Решение пробного теста по дисциплине «Методы машинного обучения» для направлений НКН, НФИ, НПИ, НБИ

15 июня 2022 г.

Задание 1 – теория (4 балла). Ответить на теоретический вопрос.

Задание 2 – регрессия (4 балла). Дан набор данных с двумя числовыми признаками (независимой переменной $\mathbf{X}$ и зависимой переменной $\mathbf{Y}$)

$$\mathbf{D} = \{(1., 2.), (3., 4.), (5., 5.), (7., 5.)\}$$

Найти параметры a и b парной линейной регрессии

$$y(x) = a\,x + b + \varepsilon$$

Решение: Параметры парной линейной регрессии вычисляются по формулам

$$a = \frac{\sigma_{\mathbf{XY}}}{\sigma^2_{\mathbf{X}}},\; b = \mu_{\mathbf{Y}} - a\,\mu_{\mathbf{X}},$$

где

- $\mu_{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\mu_{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^{n} y_i$ – средние значения
- $\sigma^2_{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu_{\mathbf{X}})^2$ – дисперсия независимой переменной $\mathbf{X}$
- $\sigma_{\mathbf{XY}} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu_{\mathbf{X}})(y_i - \mu_{\mathbf{Y}})$ – ковариация $\mathbf{X}$и $\mathbf{Y}$

Поэтому

$$\mu_{\mathbf{X}} = \frac{1}{4}(1+3+5+7) = 4,\; \mu_{\mathbf{Y}} = \frac{1}{4}(2+4+5+5) = 4,$$

$$\sum_{i=1}^{4} (x_i - \mu_{\mathbf{X}})(y_i - \mu_{\mathbf{Y}}) = (1-4)(2-4) + (3-4)(4-4) + (5-4)(5-4) + (7-4)(5-4) =$$

$$= (-3)(-2) + 0 + 1 + 3 \cdot 1 = 10,$$

$$\sum_{i=1}^{4} (x_i - \mu_{\mathbf{X}})^2 = (1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)^2 = 9+1+1+9 = 20,$$

$$a = \frac{\sum_{i=1}^{4} (x_i - \mu_{\mathbf{X}})(y_i - \mu_{\mathbf{Y}})}{\sum_{i=1}^{4} (x_i - \mu_{\mathbf{X}})^2} = \frac{10}{20} = 0.5,$$

$$b = \mu_{\mathbf{Y}} - a\,\mu_{\mathbf{X}} = 4 - 0.5 \cdot 4 = 2$$

Задание 3 – обратное дифференцирование (4 балла). Вычислить градиент функции $f(x_1, x_2) = e^{x_1 x_2} + x_2^3$ по переменным x_1, x_2 при помощи алгоритма обратного дифференцирования.

Решение: Функция $f\left(x_1, x_2\right)$ может быть вычислена при помощи следующей последовательности элементарных операций и функций:

$$w_1 = x_1,\ w_2 = x_2 \Rightarrow w_3 = w_1 w_2,\ w_4 = w_2^3 \Rightarrow w_5 = e^{w_3} \Rightarrow w_6 = w_4 + w_5$$

или

$$y = f\left(x_1, x_2\right) = e^{x_1 x_2} + x_2^3 = e^{w_1 w_2} + w_2^3 = e^{w_3} + w_4 = w_5 + w_4 = w_6$$

Таким образом, имеют место следующие прямые зависимости между переменными w_i:

- w_3 непосредственно зависит от w_1 и w_2
- w_4 непосредственно зависит от w_2
- w_5 непосредственно зависит от w_3
- w_6 непосредственно зависит от w_4 и w_5

Иначе, одни переменные w_i непосредственно влияют на другие переменные w_j следующим образом:

- переменная w_5 непосредственно влияет на w_6
- переменная w_4 непосредственно влияет на w_6
- переменная w_3 непосредственно влияет на w_5
- переменная w_2 непосредственно влияет на w_3 и w_4
- переменная w_1 непосредственно влияет на w_3

При обратном дифференцировании в цепном правиле

$$\frac{\partial y}{\partial w_j} = \sum_i \frac{\partial w_i}{\partial w_j} \frac{\partial y}{\partial w_i} = \frac{\partial w_1}{\partial w_j} \frac{\partial y}{\partial w_1} + \frac{\partial w_2}{\partial w_j} \frac{\partial y}{\partial w_2} + ...$$

остаются слагаемые только для тех переменных w_i, которые непосредственно зависят от переменной w_j, т.е. когда $\frac{\partial w_i}{\partial w_j} \neq 0$. Поэтому получаем

$$\overline{w}_6 = \frac{\partial y}{\partial w_6} = [y \equiv w_6] = \frac{\partial y}{\partial y} = 1$$

$$\overline{w}_5 = \frac{\partial y}{\partial w_5} = \frac{\partial w_6}{\partial w_5} \frac{\partial y}{\partial w_6} = 1 \cdot \frac{\partial y}{\partial w_6} = \overline{w}_6$$

$$\overline{w}_4 = \frac{\partial y}{\partial w_4} = \frac{\partial w_6}{\partial w_4} \frac{\partial y}{\partial w_6} = 1 \cdot \frac{\partial y}{\partial w_6} = \overline{w}_6$$

$$\overline{w}_3 = \frac{\partial y}{\partial w_3} = \frac{\partial w_5}{\partial w_3} \frac{\partial y}{\partial w_5} = e^{w_3} \cdot \frac{\partial y}{\partial w_5} = e^{w_3} \overline{w}_5$$

$$\overline{w}_2 = \frac{\partial y}{\partial w_2} = \frac{\partial w_3}{\partial w_2} \frac{\partial y}{\partial w_3} + \frac{\partial w_4}{\partial w_2} \frac{\partial y}{\partial w_4} = w_1 \frac{\partial y}{\partial w_3} + 3 w_2^2 \frac{\partial y}{\partial w_4} = w_1 \overline{w}_3 + 3 w_2^2 \overline{w}_4$$

$$\overline{w}_1 = \frac{\partial y}{\partial w_1} = \frac{\partial w_3}{\partial w_1} \frac{\partial y}{\partial w_3} = w_2 \frac{\partial y}{\partial w_3} = w_2 \overline{w}_3$$

Итак, градент функции $f\left(x_1, x_2\right)$, вычисленный при помощи обратного дифференцирования, равен

$$\frac{\partial y}{\partial x_1} = \overline{w}_1 = w_2 \overline{w}_3 = x_2 e^{x_1 x_2},\ \frac{\partial y}{\partial x_2} = \overline{w}_2 = x_1 e^{x_1 x_2} + 3 x_2^2$$

Задание 4 – сети прямого распространения (4 балла). Дана сеть MLP с двумя входными нейронами, скрытым слоем с тремя нейронами с функцией активации ReLU и одним нейроном в выходном слое без активации (с тождественной функцией активации). Матрица весов и вектор смещений для скрытого слоя равны

$$\mathbf{W}_h = \begin{pmatrix} 1 & -2 & -1 \\ 1 & 1 & -2 \end{pmatrix}, \, \mathbf{b}_h = \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix},$$

вектор весов и смещение для выходного слоя равны

$$\mathbf{W}_o = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix}, \, \mathbf{b}_o = -1$$

Вычислить выход нейронной сети MLP для входных данных $\mathbf{x} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

Решение: Чистые входы нейронов скрытого слоя равны

$$\mathbf{W}_h^T \mathbf{x} + \mathbf{b}_h = \begin{pmatrix} 1 & 1 \\ -2 & 1 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} + \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -4 \\ 3 \end{pmatrix} + \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -6 \\ 2 \end{pmatrix}$$

Выходы нейронов скрытого слоя равны

$$\mathbf{z} = \mathrm{ReLU}\left(\begin{pmatrix} 1 \\ -6 \\ 2 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

Тогда выход нейронной сети равен

$$\mathbf{o} = \mathbf{W}_o^T \mathbf{z} + \mathbf{b}_o = \begin{pmatrix} 1 & 3 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} - 1 = -1 - 1 = -2$$

Задание 5 – сверточные сети (4 балла). Дан фрагмент (два слоя) сверточной нейронной сети CNN, состоящий из слоя свертки с фильтром $\mathbf{W}$, смещением b, функцией активации ReLU и шагом 1 и слоем макс-пулинга с размером окна 2×2 и шагом 1. Фильтр и смещение для слоя свертки равны

$$\mathbf{W} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \, b = 1$$

Вычислить выход слоя макс-пулинга для входных данных слоя свертки

$$\mathbf{X} = \begin{pmatrix} 1 & 2 & 2 & 1 & 1 \\ 3 & 1 & 4 & 2 & 1 \\ 2 & 1 & 3 & 4 & 3 \\ 1 & 2 & 3 & 1 & 1 \\ 4 & 1 & 3 & 2 & 1 \end{pmatrix}$$

.

Решение: Свертка $\mathbf{X}$ и $\mathbf{W}$ с шагом 1 дает матрицу

$$\mathbf{X} * \mathbf{W} = \begin{pmatrix} -3 & 1 & -1 \\ -1 & -3 & 4 \\ 0 & 0 & 1 \end{pmatrix}$$

Чистые входы нейронов слоя свертки равны

$$\mathbf{X} * \mathbf{W} + b = \begin{pmatrix} -3 & 1 & -1 \\ -1 & -3 & 4 \\ 0 & 0 & 1 \end{pmatrix} + 1 = \begin{pmatrix} -2 & 2 & 0 \\ 0 & -2 & 5 \\ 1 & 1 & 2 \end{pmatrix}$$

Выходы нейронов слоя свертки равны

$$\mathbf{z}_c = \mathrm{ReLU}\left(\begin{pmatrix} -2 & 2 & 0 \\ 0 & -2 & 5 \\ 1 & 1 & 2 \end{pmatrix}\right) = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 5 \\ 1 & 1 & 2 \end{pmatrix}$$

Пулы 2×2 с шагом 1 для матрицы выходов нейронов слоя свертки равны

$$\begin{pmatrix} \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} & \begin{pmatrix} 0 & 5 \\ 1 & 2 \end{pmatrix} \end{pmatrix},$$

поэтому выход слоя макс-пулинга равен

$$\mathbf{z}_p = \begin{pmatrix} 2 & 5 \\ 1 & 5 \end{pmatrix}$$