人工智能基础 第五次作业

2.

引理: (误差向量 $(\vec{y}-\hat{y})$ 垂直于X的列空间)

最小二乘,即解超定方程: $XW = \vec{y}$, 预测向量 $\hat{y} = XW$ 。其中

$$X = egin{pmatrix} x_1 & 1 \ x_2 & 1 \ dots & dots \ x_n & 1 \end{pmatrix} = ((x_1, \cdots, x_n)^T \quad (1, \dots, 1)^T), \ \ W = egin{pmatrix} w \ b \end{pmatrix}, \ \ ec{y} = egin{pmatrix} y_1 \ y_2 \ dots \ y_n \end{pmatrix} = (y_1, \cdots, y_n)^T$$

由 $rac{\partial (ec{y}-XW)}{\partial W}=0$,可知 $W=(X^TX)^{-1}X^Tec{y}$ (或将 $ec{y}$ 分解为在X列空间的投影以及垂直分量)。因此

$$X^T(\vec{y} - \hat{y}) = X^T(\vec{y} - XW) = X^T(\vec{y} - X(X^TX)^{-1}X^T\vec{y}) = X^T\vec{y} - X^T\vec{y} = \vec{0}$$

可知 $(\vec{y} - \hat{y})$ 垂直于X的所有列向量,有

$$\hat{y}^T(\vec{y} - \hat{y}) = W^T X^T(\vec{y} - \hat{y}) = W^T \vec{0} = 0$$

$$((1, 1, \dots, 1)^T)^T (\vec{y} - \hat{y}) = (1, 1, \dots, 1)(\vec{y} - \hat{y}) = 0 \quad (\text{残差均值为0})$$

定理: (方差平方和分解)

由于 $(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$, 有

$$egin{split} \sum_{i=1}^n (y_i - ar{y})^2 &= \sum_{i=1}^n (\hat{y}_i - ar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \ &\Leftrightarrow \sum_{i=1}^n (\hat{y}_i - ar{y})(y_i - \hat{y}_i) = (\hat{y} - ar{y}(1, 1, \dots, 1)^T)^T (ar{y} - \hat{y}) = 0 \end{split}$$

由引理: $(\hat{y} - \bar{y}(1, \dots, 1)^T)^T(\vec{y} - \hat{y}) = \hat{y}^T(\vec{y} - \hat{y}) + \bar{y}(1, \dots, 1)(\vec{y} - \hat{y}) = 0 + 0 = 0$, 得证。

4.

由输入计算输出预测向量:

$$ec{\pi}(x) = (\pi_1(x), \pi_2(x), \dots, \pi_K(x)) = (rac{e^{\omega_1 x + b_1}}{\sum_{j=1}^K e^{\omega_j x + b_j}}, rac{e^{\omega_2 x + b_2}}{\sum_{j=1}^K e^{\omega_j x + b_j}}, \dots, rac{e^{\omega_K x + b_K}}{\sum_{j=1}^K e^{\omega_j x + b_j}})$$

由输出预测向量计算回传梯度:

第一层梯度:

$$egin{aligned} J &= -\sum_{j=1}^K ec{y}_j \log\left(\pi_j(x)
ight) = -\log\left(\pi_y(x)
ight) \ & rac{\partial l}{\partial \pi_j} = rac{\partial J(\pi_1, \dots, \pi_j, \dots, \pi_K)}{\partial \pi_j} = egin{cases} -rac{1}{\pi_y} & , j = y \ 0 & , j
eq y \end{cases} \end{aligned}$$

第二层梯度:

$$egin{aligned} \pi_y(x) &= e^{z_y} / \sum_{j=1}^K e^{z_j} \ rac{\partial \pi_y(x)}{\partial z_j} &= egin{cases} \pi_y - \pi_y^2 &, j = y \ -\pi_j \pi_y &, j
eq y \end{cases} \end{aligned}$$

以及最底层梯度: $z_j=\omega_j x+b_j, \; rac{dz_j}{d\omega_j}=x, \; rac{dz_j}{db_j}=1$

由反向传播, 最终有

$$egin{aligned} rac{\partial l}{\partial \omega_j} = & egin{cases} (\pi_y - 1)x &, j = y \ \pi_j x &, j
eq y \ rac{\partial l}{\partial b_j} = & egin{cases} (\pi_y - 1) &, j = y \ \pi_j &, j
eq y \end{cases} \end{aligned}$$

或者用向量表示: (设 \vec{y} 为输入(x,y)对应的独热码)

$$rac{\partial l}{\partial ec{\omega}} = x ec{\pi}(x) - x ec{y}$$
 $rac{\partial l}{\partial ec{b}} = ec{\pi}(x) - ec{y}$

3.

由题意
$$\log\left(P(Y=k)\right)=ec{eta}_kec{x}-\log Z,\ P(Y=k)=e^{ec{eta}_kec{x}}/Z$$

其中为了引入偏置, $ec{x}$ 应在末尾补1,相当于 $ec{x}=(x_1,x_2,\dots,x_n,1)$ 。
由 $\sum_{j=1}^K P(Y=k)=(\sum_{k=1}^K e^{ec{eta}_kec{x}})/Z=1$,有
$$Z=\sum_{i=1}^n P(Y=j)=\sum_{i=1}^n e^{ec{eta}_jec{x}}$$

$$P(Y=k)=e^{ec{eta}_kec{x}_k}/Z=rac{e^{ec{eta}_kec{x}}}{\sum_{i=1}^ne^{ec{eta}_jec{x}}}$$