人工智能基础 第 5 次作业

1、证明课件中一元线性回归的平方和分解公式。对于 n 组观测点 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，通过最小二乘法求得线性回归方程为 $\hat{y} = \hat{\omega}x + b$，试证明下式成立：

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

【解】：

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

只需证第 3 项为 0。

由最小二乘法，可知：

$$\hat{y}_i = \hat{\omega}x_i + b,\ \ \bar{y} = \hat{\omega}\bar{x} + b,\ \ \hat{\omega} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

从而：

$$\hat{y}_i - \bar{y} = \hat{\omega}(x_i - \bar{x})$$

$$y_i - \hat{y}_i = (y_i - \bar{y}) - (\hat{y}_i - \bar{y}) = (y_i - \bar{y}) - \hat{\omega}(x_i - \bar{x})$$

代入第 3 项，可得：

$$
\begin{aligned}
2\sum_{i=1}^{n}(\hat{y}_i-\bar{y})(y_i-\hat{y}_i) &= 2\sum_{i=1}^{n}\hat{\omega}(x_i-\bar{x})\left[(y_i-\bar{y})-\hat{\omega}(x_i-\bar{x})\right] \\
&= 2\hat{\omega}\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})-\hat{\omega}\sum_{i=1}^{n}(x_i-\bar{x})^2\right] \\
&= 2\hat{\omega}\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})-\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right] \\
&= 0
\end{aligned}
$$

因此原式得证。

2、调查某地引用水源中某重金属元素含量和该地人群患病率之间的关系，收集到的调查结果如下：

重金属含量（mg/L）	0.47	0.64	1.00	1.47	1.60	2.86	3.21	4.71
患病率（%）	22.37	23.31	25.32	22.29	28.57	35.00	46.07	46.08

（a）. 画出患病率关于重金属含量的散点图；

（b）. 利用表中数据求患病率 (y) 关于重金属含量 (x) 的一元线性回归方程和确定系数 r^2（写出计算过程，计算结果保留小数点后 4 位）；

【解】：

（a）如下图所示。

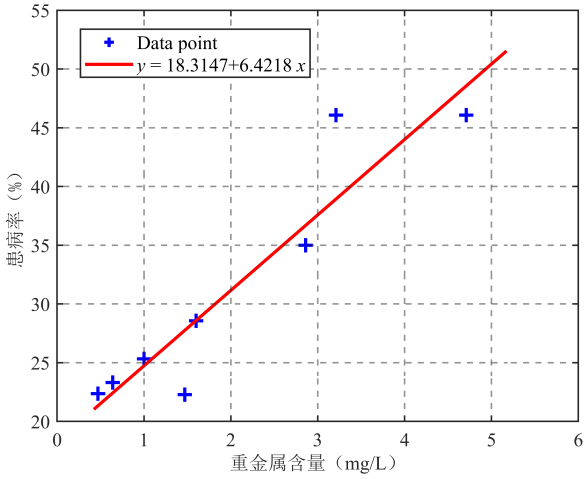

图 1: 第 2 题 (a)

（b）计算得到 $\bar{x} = 1.9950,\ \bar{y} = 31.1262$。从而：

$$S_{xx} = \sum_{i=1}^{8}(x_i - 1.995)^2 = 15.1790$$
$$S_{xy} = \sum_{i=1}^{8}(x_i - 1.995)(y_i - 31.1262) = 97.4771$$
$$S_{yy} = \sum_{i=1}^{8}(y_i - 31.1262)^2 = 718.0282$$

则斜率为：

$$\hat{\omega} = \frac{S_{xy}}{S_{xx}} = 6.4218$$

截距为：

$$\hat{b} = \bar{y} - \hat{\omega}\bar{x} = 18.3147$$

一元线性回归方程为 $\hat{y} = 6.4218x + 18.3147$。

确定系数：

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{97.4771^2}{15.1790 \times 718.0282} = 0.8718$$

4、对于多元回归问题 $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}$，如果采用最小二乘法优化平方误差损失，即：

$$J(\boldsymbol{\beta}) = \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Y}\|_2^2$$

可以得到：

$$\boldsymbol{\beta} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

而当 $\boldsymbol{X}$ 不是列满秩或某些列之间线性相关性较高时，$\boldsymbol{X}^T\boldsymbol{X}$ 接近奇异，用上述函数计算得到参数 $\boldsymbol{\beta}$ 的值误差较大，缺乏稳定性和可靠性。为了防止参数 $\boldsymbol{\beta}$ 值出现异常现象，我们在损失函数中加入一个正则化项，用损失一定的无偏性来换取参数 $\boldsymbol{\beta}$ 数值的高稳定性，即损失函数为：

$$J(\boldsymbol{\beta}) = \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Y}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2,\ \ \lambda > 0$$

此时，线性回归问题转化为了岭回归问题（Ridge Regression）。

（a）. 试证明当上述损失函数最小时，参数 $\boldsymbol{\beta} = \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}$。其中，$\boldsymbol{I}$ 是单位矩阵，

$\boldsymbol{X}^T\boldsymbol{X}+\lambda\boldsymbol{I}$ 可逆。

（提示：对于长度为 n 的列向量 $\boldsymbol{w}$，$\frac{\partial \boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w}}{\partial \boldsymbol{w}}=2\boldsymbol{A}\boldsymbol{w}$，$\frac{\partial \boldsymbol{w}^T\boldsymbol{A}}{\partial \boldsymbol{w}}=\frac{\partial \boldsymbol{A}^T\boldsymbol{\omega}}{\partial \boldsymbol{w}}=\boldsymbol{A}$）

（b）.令：

$$\boldsymbol{X}=\begin{pmatrix}1 & -1 & 2\\ 1 & 0 & 1\\ 1 & 2 & -1\\ 1 & 1 & 0\end{pmatrix},\ Y=\begin{pmatrix}1.3 & -0.5 & 2.6 & 0.9\end{pmatrix}^T$$

分别计算当损失函数中惩罚项 $\lambda=1,5,10$ 的情况下，参数 $\boldsymbol{\beta}$ 的取值。

【解】：

（a）损失函数可写为：

$$J(\boldsymbol{\beta})=(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Y})^T(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Y})+\lambda\boldsymbol{\beta}^T\boldsymbol{\beta}$$

对 $\boldsymbol{\beta}$ 求导并令为零，有：

$$\frac{\partial J(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}=2\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Y})+2\lambda\boldsymbol{\beta}=\boldsymbol{0}$$

也即：

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

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

（b）$\lambda=1$ 时：

$$\boldsymbol{\beta}=\begin{pmatrix}0.6143\\ 0.5481\\ 0.06623\end{pmatrix}$$

$\lambda=5$ 时：

$$\boldsymbol{\beta}=\begin{pmatrix}0.3909\\ 0.3721\\ 0.01879\end{pmatrix}$$

$\lambda = 10$ 时：

$$\boldsymbol{\beta} = \begin{pmatrix} 0.26875 \\ 0.266875 \\ 0.001875 \end{pmatrix}$$