Question 1

$$L(\mathbf{w}, \lambda) = \mathbf{w}^{\mathrm{T}}(\mathbf{m}_2 - \mathbf{m}_1) + \lambda(\mathbf{w}^{\mathrm{T}}\mathbf{w} - 1)$$

$$\frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, \lambda) = \mathbf{m}_2 - \mathbf{m}_1 + 2\lambda \mathbf{w} = 0$$

Hence,

$$\mathbf{w} = -rac{1}{2\lambda}(\mathbf{m}_2 - \mathbf{m}_1) \propto (\mathbf{m}_2 - \mathbf{m}_1)$$

Question 2

$$egin{aligned} J(oldsymbol{w}) &= rac{(oldsymbol{m}_2 - oldsymbol{m}_1)^2}{s_1^2 + s_2^2} \ &= rac{(oldsymbol{w}^T oldsymbol{m}_2 - oldsymbol{w}^T oldsymbol{m}_1)^2}{\sum_{n \in \mathcal{C}_1} (oldsymbol{w}^T oldsymbol{x}_1 - oldsymbol{w}^T oldsymbol{m}_1) + \sum_{n \in \mathcal{C}_2} (oldsymbol{w}^T oldsymbol{x}_2 - oldsymbol{w}^T oldsymbol{m}_2)} \ &= rac{oldsymbol{w}^T (oldsymbol{m}_1 - oldsymbol{m}_1) (oldsymbol{m}_2 - oldsymbol{m}_1)^T oldsymbol{w}}{oldsymbol{w}^T oldsymbol{S}_B oldsymbol{w}} \ &= rac{oldsymbol{w}^T oldsymbol{S}_B oldsymbol{w}}{oldsymbol{w}^T oldsymbol{S}_W oldsymbol{w}} \end{aligned}$$

Question 3

For a point ϕ_n from \mathcal{C}_k ,

$$p(\phi_n, \mathcal{C}_k) = p(\mathcal{C}_k) p(\phi_n | \mathcal{C}_k) = \pi_k p(\phi_n | \mathcal{C}_k)$$

so the likelihood function is

$$p(\{\phi_n, \mathbf{t}_n\} | oldsymbol{\pi}, oldsymbol{\mu}, \Sigma) = \prod_{n=1}^N \prod_{k=1}^K \left[\pi_k p(\phi_n | \mathcal{C}_k)
ight]^{t_{nk}}$$

Take the log-likelihood

$$\sum_{n=1}^{N}\sum_{k=1}^{K}[t_{nk}\ln\pi_k+t_{nk}\ln p(\phi_n|\mathcal{C}_k)]$$

As $\sum_{k=1}^K \pi_k = 1$, we use Lagrange Multiplier to maximize

$$egin{align} L(\pi_i,\lambda) &= \sum_{n=1}^N \sum_{k=1}^K [t_{nk} \ln \pi_k + t_{nk} \ln p(\phi_n | \mathcal{C}_k)] + \lambda (\sum_{k=1}^K \pi_k - 1) \ & rac{\partial}{\partial \lambda} L(\pi_i,\lambda) = \sum_{k=1}^K \pi_k - 1 = 0 \ & rac{\partial}{\partial \pi_i} L(\pi_i,\lambda) = rac{\sum_{n=1}^N t_{ni}}{\pi_i} + \lambda = 0 \ \end{cases}$$

$$\pi_i = -rac{\sum_{n=1}^N t_{ni}}{\lambda} = -rac{N_i}{\lambda}$$

Since

$$\sum_{k=1}^K \pi_k = \sum_{k=1}^K -\frac{N_i}{\lambda} = -\frac{N}{\lambda} = 1$$

then $\lambda = -N$

Hence,

$$\pi_i = rac{N_i}{N}$$

Question 4

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

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

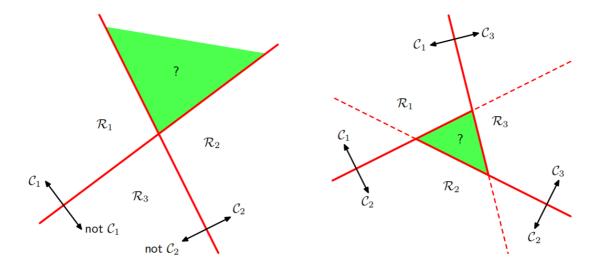
$$= \frac{e^{-a}}{1+e^{-a}} (1 - \frac{e^{-a}}{1+e^{-a}})$$

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

Question 5

$$egin{aligned} y_n &= \sigma(a_n) \ a_n &= oldsymbol{w}^T oldsymbol{\phi}_n \ rac{\partial y_n}{\partial oldsymbol{w}} &= rac{\partial y_n}{\partial oldsymbol{a}_n} rac{\partial a_n}{\partial oldsymbol{w}} = y_n (1 - y_n) \phi_n \
abla \mathbb{E}(oldsymbol{w}) &= -\sum_{n=1}^N \left(rac{t_n}{y_n} rac{\partial y_n}{\partial oldsymbol{w}} - rac{1 - t_n}{1 - y_n} rac{\partial y_n}{\partial oldsymbol{w}}
ight) \ &= -\sum_{n=1}^N \left(t_n (1 - y_n) - (1 - t_n) y_n
ight) \phi_n \ &= \sum_{n=1}^N (y_n - t_n) \phi_n \end{aligned}$$

Question 6



Question 7

Proof by contradiction

If their convex hulls intersect, then

$$\exists oldsymbol{y}, \quad s.\, t.\, oldsymbol{y} = \sum_n lpha_n oldsymbol{x}^n = \sum_m eta_n oldsymbol{z}^n$$

where lpha,eta>0 and $\sum_n lpha_n=\sum_m eta_n=1.$

Assume $\{\boldsymbol{x}^n\}$ and $\{\boldsymbol{z}^m\}$ are linearly separable, there exists a vector $\hat{\boldsymbol{w}}$ and a scalar w_0 such that $\hat{\boldsymbol{w}}^T\boldsymbol{x}^n+w_0>0$ for all \boldsymbol{x}^n , and $\hat{\boldsymbol{w}}\boldsymbol{z}^m+w_0<0$ for all \boldsymbol{z}^m .

We figure out

$$egin{aligned} \sum_n \hat{oldsymbol{w}}^T lpha_n oldsymbol{x}^n + \sum_n lpha_n w_0 &= \sum_n \hat{oldsymbol{w}}^T oldsymbol{y} + w_0 > 0 \ \sum_m \hat{oldsymbol{w}}^T eta_m oldsymbol{z}^m + \sum_m eta_m w_0 &= \sum_m \hat{oldsymbol{w}}^T oldsymbol{y} + w_0 < 0 \end{aligned}$$

It leads a contradiction.

Hence, if their convex hulls intersect, the two sets of points cannot be linearly separable, and conversely that, if they are linearly separable, their convex hulls do not intersect.