Assignment #2

Elements of Machine Learning

Saarland University – Winter Semester 2024/25

Rabin Adhikari
7072310
raad00002@stud.uni-saarland.de

Dhimitrios Duka
7059153
dhdu00001@stud.uni-saarland.de

Problem 3 (Linear & Quadratic Discriminate Analysis)

a) To begin with, we will summarize the data points for each class in the table below.

Table 1: Data for x_1, x_2, and class comparison.

Class 0		**Class 1**	
x_1	x_2	x_1	x_2
1	1	7	1
2	1	5	2
3	2	6	4
2	3	4	5
1	3	6	5

First, let's calculate the mean of each class.

$$\mu_0 = \frac{1}{5}\sum_{i=1}^{5} X_i = \frac{1}{5}\left(\begin{bmatrix}1\\1\end{bmatrix} + \begin{bmatrix}2\\1\end{bmatrix} + \begin{bmatrix}3\\2\end{bmatrix} + \begin{bmatrix}2\\3\end{bmatrix} + \begin{bmatrix}1\\3\end{bmatrix}\right) = \begin{bmatrix}1.8\\2\end{bmatrix} \tag{1}$$

Similarly, the mean of the second class μ_1 would be:

$$\mu_1 = \frac{1}{5}\sum_{i=6}^{10} X_i = \frac{1}{5}\left(\begin{bmatrix}7\\1\end{bmatrix} + \begin{bmatrix}5\\2\end{bmatrix} + \begin{bmatrix}6\\4\end{bmatrix} + \begin{bmatrix}4\\5\end{bmatrix} + \begin{bmatrix}6\\5\end{bmatrix}\right) = \begin{bmatrix}5.6\\3.4\end{bmatrix} \tag{2}$$

Now, we can calculate the covariance matrix for each class. In terms of expectations, the covariance matrix for class k is defined as follows:

$$\Sigma_k = \frac{n}{n-1}\begin{bmatrix}\mathbb{E}[(x_1^k)^2] - \mathbb{E}[x_1^k]^2 & \mathbb{E}[x_1^k x_2^k] - \mathbb{E}[x_1^k]\mathbb{E}[x_2^k] \\ \mathbb{E}[x_1^k x_2^k] - \mathbb{E}[x_1^k]\mathbb{E}[x_2^k] & \mathbb{E}[(x_2^k)^2] - \mathbb{E}[x_2^k]^2\end{bmatrix} \tag{3}$$

where the term $\frac{n}{n-1}$ is used to correct the bias in the estimation of the covariance matrix. The covariance matrix for class $k = 0$ would be:

$$\Sigma_0 = \frac{5}{4}\begin{bmatrix}0.56 & 0\\ 0 & 0.8\end{bmatrix} = \begin{bmatrix}0.7 & 0\\ 0 & 1\end{bmatrix} \tag{4}$$

Similarly, the covariance matrix for class $k = 1$ would be:

$$\Sigma_1 = \frac{5}{4}\begin{bmatrix} 1.04 & -0.84 \\ -0.84 & 2.64 \end{bmatrix} = \begin{bmatrix} 1.3 & -1.05 \\ -1.05 & 3.3 \end{bmatrix} \tag{5}$$

b) To determin in which class the new data point is going to be classified, we need to evaluate the difference between the discriminant functions for each class. The discriminant function for class k is defined as follows:

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2}\mu_k^T \Sigma^{-1} \mu_k + \log(\pi_k) \tag{6}$$

where π_k is the prior probability of class k. Because we have only two classes, we can write the difference between the discriminant functions for class $k = 0$ and $k = 1$ as follows:

$$\delta_0(x) - \delta_1(x) = x^T \Sigma^{-1} \mu_0 - \frac{1}{2}\mu_0^T \Sigma^{-1} \mu_0 + \log(\pi_0) - x^T \Sigma^{-1} \mu_1 + \frac{1}{2}\mu_1^T \Sigma^{-1} \mu_1 - \log(\pi_1) \tag{7}$$

Because the priors are equal, we ca rewrite the equation as follows:

$$\delta_0(x) - \delta_1(x) = x^T \Sigma^{-1} \mu_0 - \frac{1}{2}\mu_0^T \Sigma^{-1} \mu_0 - x^T \Sigma^{-1} \mu_1 + \frac{1}{2}\mu_1^T \Sigma^{-1} \mu_1 \tag{8}$$

If the difference is positive, it means that the new data point belongs to class $k = 0$. Otherwise, it belongs to class $k = 1$.

We must note however that the covariance matrix is the same for both discriminants. Therefore, we need to calculate the pooled covariance matrix Σ. This can be done as follows:

$$\Sigma = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1)\Sigma_i \tag{9}$$

where N is the total number of samples, k is the number of classes, and n_i is the number of samples in class i. In our case, $N = 10$, $k = 2$, and $n_0 = n_1 = 5$. Therefore, the pooled covariance matrix would be:

$$\Sigma = \frac{1}{10-2}\left(4\begin{bmatrix} 0.7 & 0 \\ 0 & 1 \end{bmatrix} + 4\begin{bmatrix} 1.3 & -1.05 \\ -1.05 & 3.3 \end{bmatrix}\right) = \begin{bmatrix} 1 & -0.525 \\ -0.525 & 2.15 \end{bmatrix} \tag{10}$$

Using numpy, we can calculate the inverse of the pooled covariance. The code is shown below:

```
import numpy as np
Sigma = np.array([[1, -0.525], [-0.525, 2.15]])
Sigma_inv = np.linalg.inv(Sigma)
```

Therefore, the inverse of the pooled covariance matrix is:

$$\Sigma^{-1} = \begin{bmatrix} 1.147 & 0.280 \\ 0.280 & 0.533 \end{bmatrix} \tag{11}$$

Now, we can calculate the difference between the discriminant functions for each class.

$$\begin{aligned} \delta_0(x) - \delta_1(x) = & \begin{bmatrix} 3.5 \\ 2 \end{bmatrix}^T \begin{bmatrix} 1.147 & 0.280 \\ 0.280 & 0.533 \end{bmatrix} \begin{bmatrix} 1.8 \\ 2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1.8 \\ 2 \end{bmatrix}^T \begin{bmatrix} 1.147 & 0.280 \\ 0.280 & 0.533 \end{bmatrix} \begin{bmatrix} 1.8 \\ 2 \end{bmatrix} \\ & - \begin{bmatrix} 3.5 \\ 2 \end{bmatrix}^T \begin{bmatrix} 1.147 & 0.280 \\ 0.280 & 0.533 \end{bmatrix} \begin{bmatrix} 5.6 \\ 3.4 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 5.6 \\ 3.4 \end{bmatrix}^T \begin{bmatrix} 1.147 & 0.280 \\ 0.280 & 0.533 \end{bmatrix} \begin{bmatrix} 5.6 \\ 3.4 \end{bmatrix} \\ \delta_0(x) - \delta_1(x) = & \ 2.218 \end{aligned} \tag{12}$$

This means that the new data point would be classified as class $k = 0$.

c) Both LDA and QDA assume that the data of each class is normally distributed. However, LDA assumes that the convariance matrix is the same for all classes, which usually is not the case. On the other hand, QDA lifts this restriction and allows for different covariance matrices for each class. This makes QDA more flexible and capable of capturing more complex decision boundaries.