NYCU Pattern Recognition, HW2

310551079 梁友誠

Part. 1, Coding:

1. mean vectors m_i (i=1, 2) of each 2 classes on training data.

```
mean vector of class 1: [2.47107265 1.97913899]
mean vector of class 2: [1.82380675 3.03051876]
```

2. within-class scatter matrix S_W on training data.

```
Within-class scatter matrix SW:
[[140.40036447  -5.30881553]
 [ -5.30881553 138.14297637]]
```

3. between-class scatter matrix S_B on training data.

```
Between-class scatter matrix SB:
[[ 0.41895314 -0.68052227]
 [-0.68052227  1.10539942]]
```

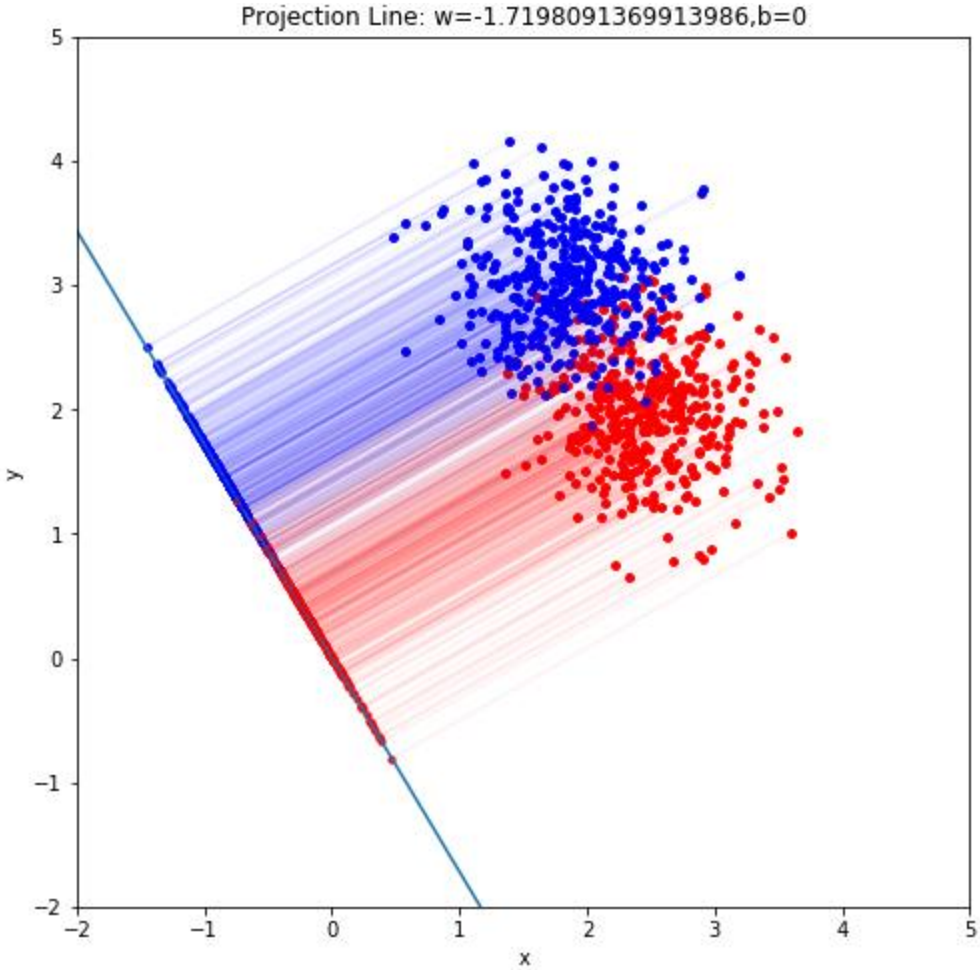
4. Fisher's linear discriminant w on training data.

```
 Fisher's linear discriminant:
[[ 0.50266214]
 [-0.86448295]]
```

5. accuracy score on testing data.

```
Accuracy of test-set 0.904
```

6. Plot the best projection line on the training data.



Part. 2, Questions:

Question 1.

We use Lagrange multiplier to enforce $w^T w = 1$, We need to maximize:

$$L(\lambda, w) = w^T(m_2 - m_1) + \lambda(w^T w - 1)$$

Calculate the derivatives:

$$\frac{\partial L(\lambda, w)}{\partial \lambda} = w^T w - 1$$

$$\frac{\partial L(\lambda, w)}{\partial w} = m_2 - m_1 + 2\lambda w$$

set the derivatives above equation to 0

$$\therefore \quad w^T w - 1 = 0 \Rightarrow w^T w = 1$$

$$m_2 - m_1 + 2\lambda w = 0 \Rightarrow w = -\frac{1}{2\lambda}(m_2 - m_1)$$

so we can get $w \propto (m_2 - m_1)$ #

Question 2.

We expand (eq 6) using (eq 1), (eq 2), (eq 3)

$$J(w) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}$$

$$= \frac{\| W^T (m_2 - m_1) \|^2}{\sum_{n \in C_1} (W^T x_n - m_1)^2 + \sum_{n \in C_2} (W^T x_n - m_2)^2}$$

The numerator can be rewritten as:

$$[W^T (m_2 - m_1)][W^T (m_2 - m_1)]^T$$

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

$$= W^T S_B W$$

And the denominator can be rewritten as:

$$\sum_{n \in C_1} [W^T (x_n - m_1)]^2 + \sum_{n \in C_2} [W^T (x_n - m_2)]^2$$

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

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

$$= W^T S_{W_1} W + W^T S_{W_2} W$$

$$= W^T S_W W$$

from above equations, we can get $J(w) = \dfrac{W^T S_B W}{W^T S_W W}$

Question 3.

$$E(w) = -\ln p(t|w) = -\sum_{n=1}^{N} \{ t_n \ln y_n + (1-t_n) \ln (1-y_n) \}$$

$$y_n = \sigma(a_n), \quad a_n = w^T \phi_n, \quad \frac{d\sigma}{da} = \sigma(1-\sigma)$$

$$\nabla E(w) = -\nabla \sum_{n=1}^{N} \{ t_n \ln y_n + (1-t_n) \ln (1-y_n) \}$$

$$= -\sum_{n=1}^{N} \nabla \{ t_n \ln y_n + (1-t_n) \ln (1-y_n) \}$$

$$= -\sum_{n=1}^{N} \frac{d \{ t_n \ln y_n + (1-t_n) \ln (1-y_n) \}}{d y_n} \cdot \frac{d y_n}{d a_n} \cdot \frac{d a_n}{d w}$$

$$= -\sum_{n=1}^{N} \left(\frac{t_n}{y_n} - \frac{1-t_n}{1-y_n} \right) \cdot y_n (1-y_n) \cdot \phi_n$$

$$= -\sum_{n=1}^{N} \frac{t_n - y_n}{y_n (1-y_n)} \cdot y_n (1-y_n) \cdot \phi_n$$

$$= -\sum_{n=1}^{N} (t_n - y_n) \cdot \phi_n$$

$$= \sum_{n=1}^{N} (y_n - t_n) \cdot \phi_n \quad \#$$