

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:

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

and

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

- Calculate the discriminant function $g_i(x)$ for each class.
- Express your discriminant functions in the form of linear discriminant functions.
- Determine and plot the decision boundaries.

Solution

- The general form of a Bayesian discriminator is given below.

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

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

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

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:

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

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).

$$\begin{aligned} g_1(\underline{x}) &= -\frac{1}{2} \left(\frac{2x[1] * 0 + 2}{1} + \frac{2x[2] * 2 + 4}{\frac{1}{9}} \right) - \frac{1}{2} \log(1 * \frac{1}{9}) + ? \\ g_2(\underline{x}) &= -\frac{1}{2} \left(\frac{2x[1] * 3 + 3}{1} + \frac{2x[2] * 1 + 1}{\frac{1}{9}} \right) - \frac{1}{2} \log(1 * \frac{1}{9}) + ? \\ g_3(\underline{x}) &= -\frac{1}{2} \left(\frac{2x[1] * 1 + 1}{1} + \frac{2x[2] * 0 + 0}{\frac{1}{9}} \right) - \frac{1}{2} \log(1 * \frac{1}{9}) + ? \end{aligned}$$

The simplified results are

$$\begin{aligned} g_1(\underline{x}) &= -18x[2] - \frac{1}{2} \log \frac{1}{9} - 19 \\ g_2(\underline{x}) &= -3x[1] + 9x[2] - \frac{1}{2} \log \frac{1}{9} - 6 \\ g_3(\underline{x}) &= -x[1] - \frac{1}{2} \log \frac{1}{9} - \frac{1}{2} \end{aligned}$$

(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:

$$g_i(\underline{x}) = W_2x[2] + W_1x[1] + W_0$$

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

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

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

$$g_3(\underline{x}) \quad W_2 = 0 \quad W_1 = -1 \quad W_0 = -\frac{1}{2} \log \frac{1}{9} - \frac{1}{2}$$

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*.

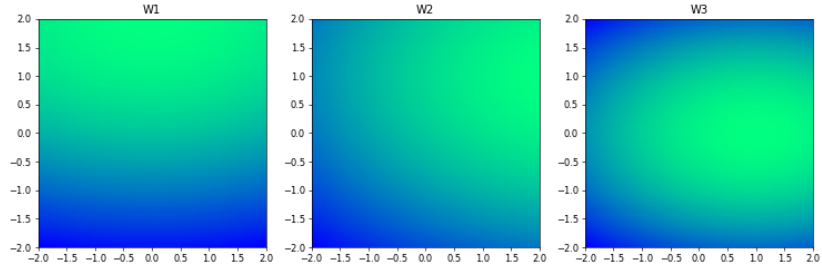


Figure 1.1: Distributions of three classes described in the problem description.

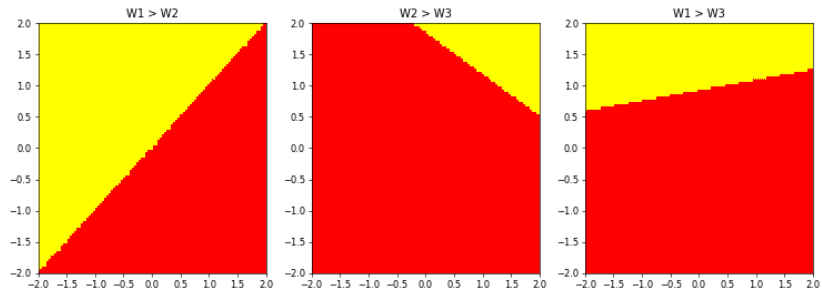


Figure 1.2: Linear discriminators of Figure 1.1 distributions.

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.

$$p(x | w_1) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad p(x | w_2) = \begin{cases} 2 - 2x & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- (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 w_1 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.

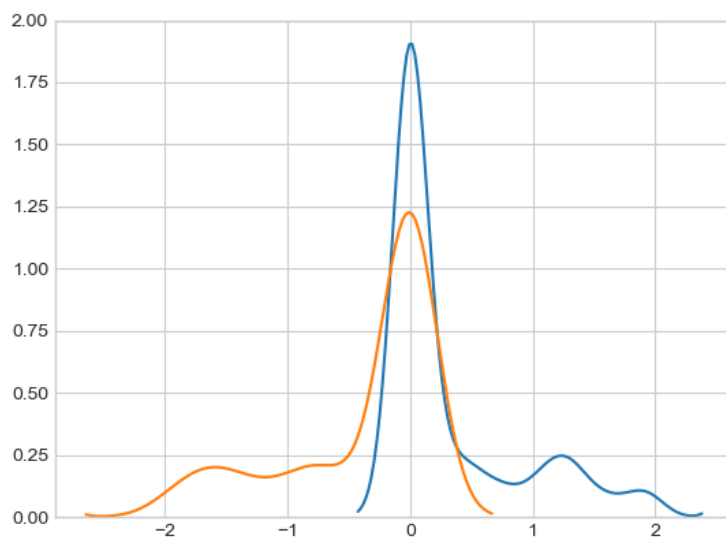


Figure 2.1: Illustration of density functions of w_1 (Blue) and w_2 (Orange).

(b) We derive 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.

$$g_i(x) \begin{matrix} \omega_i \\ > \\ \omega_j \\ < \end{matrix} g_j(x)$$

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

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

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

$$\frac{p(x | \omega_i)P(\omega_i)}{p(x)} \begin{matrix} \omega_i \\ > \\ \omega_j \\ < \end{matrix} \frac{p(x | \omega_j)P(\omega_j)}{p(x)}$$

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

$$g(x) = 4x - 2 \begin{matrix} \omega_i \\ > \\ \omega_j \\ < \end{matrix} 0 \quad (2.1)$$

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).

$$\varepsilon = \varepsilon_1 P(\omega_1) + \varepsilon_2 P(\omega_2) \quad (2.2)$$

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.

$$\varepsilon_1 = \int_{R_2} p(x | \omega_1) dx$$

$$\varepsilon_2 = \int_{R_1} p(x | \omega_2) dx$$

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:

$$\frac{p(x | \omega_1)}{p(x | \omega_2)} \stackrel{\omega_1}{>} \frac{P(\omega_2)}{P(\omega_1)}$$

$$\frac{2x}{2 - 2x} \stackrel{\omega_1}{>} \frac{3}{7}$$

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

$$g'(x) = 10x - 3 \stackrel{\omega_1}{>} \stackrel{\omega_2}{<} 0$$

3 Bayes Decision Boundary & Bhattacharyya Error Bound

Consider a two-category classification problem in two dimensions with

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

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

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

Solution

(a) The general form 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 a 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.

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

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 $e^{-\mu(s=\frac{1}{2})}$.

$$\mu\left(\frac{1}{2}\right) = \frac{1}{8} (m_2 - m_1)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (m_2 - m_1) + \frac{1}{2} \ln \frac{|\frac{\Sigma_1 + \Sigma_2}{2}|}{\sqrt{|\Sigma_1||\Sigma_2|}}$$

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

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

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

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

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

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

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

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

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 $p(x | \omega_k) \sim N(\mu_k, \Sigma_k)$.

| 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.

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

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.

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

the non-diagonal elements are computed as below.

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

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).

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

The weights are computed as following.

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

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

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

$$w_1 = [0.31 \ 1.05]$$

$$w_{10} = -1.84$$

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

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

$$w_1 = [6.42 \ 9.15]$$

$$w_{10} = -69.80$$

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