人工智能导论 作业 5

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$(xi，yi)$，通过最小二乘法求得线性回归方程为 $\hat{y}=\hat{w}x+b$，试证明下式成立：

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

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

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

故只需证明$\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})=0$.

代入$\hat{y}=\hat{w}x+b$，得到

$$\sum_{i=1}^{n}(y_i-\hat{w}x_i-b)(\hat{w}x_i+b-\bar{y})$$

再由$b=\bar{y}-\hat{w}\bar{x}$得

$$\sum_{i=1}^{n}(y_i-\hat{w}x_i-\bar{y}+\hat{w}\bar{x})(\hat{w}x_i+\bar{y}-\hat{w}\bar{x}-\bar{y})$$

$$=\sum_{i=1}^{n}(y_i-\bar{y}+\hat{w}\bar{x}-\hat{w}x_i)(\hat{w}x_i-\hat{w}\bar{x})$$

$$=\sum_{i=1}^{n}\big(y_i-\bar{y}-(\hat{w}x_i-\hat{w}\bar{x})\big)(\hat{w}x_i-\hat{w}\bar{x})$$

$$=\sum_{i=1}^{n}(y_i-\bar{y})(\hat{w}x_i-\hat{w}\bar{x})-\sum_{i=1}^{n}(\hat{w}x_i-\hat{w}\bar{x})^2$$

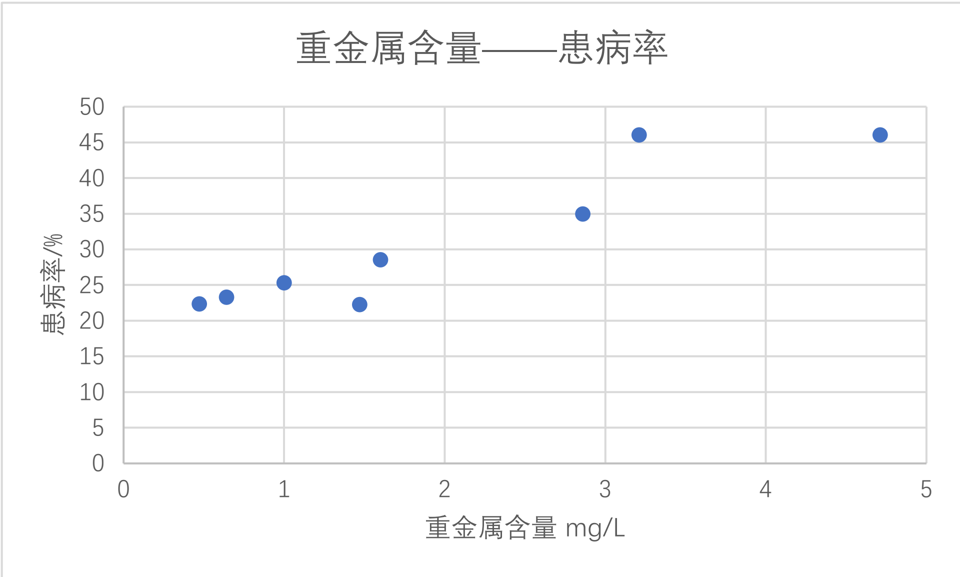
$$=\hat{w}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})-\hat{w}^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$

再由$\hat{w}=\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$，代入上式即可将上式变为 0。因此$\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})=0$，原式成立。

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

1）画出患病率关于重金属含量的散点图；



2）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）；

设$y = wx + b$

$\bar{y} = \frac{1}{8}\sum y_i = 31.12625$，$\bar{x} = \frac{1}{8}\sum x_i = 1.995$

计算$(x_i - \bar{x})$、$(y_i - \bar{y})$、$(x_i - \bar{x})^2$、$(y_i - \bar{y})(x_i - \bar{x})$、$(y_i - \bar{y})^2$如下

序号 i	1	2	3	4	5	6	7	8
$(x_i - \bar{x})$	-1.525	-1.355	-0.995	-0.525	-0.395	0.865	1.215	2.715
$(y_i - \bar{y})$	-8.75625	-7.81625	-5.80625	-8.83625	-2.55625	3.87375	14.94375	14.95375
$(x_i - \bar{x})^2$	2.325625	1.836025	0.990025	0.275625	0.156025	0.748225	1.476225	7.371225
$(y_i - \bar{y})(x_i - \bar{x})$	13.35328	10.59102	5.777219	4.639031	1.009719	3.350794	18.15666	40.59943
$(y_i - \bar{y})^2$	76.67191	61.09376	33.71254	78.07931	6.534414	15.00594	223.3157	223.6146

$$w = \frac{\sum_{i=1}^{8}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{8}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{8}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{8}(x_i - \bar{x})^2} = \frac{97.47715}{15.179} = 6.421843$$

$$b = \bar{y} - w\bar{x} = 18.3147$$

确定线性回归方程：$y = 6.4218x + 18.3147$

按照此方程估计如下：

序号 i	1	2	3	4	5	6	7	8
y_i	22.37	23.31	25.32	22.29	28.57	35.00	46.07	46.08
$\hat{y}_i$	21.33295	22.42465	24.7365	27.75475	28.58958	36.68105	38.92868	48.56138
$(\hat{y}_i - \bar{y})^2$	95.9088	75.71781	40.82891	11.36704	6.434695	30.85578	60.87788	303.9837
$(y_i - \bar{y})^2$	76.67191	61.09376	33.71254	78.07931	6.534414	15.00594	223.3157	223.6146

计算确定系数：

$$\mathrm{r}^2=\frac{\sum_{i=1}^{8}(\widehat{y}_1-\bar{y})^2}{\sum_{i=1}^{8}(y_i-\bar{y})^2}=\frac{625.9746}{718.0282}=0.8718$$

4.多元回归问题$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}$，使用以下损失函数

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

1）证明以上损失函数最小时，$\beta=(\mathrm{X}^{\mathrm{T}}\mathrm{X}+\lambda\mathrm{I})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{Y}$，这里$X^TX+\lambda T$可逆

$$J(\beta)=\left|\left|X\beta-Y\right|\right|_2^2+\lambda\left|\left|\beta\right|\right|_2^2=(X\beta-Y)^T(X\beta-Y)+\lambda\beta^T\beta$$

$$\frac{\partial J}{\partial\beta}=\frac{\partial\left((X\beta-Y)^T(X\beta-Y)\right)}{\partial\beta}+\frac{\partial(\lambda\beta^TI\beta)}{\partial\beta}$$

$$=2\frac{\partial(X\beta-Y)}{\partial\beta}(X\beta-Y)+2\lambda I\beta$$

$$=2\mathrm{X}^{\mathrm{T}}(X\beta-Y)+2\lambda I\beta$$

$$=2\mathrm{X}^{\mathrm{T}}X\beta+2\lambda I\beta-2\mathrm{X}^{\mathrm{T}}Y=2(\mathrm{X}^{\mathrm{T}}X+\lambda I)\beta-2\mathrm{X}^{\mathrm{T}}Y$$

令$\frac{\partial J}{\partial\beta}=0$，则$(\mathrm{X}^{\mathrm{T}}X+\lambda I)\beta-\mathrm{X}^{\mathrm{T}}Y=0$，即$(\mathrm{X}^{\mathrm{T}}X+\lambda I)^{-1}\mathrm{X}^{\mathrm{T}}Y$

2）给定 X、Y，分别计算λ取 1、5、10 时的β。

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

代入公式$\beta=(\mathrm{X}^{\mathrm{T}}\mathrm{X}+\lambda\mathrm{I})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{Y}$即可，结果如下

当λ取 1 时，$\beta=(0.6143,\ 0.5481,\ 0.0662)^{\mathrm{T}}$；

当λ取 5 时，$\beta=(0.3909,\ 0.3721,\ 0.0188)^{\mathrm{T}}$；

当λ取 10 时，$\beta=(0.2687,\ 0.2669,\ 0.0019)^{\mathrm{T}}$；