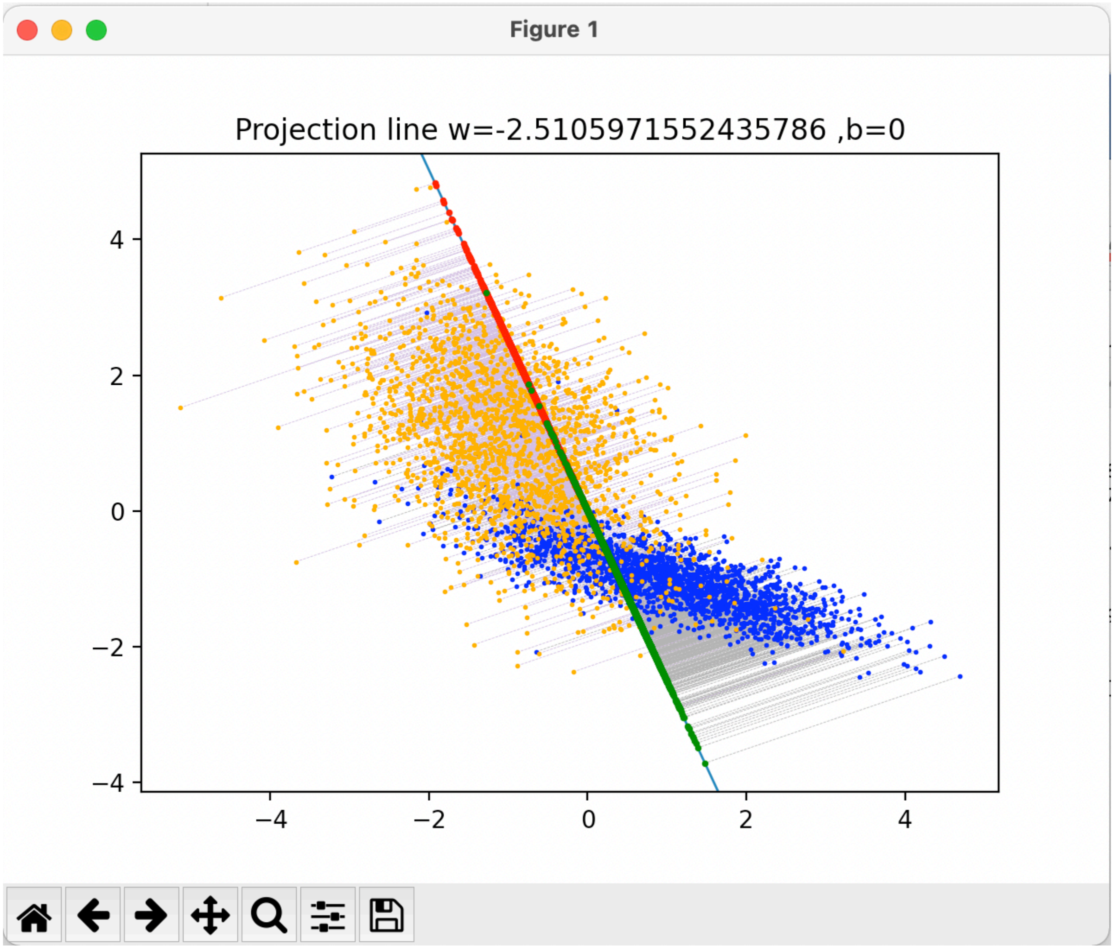
NYCU Introduction to Machine Learning, Homework 2

109550116 楊傑宇

Part. 1, Coding (60%):

1. mean vector of class 1: (0.9925313591759756, -0.9911548083897498)
   mean vector of class 2: (-0.9888012040932153, 1.0052277794804598)
2. Within-class scatter matrix SW: [[4337.38546493 -1795.55656547][-1795.55656547 2834.75834886]]
3. Between-class scatter matrix SB: [[3.92567873 -3.95549783][-3.95549783 3.98554344]]
4. Fisher's linear discriminant: [0.37003809 -0.92901658]
5. K= 1 :Accuracy of test-set 0.8488
   K= 2 :Accuracy of test-set 0.8488
   K= 3 :Accuracy of test-set 0.8792
   K= 4 :Accuracy of test-set 0.8824
   K= 5 :Accuracy of test-set 0.8912
6.



Part. 2, Questions (40%):

1.

- PCA： PCA 是希望讓資料投影後的分散量最大化，並需要須分類別，因此會避免重疊
- FLD: FLD 則是希望“不同類別的”分散量越大越好

2.

- 我認為從兩組要延伸至多組，最關鍵的差異就是在 Sw 與 Sb 的計算 Sw 從原本的兩個組內分散量到多個組內分散量，而 Sb 從兩組的組間分散量到各組間的分散量相加

3.

3.

Eq(6). $$\frac{(m_2-m_1)^2}{s_1^2+s_2^2}$$

$$\xRightarrow{Eq(3)} \frac{[W^T(m_2-m_1)]^2}{s_1^2+s_2^2} \Rightarrow \frac{W^T(m_2-m_1)(m_2-m_1)^T W^{T^T}}{s_1^2+s_2^2}$$

$$\Rightarrow \frac{W^T S_B W}{s_1^2+s_2^2} \xRightarrow{Eq(5)} \frac{W^T S_B W}{\sum_n (y_n-m_1)^2+\sum_{n'}(y_{n'}-m_2)^2}$$

$$\xRightarrow{Eq(4)\,Eq(1).} \frac{W^T S_B W}{\sum_n (W^T x_n - W^T m_1)^2 + \sum_{n'}(W^T x_{n'} - W^T m_2)^2}$$

$$\Rightarrow \frac{W^T S_B W}{\sum_n (W^T x_n - W^T m_1)(W^T x_n - W^T m_1)^T + \sum_{n'}(W^T x_{n'} - W^T m_2)(W^T x_{n'} - W^T m_2)^T}$$

$$\Rightarrow \frac{W^T S_B W}{\sum_n W^T (x_n-m_1)(x_n-m_1)^T W + \sum_{n'} W^T (x_{n'}-m_2)(x_{n'}-m_2)^T W}$$

$$\Rightarrow \frac{W^T S_B W}{W^T \left(\sum_n (x_n-m_1)(x_n-m_1)^T + \sum_{n'}(x_{n'}-m_2)(x_{n'}-m_2)^T\right) W}$$

$$\Rightarrow \frac{W^T S_B W}{W^T S_W W} \#$$

4、5.

4.

$$\frac{\partial E}{\partial a} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial a} \qquad \frac{dy}{da} = y(1-y)$$

$$= \Sigma\left(\frac{\partial(-t\log y - (1-t)\log(1-y))}{\partial y} \times \frac{dy}{da}\right)$$

$$= \Sigma\left[\left(\frac{-t}{y} + \frac{1-t}{1-y}\right) \times y(1-y)\right]$$

$$= \Sigma\left(\frac{y-t}{y(1-y)} \times y(1-y)\right) = \Sigma(y-t)$$

So for $\frac{\partial E}{\partial a_k}$ we take the part of k.

$$= (y_k - t_k) \ \#$$

5

for multiclass class the likelihood function:

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

取 log $= \prod_{n=1}^{N}\prod_{k=1}^{K} t_{nk} \ln y_{nk}$ — ①

cross entropy: $-\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln y_{nk}$ — ②

So we maximizing ① is equivalent minimizing ②.