Assignment III: Linear Models for Classification

Select the equations that fit in the blanks in the question text. The selected answers ((A)–(D)) must be submitted via the link to the assignment that appears in the “General” channel of “機械学習 2024 KA240201-teams.”

Questions 1 and 2

Consider a generative classification model for K classes defined by prior class probabilities $p(\mathcal{C}_k) = \pi_k$ and general class-conditional densities $p(\boldsymbol{\phi}|\mathcal{C}_k)$ where $\boldsymbol{\phi}$ is the input feature vector. Suppose we are given a training data set $\{\boldsymbol{\phi}_n, \mathbf{t}_n\}$, where $n = 1, \ldots, N$, and $\mathbf{t}_n$ is a binary target vector of length K that uses the 1-of-K coding scheme, so that it has components $t_{nj} = I_{jk}$ if pattern n is from class $\mathcal{C}_k$. Assuming that the data points are drawn independently from this model, we show that the maximum-likelihood solution for the prior probabilities is given by

$$\pi_k = \frac{N_k}{N}, \tag{a}$$

where N_k is the number of data points assigned to class $\mathcal{C}_k$.

The likelihood function is given by

$$p(\{\boldsymbol{\phi}_n, \mathbf{t}_n\}|\{\pi_k\}) = \prod_{n=1}^{N}\prod_{k=1}^{K}\{\boxed{\quad(1)\quad}\pi_k\}^{t_{nk}}$$

and taking the logarithm, we obtain

$$\ln p(\{\boldsymbol{\phi}_n, \mathbf{t}_n\}|\{\pi_k\}) = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\{\ln\boxed{\quad(1)\quad} + \ln \pi_k\}.$$

In order to maximize the log likelihood with respect to π_k, we need to preserve the constraint $\sum_{k=1}^{K}\pi_k = 1$. This can be done by introducing a Lagrange multiplier λ and maximizing

$$\ln p(\{\boldsymbol{\phi}_n, \mathbf{t}_n\}|\{\pi_k\}) + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right).$$

Setting the derivative with respect to π_k equal to zero, we obtain

$$\frac{1}{\pi_k}\boxed{\quad(2)\quad} + \lambda = 0.$$

Re-arranging then gives

$$-\pi_k \lambda = \boxed{\quad (2) \quad} = N_k.$$

Summing both sides over k, we find that $\lambda = -N$, and using this to eliminate λ, we obtain Eq. (a).

Question 1. Select the equation that fills in the blank (1).

(A) $p(\boldsymbol{\phi}_n)$

(B) $p(\mathcal{C}_k)$

(C) $p(\boldsymbol{\phi}_n, \mathcal{C}_k)$

(D) $p(\boldsymbol{\phi}_n | \mathcal{C}_k)$

Question 2. Select the equation that fills in the blank (2).

(A) t_{nk}

(B) $\sum_{n=1}^{N} t_{nk}$

(C) $\sum_{k=1}^{K} t_{nk}$

(D) $\sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk}$

Questions 3 and 4

We show that the derivatives of the following softmax function

$$p(\mathcal{C}_k|\boldsymbol{\phi}) = y_k(\boldsymbol{\phi}) = \frac{e^{a_k}}{\sum_j e^{a_j}}, \tag{b}$$

where the a_k are defined by $a_k = \mathbf{w}_k^{\mathrm{T}} \boldsymbol{\phi}$, are given by

$$\frac{\partial y_k}{\partial a_j} = y_k (I_{kj} - y_j), \tag{c}$$

where I_{kj} are the elements of the identity matrix.

Using Eq. (b), we have

$$\frac{\partial y_k}{\partial a_k} = \frac{\boxed{\quad (3) \quad}}{(\sum_i e^{a_i})^2} = y_k(1 - y_k),$$

$$\frac{\partial y_k}{\partial a_j} = -\frac{\boxed{\quad (4) \quad}}{(\sum_i e^{a_i})^2} = -y_k y_j, \qquad j \neq k.$$

Combining these results, we obtain Eq. (c).

Question 3. Select the equation that fills in the blank (3).

(A) $(e^{a_k} \sum_i e^{a_i}) - e^{a_k}$

(B) $(e^{a_k} \sum_i e^{a_i}) - e^{2a_k}$

(C) $(e^{2a_k} \sum_i e^{a_i}) - e^{a_k}$

(D) $(e^{2a_k} \sum_i e^{a_i}) - e^{2a_k}$

Question 4. Select the equation that fills in the blank (4).

(A) $e^{a_k} e^{a_j}$

(B) $e^{2a_k} e^{a_j}$

(C) $e^{a_k} e^{2a_j}$

(D) $e^{2a_k} e^{2a_j}$