人工智能基础

Homework 5

1. 一元线性回归的平方和分解公式

证明:

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i-\bar{y}) &= \sum_{i=1}^{n}((y_i-\hat{y}_i)+(\hat{y}_i-\bar{y}))^2 \\
&= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+2\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})
\end{aligned} \tag{1}
$$

其中，利用最小二乘法的定义，$\hat{y}_i=\hat{\omega}x+\hat{b}$，平方误差和最小，则:

$$
\begin{aligned}
|L_2| &= \sum_{i=1}^{n}(y_i-(\hat{\omega}x_i+\hat{b}))^2 \\
\frac{\partial|L_2|}{\partial\hat{b}} &= -2\sum_{i=1}^{n}(y_i-\hat{y}_i)=0 \\
\frac{\partial|L_2|}{\partial\hat{\omega}} &= -2\sum_{i=1}^{n}x_i(y_i-\hat{y}_i)=0
\end{aligned} \tag{2}
$$

将（2）代入（1），即得:

$$
\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})=\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{\omega}x_i+\hat{b}-\bar{y})=0 \tag{3}
$$

故:

$$
\sum_{i=1}^{n}(y_i-\bar{y})=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 \tag{4}
$$

证毕。

3. 正交最小二乘

(a)

不适用，普通最小二乘法只适用于仅考虑因变量y存在误差的情况。这时候，需要用正交最小二乘，同时考虑x和y的观测误差。

(b)

很显然，每个点$\boldsymbol{x}$（这里为坐标向量而非横坐标）到直线的距离为:

$$
d=|(\boldsymbol{x}-\boldsymbol{x_0})\cdot\boldsymbol{n}| \tag{5}
$$

因此平方误差损失即为:

$$|L_2| = \sum_{i=1}^{n}[(\boldsymbol{x_i} - \boldsymbol{x_0}) \cdot \boldsymbol{n}]^2 \tag{6}$$

(c)

当平方误差取到最小时，有：

$$\frac{\partial |L_2|}{\partial \boldsymbol{x_0}} = -2\boldsymbol{n}\sum_{i=1}^{n}(\boldsymbol{x_i} - \boldsymbol{x_0}) \cdot \boldsymbol{n} = 0 \tag{7}$$

显然，有一个可行的取值为：

$$\boldsymbol{x_0} = \bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x_i}) \tag{8}$$

此时，平方误差损失可以化为二次型：

$$|L_2| = \sum_{i=1}^{n}[(\boldsymbol{x_i} - \boldsymbol{x_0}) \cdot \boldsymbol{n}]^2 = \boldsymbol{n^T S n} \tag{9}$$

其中$\boldsymbol{S} = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix}$为协方差矩阵。由于其实对称，故可以进行正交分解：

$$\boldsymbol{S} = \boldsymbol{Q \Lambda Q^T} \tag{10}$$

$\boldsymbol{\Lambda}$为特征值组成的对角矩阵，$\boldsymbol{Q}$为单位特征向量组成的正交矩阵。（9）式表明，最小平方损失等同于在所有的向量中寻找与$\boldsymbol{S}$的Rayleigh商中最小的一个，设特征值$\lambda_1 > \lambda_2$，对应的特征向量分别为$\boldsymbol{q_1}, \boldsymbol{q_2}$，则当且仅当

$$\boldsymbol{n^T}\begin{bmatrix}\boldsymbol{q_1} & \boldsymbol{q_2}\end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{11}$$

也即：

$$\boldsymbol{n} = \boldsymbol{q_2} \tag{12}$$

Rayleigh商取得最小值λ_2。

综上所述，直线的法向量$\boldsymbol{n}$应该取协方差矩阵的最小特征值所对应的特征向量，而$\boldsymbol{x_0}$可以取各数据点的均值$\bar{\boldsymbol{x}}$（或该点与法向量确定直线上的其他点，但均值是最容易计算也最直观的），此时平方误差损失最小。

4. 多元线性回归

(a)

$$J_{(\boldsymbol{\beta})} = ||\boldsymbol{X\beta} - \boldsymbol{Y}||_2^2 + \lambda||\boldsymbol{\beta}||_2^2 = (\boldsymbol{\beta^T X^T} - \boldsymbol{Y^T})(\boldsymbol{X\beta} - \boldsymbol{Y}) + \lambda\boldsymbol{\beta^T\beta} \tag{13}$$

直接求导，得到：

$$\frac{\partial J_{(\boldsymbol{\beta})}}{\partial \boldsymbol{\beta}} = 2\boldsymbol{X^T X \beta} - 2\boldsymbol{X^T Y} + 2\boldsymbol{\beta} \tag{14}$$

取最小值时，导数为零，且$\boldsymbol{X^T X} + \lambda\boldsymbol{I}$必可逆，因此参数为：

$$\boldsymbol{\beta} = (\boldsymbol{X^T X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X^T Y} \tag{15}$$

(b)

$\lambda = 1$，有:

$$\boldsymbol{\beta} = \begin{bmatrix} 0.6143 \\ 0.5481 \\ 0.0662 \end{bmatrix} \tag{16}$$

$\lambda = 5$，有:

$$\boldsymbol{\beta} = \begin{bmatrix} 0.3909 \\ 0.3721 \\ 0.0188 \end{bmatrix} \tag{17}$$

$\lambda = 10$，有:

$$\boldsymbol{\beta} = \begin{bmatrix} 0.2687 \\ 0.2669 \\ 0.0019 \end{bmatrix} \tag{18}$$

可以发现，λ越大，求得的参数各分量越小，而总体而言具有较好的稳定性。