Machine Learning

BLG 527E

Homework 2

Seyyid Osman Sevgili
504221565

2024 - Spring

Contents

Grade Expectation Table

		Q1	Q2	Q3	Total
Grade	Max	4	3	3	10
	Expected	4	3	3	10

Table 1: Grade Expectation Table

1 Question 1

1.1 a

In this section, i generated the data from given mean and variance matricies. Figure 1 shows the generated data on the 2D space.

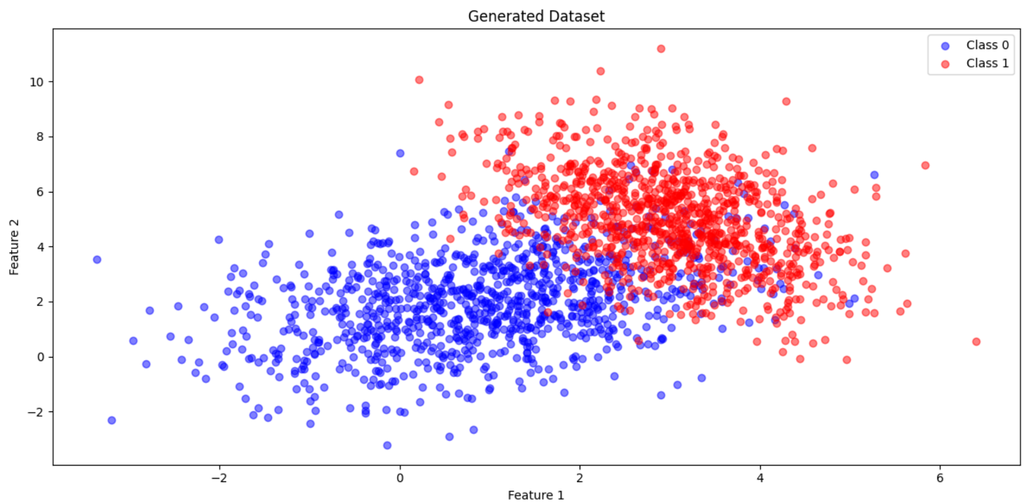

Figure 1: Generated Dataset

1.2 b

In this section, I partitioned the data into training and test sets. After that, I wrote custom functions to calculate the mean and covariance matrix from scratch. The formulas used for this step as follows.

$$E[\mathbf{x}] = \boldsymbol{\mu} = [\mu_1, \ldots, \mu_d]^T \tag{1}$$

$$\sigma_{ij} \equiv \mathrm{Cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = E[X_i X_j] - \mu_i \mu_j \tag{2}$$

$$\Sigma \equiv \mathrm{Cov}(\mathbf{X}) = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] = E[\mathbf{X}\mathbf{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T \tag{3}$$

Please refer to `src/notebooks/q1.ipynb` and `src/utils.py` files for detailed implementation steps.

1.3 c

In this section, I designed the Quadratic Discriminant Analysis algorithm from scratch to accurately predict the classes of the given features. The formula used for this step as follows.

$$g_i(x) = -\frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(x - m_i)^T \Sigma_i^{-1}(x - m_i) + \log \hat{P}(C_i) \tag{4}$$

As a result obtained training and test errors shown in Table 2.

Error Type	**1 - Accuracy**
Train Error (QDA)	0.09125
Test Error (QDA)	0.09

Table 2: QDA Results

Figure 2 shows the decision boundaries that is derived from QDA.

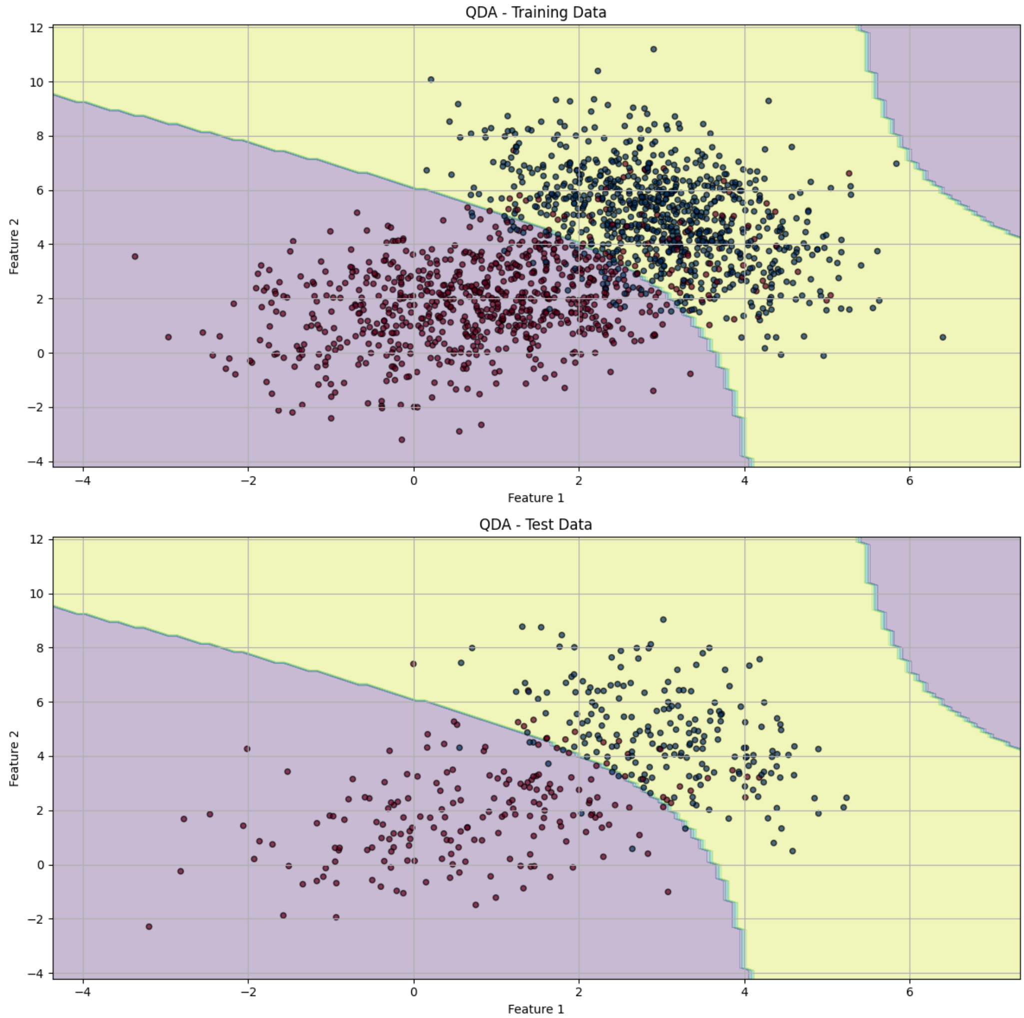

Figure 2: QDA Decision Boundaries

Please refer to `src/notebooks/q1.ipynb` and `src/qda.py` files for detailed implementation steps.

1.4 d

In this section, I designed the Lineer Discriminant Analysis algorithm from scratch to accurately predict the classes of the given features. The formula used for this step as follows.

$$g_i(x) = -\frac{1}{2}(x - m_i)^T \Sigma^{-1}(x - m_i) + \log \hat{P}(C_i) \tag{5}$$

As a result obtained training and test errors shown in Table 3.

Error Type	1 - Accuracy
Train Error (QDA)	0.10125
Test Error (QDA)	0.09

Table 3: LDA Results

Figure 3 shows the decision boundaries that is derived from LDA.

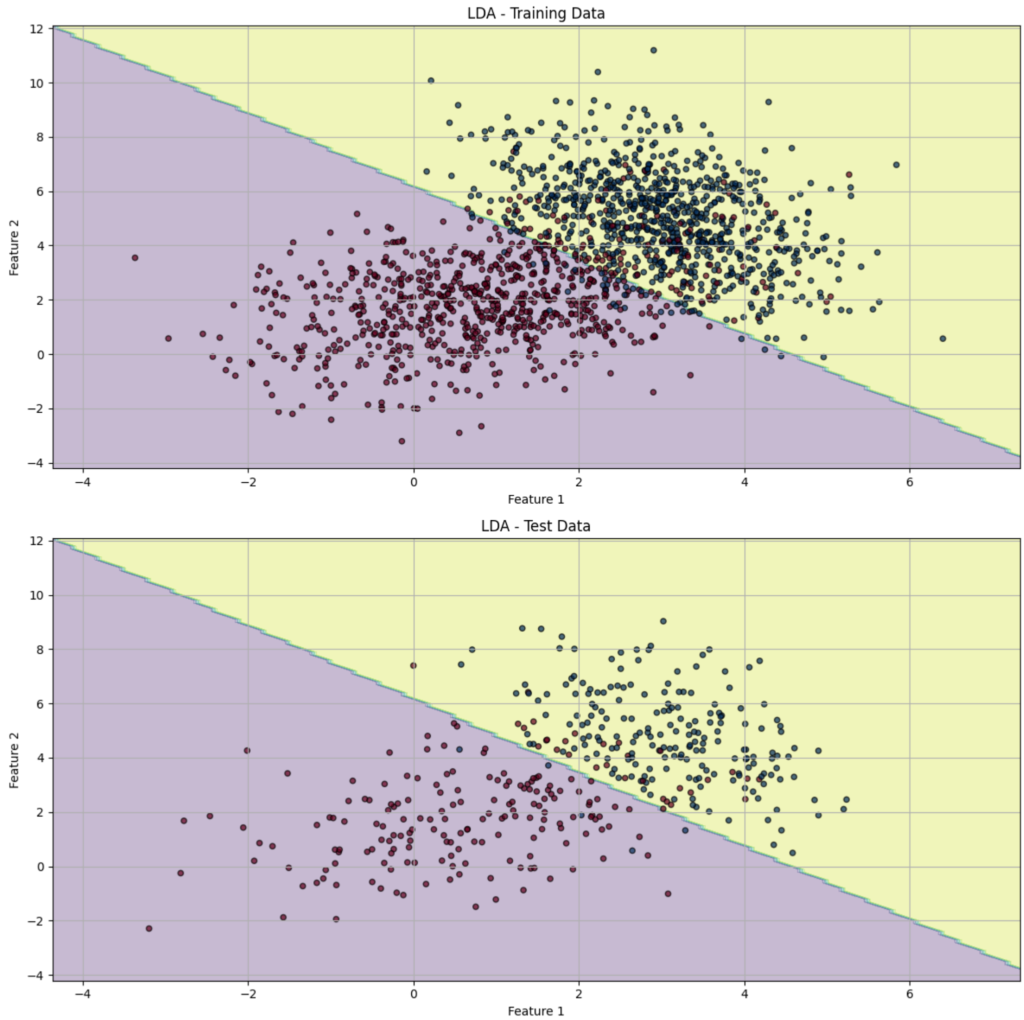

Figure 3: LDA Decision Boundaries

As we can see from Figure 3 and Table 3, the training set error increased because the data can be split more effectively using quadratic functions. When we used a linear decision boundary, the error increased. On the other hand, the test error remained nearly the same, as the test data can be separated using a linear decision boundary.

Please refer to `src/notebooks/q1.ipynb` and `src/lda.py` files for detailed implementation steps.

2 Question 2

In this question i implemented the Principal Compenent Aanalysis (PCA) from scratch for given the opdigits dataset. The steps for PCA can be written as follows.

1. Standardize the data
$$x'_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{6}$$

2. Compute the covariance matrix
$$C = \frac{1}{n-1} X'^{\top} X' \tag{7}$$

3. Compute eigenvalues and eigenvectors
$$C v_k = \lambda_k v_k \tag{8}$$

4. Sort and select principal components
$$V_k = \text{Sorted eigenvectors corresponding to top } k \text{ eigenvalues} \tag{9}$$

5. Transform the data
$$Z = X' V_k \tag{10}$$

As a result, Figure 4 shows the dataset represented using the top 2 principal components (eigenvectors). The figure indicates that by using only these 2 components, we can mostly distinguish between the labels. A key objective of PCA is to capture as much variance as possible while reducing the dimensionality of the data, and this representation demonstrates that even with just 2 components, much of the original variance is retained.

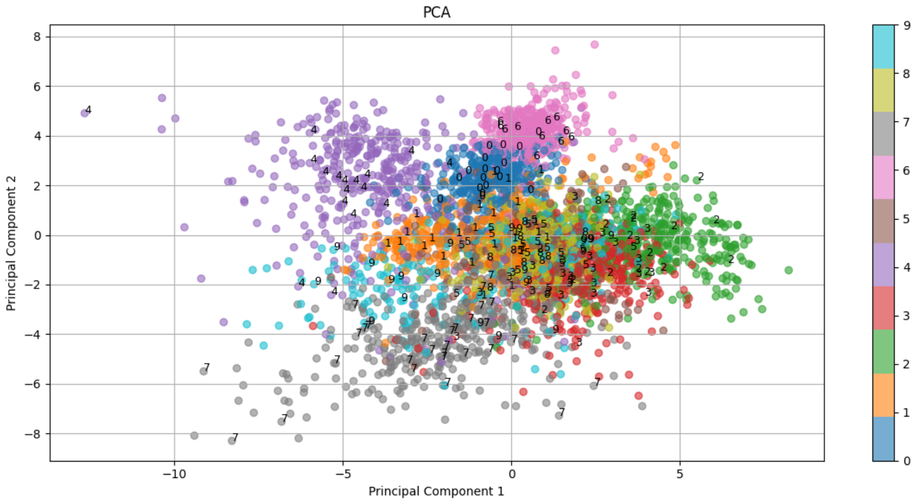

Figure 4: PCA

3 Question 3

3.1 a

I wrote the steps for the Expectation-Maximization algorithm based on the knowledge I gained from the course. The equations as follows.

$X = \{x^{(i)}\} \rightarrow$ observed data

$Z = \{z^{(i)}\} \rightarrow$ unobserved data

$\rightarrow$ log likelihood maximize it

$$L(\theta) = \sum_i \log p(x^{(i)}; \theta) = \sum_i \log \sum_{z^{(i)}} p(x^{(i)}, z^{(i)}; \theta)$$

E-Step $\rightarrow$ Computes the expected value of the complete log likelihood with respect to posterior distribution of the Z

$$Q(\theta, \theta^{(t)}) = \sum_i \sum_{z^{(i)}} p(z^{(i)} | x^{(i)}; \theta^{(t)}) \log p(x^{(i)}, z^{(i)}; \theta)$$

$$p(z^{(i)} | x^{(i)}; \theta) = \frac{p(x^{(i)}, z^{(i)}; \theta^{(t)})}{p(x^{(i)}; \theta^{(t)})} = \frac{p(x^{(i)}, z^{(i)}; \theta^{(t)})}{\sum_{z^{(i)}} p(x^{(i)}, z^{(i)}; \theta^{(t)})}$$

M-Step

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta, \theta^{(t)})$$

$$= \underset{\theta}{\operatorname{argmax}} \sum_i \sum_{z^{(i)}} p(z^{(i)} | x^{(i)}; \theta^{(t)}) \log p(x^{(i)}, z^{(i)}; \theta)$$

3.2 b

For the given problem on the homework description solution can be written as follows.

$\theta_A \rightarrow$ probability of getting Heads when coin "A" chosen.

$\theta_B \rightarrow$ probability of getting Heads when coin "B" chosen.

$\pi A \rightarrow$ probability of choosing coin "A"

$\pi B \rightarrow$ probability of choosing coin "B"

Step1: Assign random values initially

Lets, $\theta_A^{(0)} = 0.60$ $\quad \pi A^{(0)} = 0.5$

$\theta_B^{(0)} = 0.50$ $\quad \pi B^{(1)} = 0.5$

Step2: Expectation Step

calculate, $P(A \mid E)$ and $P(B \mid \underbrace{E})$ → Current experiment

$$P(\pi_A \mid E) = \frac{P(E \mid \pi_A)\ P(\pi_A)}{P(E)} = \frac{P(E \mid \pi_A)\ P(\pi_A)}{P(E \mid \pi_A) + P(E \mid \pi_B)}$$

$$P(E \mid \pi_A) = \binom{n}{x} \theta_A^{x} (1-\theta_A)^{n-x}$$

do same steps for coin "B". Then,

update the π_A and π_B values.

$$\pi_A = \frac{\sum_{i=0}^{n} P(\pi_A \mid E_{(i)})}{n}$$

$$\pi_B = 1 - \pi_A$$

Step 3 : Maximization Step

Update θ_A and θ_B

$$\theta_A = \frac{\sum_{i=0}^{n} P(\pi_A \mid E^{(i)}) \times X^{(i)}}{\sum_{i=0}^{n} P(\pi_A \mid E^{(i)}) \times n^{(i)}}$$

→ number of heads at this experiment

→ total coin flips

Same for the θ_B

Iterate until change is very small

I wrote the code for the explained algorithm above. The results for the initial values $Q_a = 0.6$, $Q_b = 0.5$, $\pi_a = 0.5$, and $\pi_b = 0.5$ are mentioned in Table 4.

Parameter	**Value**
Q_a	0.7934
Q_b	0.5139
π_a	0.5228
π_b	0.4772

Table 4: EM Results