學號：410551027
姓名：江衍涵

Part 2：

1. Using Lagrange multiplier to constraint with $w^T w = 1$.

We now need to maximize

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

derivatives:

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

and

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

Setting the derivative above equals to 0, which gives:

$$w = -\frac{1}{2\lambda}(m_2 - m_1) \propto (m_2 - m_1)$$

2.

Using eq 2, eq 3 to expand eq 6,

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

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

The numerator can be written as:

$$\text{numerator} = [w^T(m_2 - m_1)][w^T(m_2 - m_1)]^T = w^T S_B w$$

Where we have defined:

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

And where it's the same for denominator,

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

$$= w^T S_{w_1} w + w^T S_{w_2} w$$

$$= w^T S_w w$$

Where we have defined:

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

3.

Follow the hint:

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

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

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

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

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

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

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

Where we have used $y_n = \sigma(a_n)$, $a_n = w^T \phi_n$, the chain rules and eq 8.