Machine Learning (H)
Southern University of Science and Technology
Mengxuan Wu
12212006

Assignment 4

Mengxuan Wu

Question 1

The function with Lagrange multiplier is

$$C(w, \lambda) = w^T(\mathbf{m_2} - \mathbf{m_1}) + \lambda(w^T w - 1)$$

Take the derivative with respect to w and set it to zero, we have

$$\frac{\partial C}{\partial w} = \mathbf{m_2} - \mathbf{m_1} + 2\lambda w = 0$$

Solve the equation, we have

$$w = \frac{\mathbf{m_1} - \mathbf{m_2}}{2\lambda}$$

Thus, we have $w \propto \mathbf{m_1} - \mathbf{m_2}$.

Question 2

$$\begin{aligned} J(w) &= \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2} \\ &= \frac{w^T(\mathbf{m_2} - \mathbf{m_1})(\mathbf{m_2} - \mathbf{m_1})^T w}{\sum_{n \in C_1}(w^T(\mathbf{x_n} - \mathbf{m_1}))^2 + \sum_{n \in C_2}(w^T(\mathbf{x_n} - \mathbf{m_2}))^2} \\ &= \frac{w^T(\mathbf{m_2} - \mathbf{m_1})(\mathbf{m_2} - \mathbf{m_1})^T w}{w^T\left(\sum_{n \in C_1}(\mathbf{x_n} - \mathbf{m_1})(\mathbf{x_n} - \mathbf{m_1})^T + \sum_{n \in C_2}(\mathbf{x_n} - \mathbf{m_2})(\mathbf{x_n} - \mathbf{m_2})^T\right) w} \end{aligned}$$

Since we know

$$S_B = (\mathbf{m_2} - \mathbf{m_1})(\mathbf{m_2} - \mathbf{m_1})^T$$

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

Thus, we have

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

Question 3

For each data point, we have

$$p(\{\phi, t\}) = \prod_{k=1}^{K} p(\phi|C_k)p(C_k) = \prod_{k=1}^{K} [\pi_k p(\phi|C_k)]^{t_k}$$

Thus, for the whole data set, we have

$$p(\{\phi_n, t_n\}) = \prod_{n=1}^{N} \prod_{k=1}^{K} [\pi_k p(\phi_n|C_k)]^{t_{nk}}$$

Take the log of the likelihood, we have

$$\log p(\{\phi_n, t_n\}) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \log \pi_k + t_{nk} \log p(\phi_n|C_k)$$

The extra constraint is

$$\sum_{k=1}^{K} \pi_k = 1$$

Thus, the Lagrange function is

$$L(\pi, \lambda) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \log \pi_k + t_{nk} \log p(\phi_n|C_k) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$$

Take the derivative with respect to π_k and set it to zero, we have

$$\frac{\partial L}{\partial \pi_k} = \sum_{n=1}^{N} \frac{t_{nk}}{\pi_k} + \lambda = 0$$

Solve the equation, we have

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

Since $\sum_{k=1}^{K} \pi_k = 1$, we have

$$\sum_{k=1}^{K} \pi_k = -\frac{1}{\lambda} \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} = 1$$

Thus, we have

$$\lambda = -N$$

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} t_{nk} = \frac{N_k}{N}$$

where N_k is the number of data points in class C_k.

Question 4

Taking the derivative of the sigmoid function, we have

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

Question 5

$$\frac{\partial E(w)}{\partial w} = \frac{\partial}{\partial w} - \sum_{n=1}^{N}[t_n \log y_n + (1-t_n)\log(1-y_n)]$$

Since $y_n = \sigma(w^T\phi_n)$, combining with the result in Question 4, we have

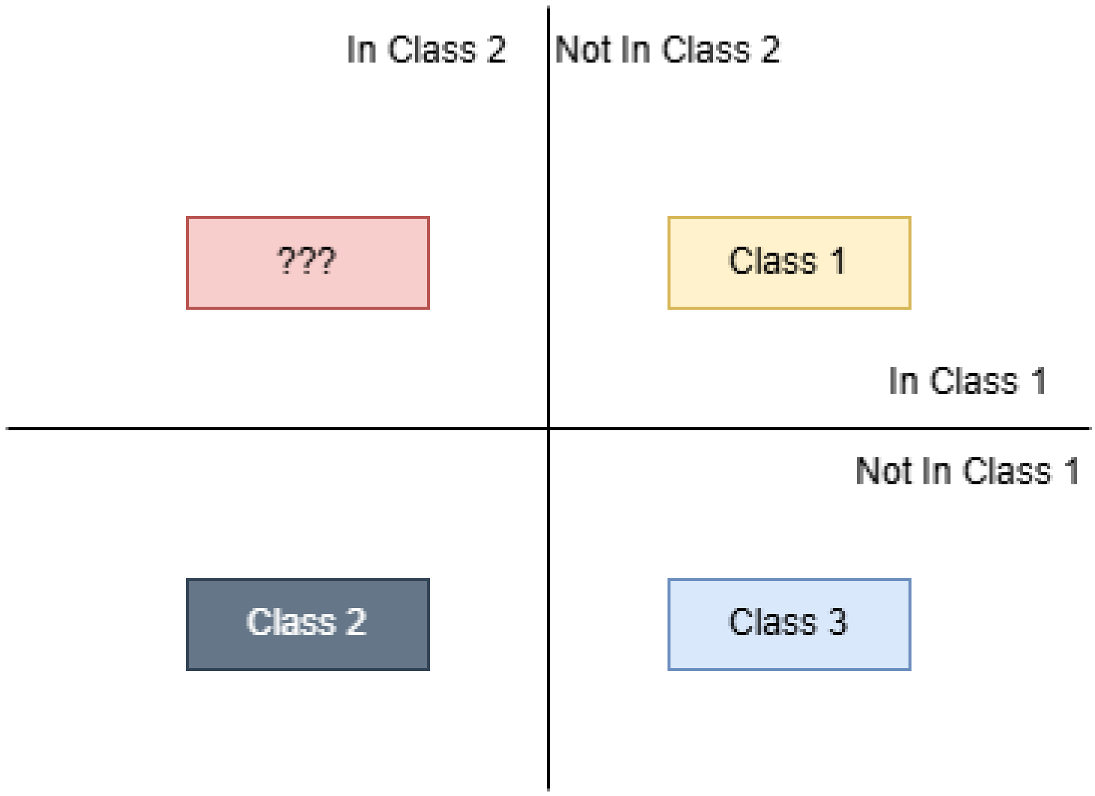
$$\begin{aligned}\frac{\partial y_n}{\partial w} &= \frac{\partial}{\partial w}\sigma(w^T\phi_n) \\ &= \sigma(w^T\phi_n)(1-\sigma(w^T\phi_n))\frac{\partial}{\partial w}w^T\phi_n \\ &= y_n(1-y_n)\phi_n\end{aligned}$$

Thus, we have

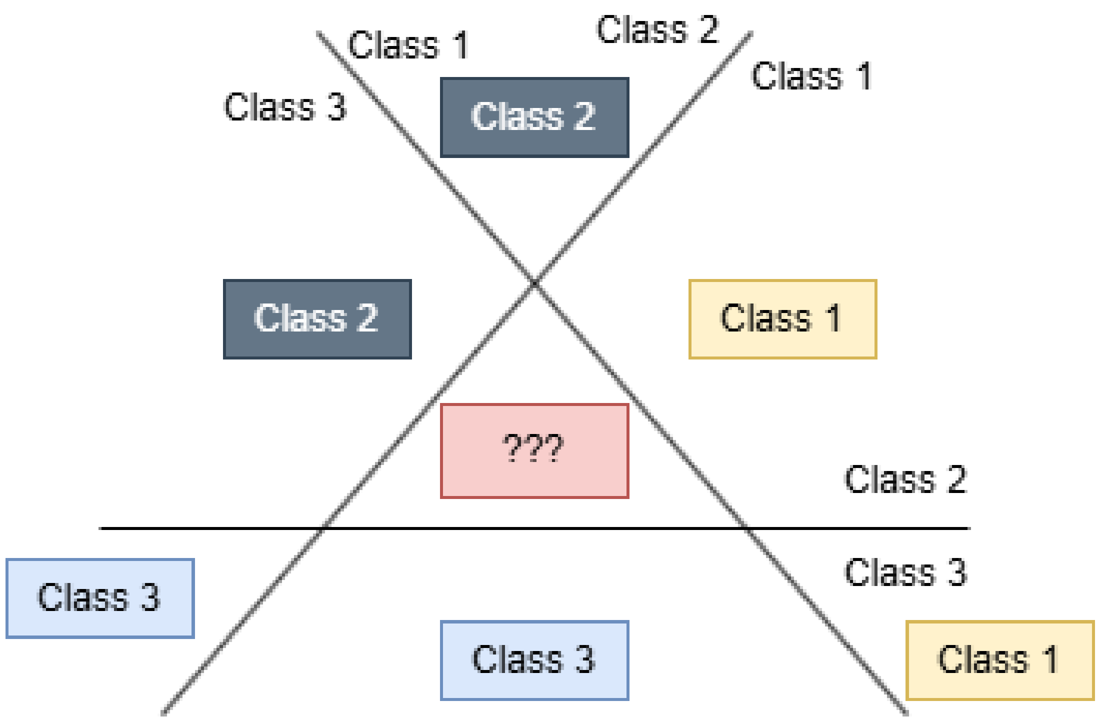
$$\begin{aligned}\frac{\partial E(w)}{\partial w} &= -\sum_{n=1}^{N}\left[t_n\frac{1}{y_n}y_n(1-y_n)\phi_n - (1-t_n)\frac{1}{1-y_n}y_n(1-y_n)\phi_n\right] \\ &= -\sum_{n=1}^{N}[t_n(1-y_n)\phi_n - (1-t_n)y_n\phi_n] \\ &= -\sum_{n=1}^{N}[t_n - y_n]\,\phi_n \\ &= \sum_{n=1}^{N}[y_n - t_n]\,\phi_n\end{aligned}$$

Question 6

For the first approach, we have



For the second approach, we have



Question 7

Convex hulls intersect ⇒ Sets are not linearly separable

If two convex hulls intersect, then there must be at least a point that is in both convex hulls. Let the point be $\mathbf{x}$, then we have

$$\mathbf{x} = \sum_{n=1}^{N} \alpha_n \mathbf{x}^n = \sum_{m=1}^{M} \beta_m \mathbf{z}^m$$

where $\alpha_n \geq 0$, $\sum_{n=1}^{N} \alpha_n = 1$, $\beta_m \geq 0$, $\sum_{m=1}^{M} \beta_m = 1$.
Suppose there exists a vector $\hat{w}$ and a scalar w_0 that can linearly separate the two convex hulls, then we have

$$\hat{w}^T \mathbf{x}^n + w_0 > 0 \quad \forall n$$
$$\hat{w}^T \mathbf{z}^m + w_0 < 0 \quad \forall m$$

Thus, we have

$$\hat{w}^T \mathbf{x} = \hat{w}^T \sum_{n=1}^{N} \alpha_n \mathbf{x}^n = \sum_{n=1}^{N} \alpha_n \hat{w}^T \mathbf{x}^n > 0$$
$$\hat{w}^T \mathbf{x} = \hat{w}^T \sum_{m=1}^{M} \beta_m \mathbf{z}^m = \sum_{m=1}^{M} \beta_m \hat{w}^T \mathbf{z}^m < 0$$

There is a contradiction, thus the two convex hulls cannot be linearly separated.

Sets are linearly separable ⇒ Convex hulls do not intersect

This is the contrapositive of the first part, thus it has the same truth value, which is true.