Part. 1, Coding (60%):

Q1:

```
mean vector of class 1: [ 0.99253136 -0.99115481]
mean vector of class 2: [-0.9888012   1.00522778]
```

Q2:

```
Within-class scatter matrix SW:
[[ 4337.38546493 -1795.55656547]
 [-1795.55656547  2834.75834886]]
```

Q3:

```
Between-class scatter matrix SB:
[[ 3.92567873 -3.95549783]
 [-3.95549783  3.98554344]]
```

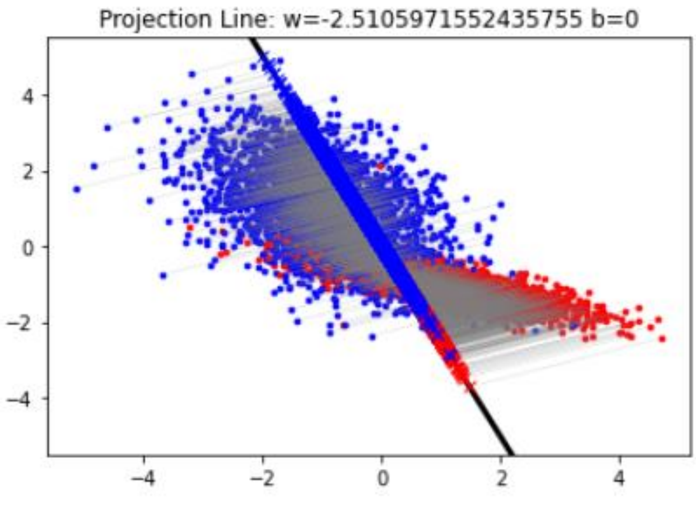
Q4:

```
Fisher’s linear discriminant: [ 0.37003809 -0.92901658]
```

Q5:

```
K=1 : Accuracy of test-set 0.8488
K=2 : Accuracy of test-set 0.8488
K=3 : Accuracy of test-set 0.88
K=4 : Accuracy of test-set 0.8824
K=5 : Accuracy of test-set 0.8904
```

Q6:



Part. 2, Questions (40%):

Q1:

PCA 是 unsupervised learning，要找到投影軸讓資料投影下去的分散量最大化，不過 PCA 不用知道資料的類別；LDA 則是 supervised learning，要找到資料投影下去後，最大的「組間分散量」(不同 class 間的差距越大越好)。

Q2:

2. Assume the dimension of input space is D, which is greater than K $(K>2)$

$$y = W^T x$$

when $K=2$

$$S_W = \sum_{n \in C_1} (x_n - m_1)(x_n - m_1)^T + \sum_{n \in C_2} (x_n - m_2)(x_n - m_2)^T$$

now:

$$S_W = \sum_{k=1}^{K} S_k, \text{ where } S_k = \sum_{n \in C_k} (x_n - m_k)(x_n - m_k)^T \text{ and } m_k = \frac{1}{N_k} \sum_{n \in C_k} x_n$$

when $K=2$

$$S_B = (m_2 - m_1)(m_2 - m_1)^T$$

now:

$$S_B = \sum_{k=1}^{K} N_k (m_k - m)(m_k - m)^T \text{ where } m = \frac{1}{N} \sum_{n=1}^{N} x_n$$

when $D' = 1$

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

The optimal w is the eigenvector of $S_W^{-1} S_B$ that corresponds to the largest eigenvalue

now:

$$J(W) = \mathrm{Tr}\{(W S_W W^T)^{-1} (W S_B W^T)\}$$

The columns of the optimal W are the eigenvectors of $S_W^{-1} S_B$ that correspond to the D' largest eigenvalues

Q3:

3. Since $J(w) = \dfrac{(m_2 - m_1)^2}{s_1^2 + s_2^2}$

$$= \frac{(w^T m_2 - w^T m_1)^2}{s_1^2 + s_2^2} \quad \cdots \text{ by } ①$$

$$= \frac{(w^T m_2 - w^T m_1)^2}{w^t S_1 w + w^t S_2 w} \quad \cdots \text{ by } ②$$

$$= \frac{w^t (m_1 - m_2)(m_1 - m_2)^t w}{w^t (S_1 + S_2) w}$$

$$= \frac{w^t S_B w}{w^t S_W w}$$

$$s_k^2 = \sum (y_n - m_k)^2$$

$$= \sum (w^t x - w^t m_k)^2$$

$$= \sum w^t (x - m_i)(x - m_i)^t w$$

$$= w^t S_k w \quad \cdots ②$$

$$S_B = (m_1 - m_2)(m_1 - m_2)^T \quad ①$$

$$S_W = (S_1 + S_2)$$

Q4:

4. $\dfrac{\partial E}{\partial a_k} = -t_k \dfrac{1}{y_k} \dfrac{\partial y_n}{\partial a_n} + (1 - t_k) \dfrac{1}{1 - y_k} \dfrac{\partial y_k}{\partial a_k}$

$$\frac{\partial y_k}{\partial a_k} = y_k (1 - y_k)$$

$$\frac{\partial E}{\partial a_k} = -t_n \frac{y_n (1 - y_n)}{y_n} + (1 - t_n) \frac{y_n (1 - y_n)}{(1 - y_n)}$$

$$= y_n - t_n$$

Q5:

5. $p(t \mid w_1, \ldots, w_k) = \prod_{k=1}^{K} y_k^{t_k}$

for N points, likelihood func would be

$$p(T \mid w_1, \ldots w_k) = \prod_{n=1}^{N} \prod_{k=1}^{K} y_{nk}^{t_{nk}}$$

taking the negative logarithm in order to derive an error function, we get

$$E(w) = -\sum_{n=1}^{N} \sum_{k=1}^{K} t_{kn} \ln y_k (x_n, w)$$