第五次作业

2. 试证明课件 26 页的一元线性回归的平方和分解公式。即对于n组观测点$(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$，通过最小二乘法已求得线性回归方程为$\hat{y}=\hat{\omega}x+b$，证明下面的等式成立：

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

其中$\overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$。

左侧：

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

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

将 $\hat{y}=\hat{\omega}x+\hat{b}$ 代入，

令 $f=\sum_{i=1}^{n}[y_i-(\hat{\omega}x+\hat{b})]^2$

由平方误差最小，有：

$$\frac{\partial f}{\partial \hat{b}}=-2\sum_{i=1}^{n}(y_i-\hat{y}_i)=0$$

$$\frac{\partial f}{\partial \hat{\omega}}=-2\sum_{i=1}^{n}x_i(y_i-\hat{y}_i)=0$$

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

$$=0$$

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

#

3. 设在一个 K 分类问题中，一个样例预测为第 k 类的概率建模为如下的对数线性模型

$$\log P(Y=k)=\boldsymbol{\beta}_k\boldsymbol{x}-\log Z$$

其中$P(Y=k)$表示样例预测为第 k 类的概率，$\boldsymbol{x}$是输入的样例数据，$\boldsymbol{\beta}_k$为权重，二者都是向量，等式右边补充了一项$-\log Z$来保证模型预测的所有类别的概率集合构成一个概率分布，即模型预测的所有类别的概率之和为 1。

试推导如下结论：通过该对数线性模型，将该样例预测为第 k 类的概率为

$$P(Y=k)=\frac{e^{\boldsymbol{\beta}_k\boldsymbol{x}}}{\sum_{j=1}^{K}e^{\boldsymbol{\beta}_j\boldsymbol{x}}}$$

即我们熟悉的 Softmax 回归模型。

$$\log\frac{P(Y=1)}{P(Y=K)}=\log P(Y=1)-\log P(Y=K)=(\beta_1-\beta_K)x$$

$$\cdots\cdots$$

$$\log\frac{P(Y=K-1)}{P(Y=K)}=\log P(Y=K-1)-\log(Y=K)=(\beta_{K-1}-\beta_K)x$$

$$\Longrightarrow\quad P(Y=i)=P(Y=K)e^{(\beta_i-\beta_K)x}$$

$$\text{又}\quad P(Y=K)+\sum_{i=1}^{K-1}P(Y=i)=1$$

$$\therefore\quad P(Y=K)=\frac{1}{1+\sum_{i=1}^{K-1}e^{(\beta_i-\beta_K)x}}$$

$$\therefore\quad P(Y=i)=\frac{e^{(\beta_i-\beta_K)x}}{1+\sum_{j=1}^{K-1}e^{(\beta_j-\beta_K)x}}\quad(\text{同乘}\ e^{\beta_K x})$$

$$=\frac{e^{\beta_i x}}{\sum_{j=1}^{K}e^{\beta_j x}}$$

另解 $\sum_{i=1}^{K}P(Y=i)=1$ 有 $Z=\sum_{j=1}^{K}e^{\beta_j x}$

则 $P(Y=i)=\dfrac{e^{\beta_i x}}{\sum_{j=1}^{K}e^{\beta_j x}}$