0816170 郭建良 NYCU Introduction to Machine Learning, Homework 2

Part 1

1. Mean vectors mi (i=1, 2) of each 2 classes

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

2. Within-class scatter matrix SW

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

3. Between-class scatter matrix SB

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

4. Fisher's linear discriminant W

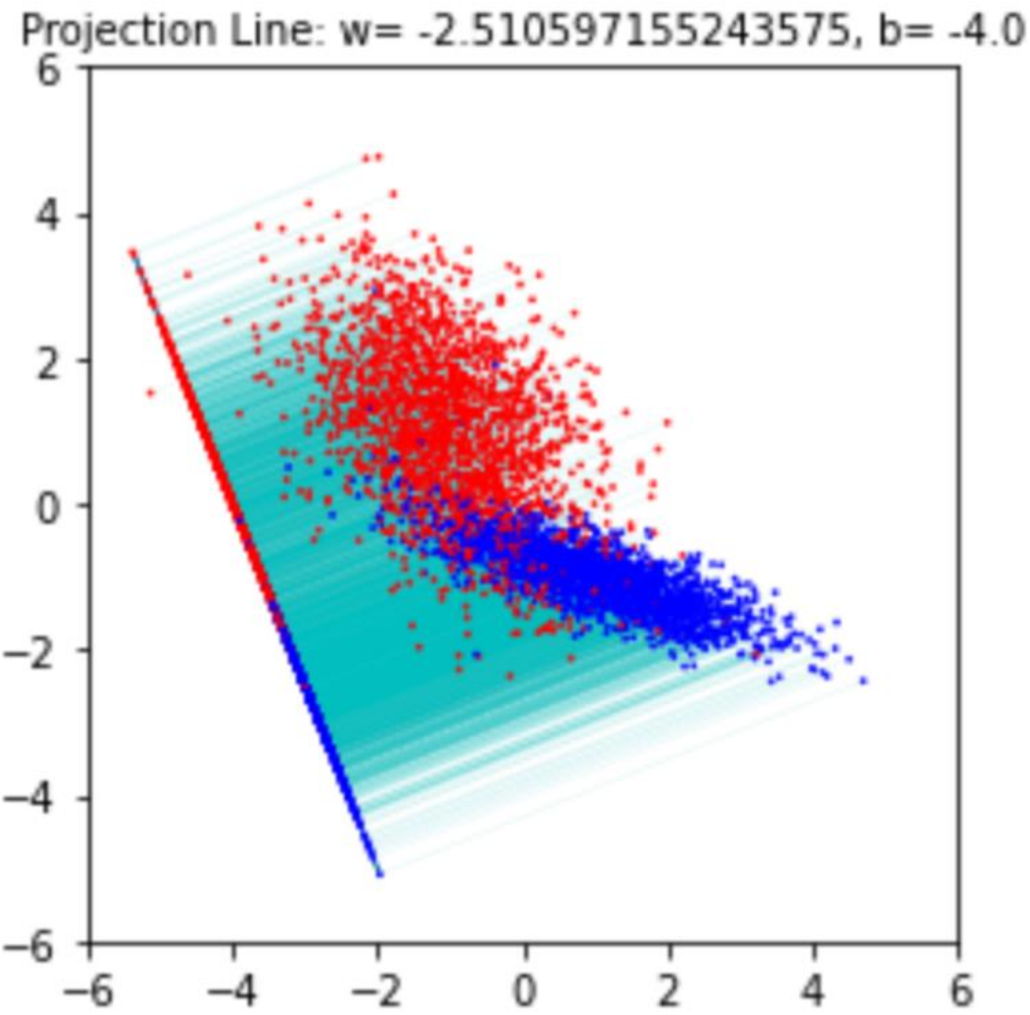
```
Fisher’s linear discriminant:
[[-0.000224  ]
 [ 0.00056237]]
After Normalization:
[[-0.37003809]
 [ 0.92901658]]
```

5. Accuracy score with K values from 1 to 5

(若 k 為偶數且兩種 class 數目相等，歸類於距離最近的 training node 的 class)

```
Accuracy of test-set when k = 1: 0.8488
Accuracy of test-set when k = 2: 0.8488
Accuracy of test-set when k = 3: 0.8792
Accuracy of test-set when k = 4: 0.8824
Accuracy of test-set when k = 5: 0.8912
```

6. Plot the best projection line, colorize the data and project all data points



Part 2

1 PCA maximizes the amount of information carried over onto smaller dimensions, and uses the principal components found through singular value decomposition.
On the other hand, Fisher's Linear Discriminant takes the categories in the data into account, while PCA does not.

2

2-class

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

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

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

$$\frac{\partial}{\partial w} J(w) = 0$$

$$\Rightarrow w \propto S_w^{-1}(m_2 - m_1)$$

mulit-class

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

$$\longrightarrow S_B = \sum_{k=1}^{K} N_k (m_k - m)(m_k - m)^T$$

$$\longrightarrow S_w = \sum_{k=1}^{K} S_k$$

$$S_k = \sum_{n \in C_k} (x_n - m_k)(x_n - m_k)^T$$

$\longrightarrow$ Find the weight vector w through taking the eigenvectors of $S_w^{-1} S_B$ that correspond to the largest eigenvalue

$$W = \max_{D'}(\mathrm{eig}(S_w^{-1} S_B))$$

3 According to Eq. 4

$$(m_2 - m_1)^2 = W^T (m_2 - m_1)(m_2 - m_1)^T W$$
$$= W^T S_B W \cdots\cdots ①$$

According to Eq. 1, 4, 5

$$S_k^2 = \sum_{n \in C_k} (y_n - m_k)^2$$
$$= \sum_{n \in C_k} (W^T x_n - W^T m_k)^2$$
$$= \sum_{n \in C_k} W^T (x_n - m_k)(x_n - m_k)^T W$$
$$= W^T S_k W \cdots\cdots ②$$

According to ②

$$S_1^2 + S_2^2 = W^T S_1 W + W^T S_2 W = W^T S_W W \cdots ③$$

$$\frac{①}{③} \Rightarrow \frac{(m_2 - m_1)^2}{S_1^2 + S_2^2} = \frac{W^T S_B W}{W^T S_W W}$$

4

$$\frac{\partial E}{\partial a_k} = \boxed{\frac{\partial E}{\partial y_k}} \frac{\partial y_k}{\partial a_k}$$

Softmax derivative

$$y_k = \frac{e^{a_k}}{\sum_N e^{a_n}} \qquad \frac{\partial y_k}{\partial a_k} = \frac{\left(e^{a_k} \sum_n e^{a_n}\right) - e^{a_k} \times e^{a_k}}{\left(\sum_n e^{a_n}\right)^2}$$

$$= \frac{e^{a_k}}{\sum_n e^{a_n}} - \left(\frac{e^{a_k}}{\sum_n e^{a_n}}\right)^2$$

$$= y_k - y_k^2$$

$$= y_k(1-y_k)$$

$$\frac{\partial E}{\partial y_k} = \frac{\partial}{\partial y_k}\left(-\sum_{n=1}^{N}\{t_n \ln y_n + (1-t_n)\ln(1-y_n)\}\right)$$

$$= -\frac{\partial}{\partial y_k}\left(\sum_{n=1}^{N}\{t_n \ln y_n + (1-t_n)\ln(1-y_n)\}\right)$$

$$= -\frac{\partial}{\partial y_k}\left(t_k \ln y_k + (1-t_k)\ln(1-y_k)\right)$$

$$= -\left(\frac{t_k}{y_k} - \frac{1-t_k}{1-y_k}\right) = -\frac{t_k}{y_k} + \frac{1-t_k}{1-y_k}$$

$$\Rightarrow \frac{\partial E}{\partial a_k} = \left(-\frac{t_k}{y_k} + \frac{1-t_k}{1-y_k}\right)\left(y_k(1-y_k)\right)$$

$$= -t_k + t_k y_k + y_k - t_k y_k$$

$$= y_k - t_k$$

5

$$y_k(x_n, w) = p(t_k = 1 \mid x_n)$$

$$p(T \mid w_1, w_2, \ldots w_k) = \prod_{n=1}^{N} \prod_{k=1}^{K} p(t_k = 1 \mid x_n)^{t_{kn}}$$

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} y_k(x_n, w)^{t_{kn}}$$

$$E(w) = -\ln p(T \mid w_1, w_2, \ldots w_k)$$

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

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

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

$\Downarrow$ minimize

maximize $p(T \mid w_1, w_2, \ldots w_k)$