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

$$X=\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}=((x_1,\cdots,x_n)^T \quad (1,\ldots,1)^T),\ \ W=\begin{pmatrix} w \\ b \end{pmatrix},\ \ \vec{y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}=(y_1,\cdots,y_n)^T$$

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

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

可知$(\vec{y}-\hat{y})$垂直于X的所有列向量，有

$$\hat{y}^T(\vec{y}-\hat{y})=W^TX^T(\vec{y}-\hat{y})=W^T\vec{0}=0$$
$$\big((1,1,\ldots,1)^T\big)^T(\vec{y}-\hat{y})=(1,1,\ldots,1)(\vec{y}-\hat{y})=0\ \ (\text{残差均值为}0)$$

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

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$
$$\Leftrightarrow\sum_{i=1}^{n}(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)=(\hat{y}-\bar{y}(1,1,\ldots,1)^T)^T(\vec{y}-\hat{y})=0$$

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

4.

由输入计算输出预测向量：

$$\vec{\pi}(x)=(\pi_1(x),\pi_2(x),\ldots,\pi_K(x))=(\frac{e^{\omega_1x+b_1}}{\sum_{j=1}^{K}e^{\omega_jx+b_j}},\frac{e^{\omega_2x+b_2}}{\sum_{j=1}^{K}e^{\omega_jx+b_j}},\ldots,\frac{e^{\omega_Kx+b_K}}{\sum_{j=1}^{K}e^{\omega_jx+b_j}})$$

由输出预测向量计算回传梯度：

第一层梯度：

$$J=-\sum_{j=1}^{K}\vec{y}_j\log\left(\pi_j(x)\right)=-\log\left(\pi_y(x)\right)$$
$$\frac{\partial l}{\partial \pi_j}=\frac{\partial J(\pi_1,\ldots,\pi_j,..,\pi_K)}{\partial \pi_j}=\begin{cases}-\frac{1}{\pi_y} & ,j=y\\ 0 & ,j\neq y\end{cases}$$

第二层梯度：

$$\pi_y(x) = e^{z_y} / \sum_{j=1}^{K} e^{z_j}$$

$$\frac{\partial \pi_y(x)}{\partial z_j} = \begin{cases} \pi_y - \pi_y^2 & ,j = y \\ -\pi_j \pi_y & ,j \neq y \end{cases}$$

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

由反向传播，最终有

$$\frac{\partial l}{\partial \omega_j} = \begin{cases} (\pi_y - 1)x & ,j = y \\ \pi_j x & ,j \neq y \end{cases}$$

$$\frac{\partial l}{\partial b_j} = \begin{cases} (\pi_y - 1) & ,j = y \\ \pi_j & ,j \neq y \end{cases}$$

或者用向量表示：（设$\vec{y}$为输入(x, y)对应的独热码）

$$\frac{\partial l}{\partial \vec{\omega}} = x\vec{\pi}(x) - x\vec{y}$$

$$\frac{\partial l}{\partial \vec{b}} = \vec{\pi}(x) - \vec{y}$$

3.

由题意$\log(P(Y = k)) = \vec{\beta}_k \vec{x} - \log Z$，$P(Y = k) = e^{\vec{\beta}_k \vec{x}} / Z$

其中为了引入偏置，$\vec{x}$应在末尾补1，相当于$\vec{x} = (x_1, x_2, .., x_n, 1)$。

由$\sum_{j=1}^{K} P(Y = k) = (\sum_{k=1}^{K} e^{\vec{\beta}_k \vec{x}}) / Z = 1$，有

$$Z = \sum_{j=1}^{n} P(Y = j) = \sum_{j=1}^{n} e^{\vec{\beta}_j \vec{x}}$$

$$P(Y = k) = e^{\vec{\beta}_k \vec{x}_k} / Z = \frac{e^{\vec{\beta}_k \vec{x}}}{\sum_{j=1}^{n} e^{\vec{\beta}_j \vec{x}}}$$