Assignment II: Linear Models for Regression

Select the equations or calculate the values that fit in the blanks in the question text. The selected answers ((A)–(D)) or the calculated values 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

We show that a general linear combination of logistic sigmoid functions of the form

$$y(x, \mathbf{w}) = w_0 + \sum_{j=1}^{M} w_j \sigma\left(\frac{x - \mu_j}{s}\right) \tag{a}$$

is equivalent to a linear combination of 'tanh' functions of the form

$$y(x, \mathbf{u}) = u_0 + \sum_{j=1}^{M} u_j \tanh\left(\frac{x - \mu_j}{s}\right)$$

and find expressions to relate the new parameters $u_1, \ldots, u_M$ to the original parameters $w_1, \ldots, w_M$.

If we now take $a_j = (x - \mu_j)/2s$, we can rewrite Eq. (a) as

$$\begin{aligned} y(x, \mathbf{w}) &= w_0 + \sum_{j=1}^{M} \frac{w_j}{2} (\boxed{\ (1)\ } + 1) \\ &= u_0 + \sum_{j=1}^{M} u_j \tanh(a_j) \end{aligned}$$

where $u_j = w_j/2$ for $j = 1, \ldots, M$, and $u_0 = \boxed{\ (2)\ }$.

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

(A) $\sigma(a_j) - 1$

(B) $\sigma(2a_j) - 1$

(C) $2\sigma(a_j) - 1$

(D) $2\sigma(2a_j) - 1$

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

(A) $w_0/2$

(B) w_0

(C) $w_0 + \sum_{j=1}^{M} w_j$

(D) $w_0 + \sum_{j=1}^{M} w_j/2$

Questions 3 and 4

Consider a linear basis function regression model for a multivariate target variable $\mathbf{t}$ having a Gaussian distribution of the form

$$p(\mathbf{t}|\mathbf{W}, \boldsymbol{\Sigma}) = \mathcal{N}(\mathbf{t}|\mathbf{y}(\mathbf{x}, \mathbf{W}), \boldsymbol{\Sigma})$$

where

$$\mathbf{y}(\mathbf{x}, \mathbf{W}) = \mathbf{W}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x})$$

together with a training data set comprising input basis vectors $\boldsymbol{\phi}(\mathbf{x}_n)$ and corresponding target vector $\mathbf{t}_n$, with $n = 1, \ldots, N$. Here we derive the maximum likelihood solution $\mathbf{W}_{\mathrm{ML}}$ for the parameter matrix $\mathbf{W}$.

We first write down the log likelihood function which is given by

$$\ln L(\mathbf{W}, \boldsymbol{\Sigma}) = -\frac{N}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \left(\mathbf{t}_n - \mathbf{W}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n)\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\mathbf{t}_n - \mathbf{W}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n)\right)$$

First of all, we set the derivative with respect to $\mathbf{W}$ equal to zero, giving

$$0 = \boxed{\quad (3) \quad}.$$

Multiplying through by $\boldsymbol{\Sigma}$ and introducing the design matrix $\boldsymbol{\Phi}$ and the target data matrix $\mathbf{T}$, we have

$$\boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi} \mathbf{W} = \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{T}.$$

Solving for $\mathbf{W}$ then gives

$$\mathbf{W}_{\mathrm{ML}} = \boxed{\quad (4) \quad}.$$

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

(A) $-\sum_{n=1}^{N} \boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}(\mathbf{x}_n)(\mathbf{t}_n - \mathbf{W}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n))$

(B) $-\sum_{n=1}^{N} \boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}}(\mathbf{t}_n - \mathbf{W}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n))$

(C) $-\sum_{n=1}^{N} \boldsymbol{\Sigma}^{-1} (\mathbf{t}_n - \mathbf{W}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n))\boldsymbol{\phi}(\mathbf{x}_n)$

(D) $-\sum_{n=1}^{N} \boldsymbol{\Sigma}^{-1} (\mathbf{t}_n - \mathbf{W}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n))\boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}}$

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

(A) $\left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{T}$

(B) $\mathbf{\Phi}^{\mathrm{T}}\mathbf{T}\left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1}$

(C) $\left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$

(D) $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{T}\right)^{-1}$

Question 5

Suppose that we have a training data set, in which the input data set is $\{1, 2, 3, 4, 5\}$ and their corresponding target data set is $\{0.80, 2.45, 4.00, 6.03, 7.83\}$. To a linear regression model $t = w_1 x + w_0 + \epsilon$, where ϵ is the noise following a zero-mean Gaussian distribution, we can calculate the maximum likelihood solutions of w_1 and w_0 using the above training data set. If we use this model and these maximum likelihood solutions, the predicted target value $\hat{t}$ for a new input $x_{\text{new}} = 2.5$ is $\boxed{\;(5)\;}$.

Question 5. Calculate the value that fills in the blank (5).