Q1. Using a Largrange multiplier, we have

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

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

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

Since $w^T w = 1$, then by setting $\frac{\partial}{\partial w} L(\lambda, w) = 0$, we can get the maximization.

$$m_2 - m_1 + 2\lambda w = 0$$

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

Q2 Let's define some notations.

$$\begin{cases} S_k = \sum_{n \in C_k} (y_n - m_k)^2 \\ S_W = S_1 + S_2 \\ S_B = (m_2 - m_1)(m_2 - m_1)^T \end{cases}$$

Since

$$\begin{aligned} s_k^2 &= \sum_{n \in C_k} (y_n - m_k)^2 \\ &= \sum_{n \in C_k} (w^T x - w^T m_k)^2 \\ &= \sum_{n \in C_k} w^T (x - m_k)(x - m_k)^T w \\ &= w^T S_k w \end{aligned}$$

then $s_1^2 + s_2^2 = w^T S_1 w + w^T S_2 w = w^T S_W w$

Since

$$\begin{aligned} (m_2 - m_1)^2 &= (w^T m_2 - w^T m_1)^2 \\ &= w^T (m_2 - m_1)(m_2 - m_1)^T \cdot w \\ &= w^T S_B w \end{aligned}$$

then $J(w) = \dfrac{(m_2 - m_1)^2}{s_1^2 + s_2^2} = \dfrac{w^T S_B w}{w^T S_W w}$

Q3. The maximum-likelihood function can be written as:

$$p(\{\Phi_n, t_n\} \mid \pi_1, \pi_2, \cdots, \pi_K) = \prod_{n=1}^{N}\prod_{k=1}^{K}\left[p(\Phi_n \mid C_k)\cdot p(C_k)\right]^{t_{nk}}$$

$$= \prod_{n=1}^{N}\prod_{k=1}^{K}\left[\pi_k \cdot p(\Phi_n \mid C_k)\right]^{t_{nk}}$$

$$\Rightarrow \ln p = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\left[\ln \pi_k + \ln p(\Phi_n \mid C_k)\right]$$

$$\propto \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\cdot \ln \pi_k$$

Using a Largrange multiplier, we have

$$L(\lambda, \pi_k) = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\cdot \ln \pi_k + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)$$

$$\Rightarrow \frac{\partial}{\partial \pi_k} L(\lambda, \pi_k) = \sum_{n=1}^{N}\frac{t_{nk}}{\pi_k} + \lambda$$

Let $\frac{\partial}{\partial \pi_k} L(\lambda, \pi_k) = 0$, we have

$$\sum_{n=1}^{N}\frac{t_{nk}}{\pi_k} + \lambda = 0 \Rightarrow \pi_k = -\frac{1}{\lambda}\sum_{n=1}^{N} t_{nk} = -\frac{N_k}{\lambda}$$

$$\Rightarrow \sum_{k=1}^{K}\pi_k = \sum_{k=1}^{K} -\frac{N_k}{\lambda}$$

$$\Rightarrow 1 = -\frac{N}{\lambda}$$

$$\Rightarrow \lambda = -N$$

Thus, we have $\pi_k = -\frac{N_k}{\lambda} = \frac{N_k}{N}$

Q4.

$$\sigma(a) = \frac{1}{1+e^{-a}}$$

$$\frac{d\sigma}{da} = -(1+e^{-a})^{-2}\cdot(-e^{-a})$$

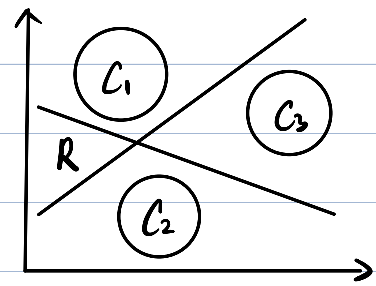
$$= \sigma^2(a)\cdot\left(\frac{1}{\sigma(a)} - 1\right)$$

$$= \sigma(a)\cdot(1-\sigma(a))$$

$$= \sigma(1-\sigma)$$

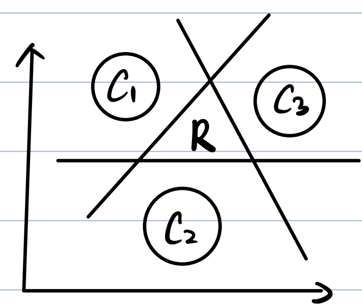
Q5.
$$\begin{aligned}\nabla E(w) &= -\nabla \sum_{n=1}^{N} \{t_n \ln y_n + (1-t_n)\cdot \ln(1-y_n)\} \\ &= -\sum_{n=1}^{N} \nabla \{t_n \ln y_n + (1-t_n)\cdot \ln(1-y_n)\} \\ &= -\sum_{n=1}^{N} \frac{\partial}{\partial y_n}\{t_n \ln y_n + (1-t_n)\cdot \ln(1-y_n)\}\cdot \frac{\partial y_n}{\partial a_n}\cdot\frac{\partial a_n}{\partial w} \\ &= -\sum_{n=1}^{N} \left(\frac{t_n}{y_n} - \frac{1-t_n}{1-y_n}\right)\cdot y_n(1-y_n)\cdot \phi_n \\ &= -\sum_{n=1}^{N} \frac{t_n - y_n}{y_n(1-y_n)}\cdot y_n(1-y_n)\cdot \phi_n \\ &= \sum_{n=1}^{N} (y_n - t_n)\cdot \phi_n\end{aligned}$$

Q6. ①



The region R is ambiguous.

②



The region R is ambiguous.

Q7. ① Proof: If convex hulls intersect, then not linearly separable

Let's write down the convex hull for both x and z

$$x = \sum_n \alpha_n \cdot x^n \qquad z = \sum_m \beta_m \cdot z^m$$

If intersect, we have $\begin{cases} y = \sum_n \alpha_n \cdot x^n \\ y = \sum_m \beta_m \cdot z^m \end{cases}$

Then $\hat{w}^T \cdot y + w_0 = w^T \cdot (\sum_n \alpha_n \cdot x^n) + (\sum_n \alpha_n) \cdot w_0$.

$$= \sum_n \alpha_n \cdot \hat{w}^T \cdot x^n + \sum_n \alpha_n \cdot w_0$$

$$= \sum_n \alpha_n (\hat{w}^T \cdot x^n + w_0)$$

Suppose $\{x^n\}$, $\{z^m\}$ are linearly separable, then

$\hat{w}^T \cdot x^n + w_0 > 0$ and $\hat{w}^T \cdot z^m + w_0 < 0$ for $\forall x^n, z^m$.

Since $\alpha_n \geq 0$, then $\hat{w}^T y + w_0 > 0$

We can also write $\hat{w}^T y + w_0$ as $\sum_m \beta_m (\hat{w}^T \cdot z^m + w_0)$

then we will get $\hat{w}^T y + w_0 < 0$

A contradiction occurs, which means if their convex hulls intersect, the two sets of points cannot be linearly separable

② Proof: If linearly separable, then convex hulls do not intersect.

Obviously, its contrapositive is:

If their convex hulls intersect, they are not linearly separable

which has been proved in part ①