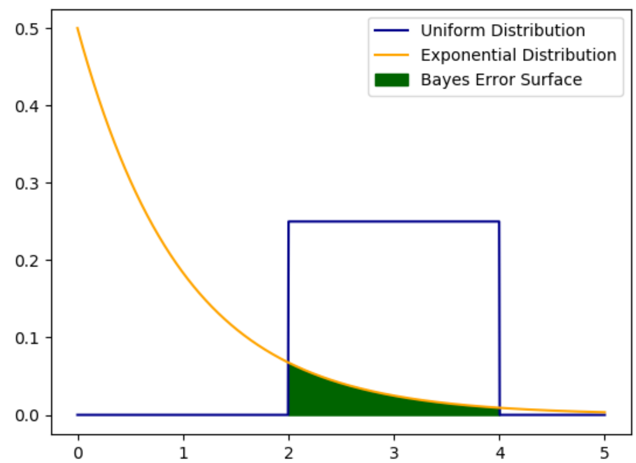

Figure 5.2: Bayes Error Surface Plot.

$$\varepsilon = \int_2^4 \frac{1}{2} \exp(-x) = -0.058$$

(d) We repeat the whole process with $a = 2$ and $b = 22$ again. The decision boundary will be as following:

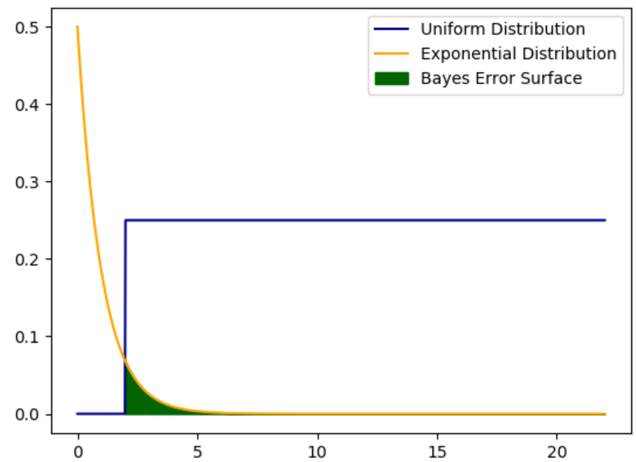

Figure 5.3: Bayes Error Surface Plot.

$$d(x) = x - 2.995 \underset{c_2}{\overset{c_1}{\gtrless}}$$

Here is the results after plotting the decision boundary(green line). The Bayes error for this condition is given below. The interval is $2 < x < 22$:

$$\varepsilon = \int_{2}^{22} \frac{1}{2} \exp(-x) = -0.067$$

6 Linear Discriminant Analysis & Bayes Classifier

Consider a 2-class classification problem with d-dimensional real-valued inputs x, where the class-conditional densities, $p(x \mid c_1)$ and $p(x \mid c_2)$ are multivariate Gaussian with different means $\underline{\mu}_1$ and $\underline{\mu}_2$ and a common covariance matrix Σ, with class probabilities $P(c_1)$ and $P(c_2)$.

(a) Write the discriminant function for this problem in the form of $g_1(\underline{x}) = \log p(\underline{x} \mid c_1) + \log P(c_1)$.

(b) prove that the optimal decision boundary can be written in the form of a linear discriminant, $\underline{w}\underline{x} + w_0 = 0$, where $\underline{w}$ is a d-dimensional weight vector and w_0 is a scalar, and clearly indicate what are $\underline{w}$ and w_0 are in terms of parameters of the classification model.

Solution

The multivariate normal density function is defined as

$$p(\underline{x} \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{\frac{-1}{2}(\underline{x}-\mu)^T\Sigma^{-1}(\underline{x}-\mu)} \tag{6.1}$$

Using Bayes rule, the *MAP* discriminant function becomes

$$g_1(\underline{x}) = p(\omega_i \mid \underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{\frac{-1}{2}(\underline{x}-\underline{\mu}_1)^T\Sigma^{-1}(\underline{x}-\underline{\mu}_1)} P(\omega_i)\frac{1}{p(\underline{x})}$$

Eliminating the constant term($p(\underline{x})$ which is not considered in discrimination) and taking natural log(since the logarithm is monotonically increasing function)

$$g_1(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_1)\Sigma^{-1}(\underline{x} - \underline{\mu}_1) - \frac{1}{2}\log(|\Sigma|) + \log(P(\omega_1))$$

Almost the same condition happens for the $g_2(\underline{x})$ with minor index changes:

$$g_2(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_2)\Sigma^{-1}(\underline{x} - \underline{\mu}_2) - \frac{1}{2}\log(|\Sigma|) + \log(P(\omega_1))$$

(b) This can be gone further by removing the constant term of $\log(|\Sigma|)$ from the above discriminants. We'll can derive the following results.

$$g_i(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_2)\Sigma^{-1}(\underline{x} - \underline{\mu}_2) + \log(P(\omega_1))$$

Expanding the quadratic term yields

$$g_i(\underline{x}) = -\frac{1}{2}(\underline{x}^T\Sigma^{-1}\underline{x} - 2\underline{\mu}_i^T\Sigma^{-1}\underline{x} + \underline{\mu}_i^T\Sigma^{-1}\underline{\mu}_i) + \log(P(\omega_1))$$

Removing the term $\underline{x}^T\Sigma^{-1}\underline{x}$ which is constant for all classes. Reorganizing terms we obtain

$$g_i(\underline{x}) = w_i^T x + w_{i0} \tag{6.2}$$

where w_i and w_{i0} are defined as:

$$w_i(\underline{x}) = \Sigma^{-1}\underline{\mu}_i \qquad w_{i0} = -\frac{1}{2}\underline{\mu}_i^T\Sigma^{-1}\underline{\mu}_i + \log P(\omega_i)$$

7 Bayesian Decision Plot

Solution

Here is the result after plotting the both points and Bayesian discriminator. Please refer to the *plot.py* in the directory of problem 7.

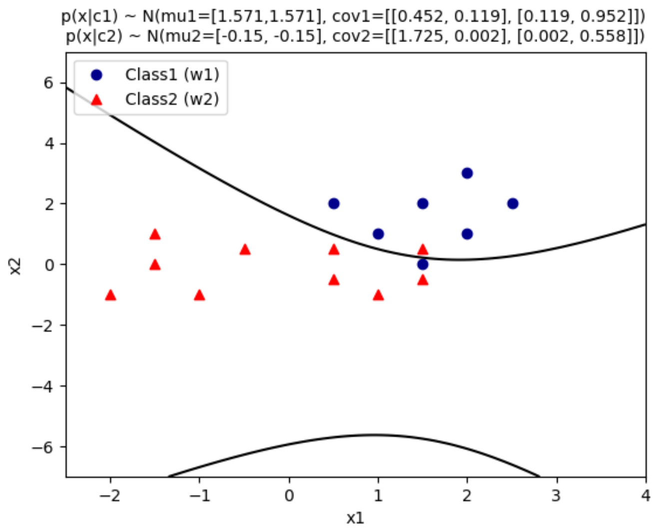

Figure 7.1: Decision region for the given scattered points.