# 第五次作业

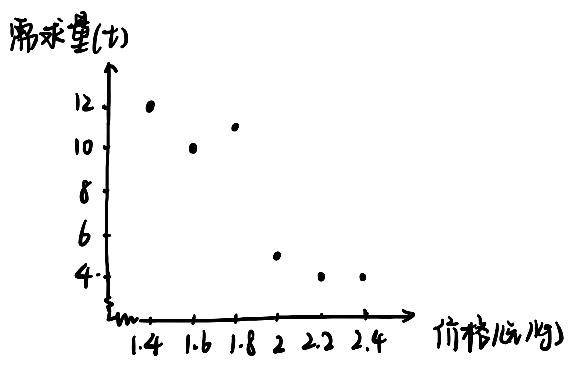
1.2 各为选1题，3,4各为1题

1. 调查到某地区一种商品的价格和需求量之间的几组历史数据，结果如下：

| | | | | | | |
|---|---|---|---|---|---|---|
| $x$ | 价格(元/kg) | 1.4 | 1.6 | 1.8 | 2 | 2.2 | 2.4 |
| $y$ | 需求量(t) | 12 | 10 | 11 | 5 | 4 | 4 |

(1) 请画出该商品需求量关于价格的散点图；
(2) 试利用表中数据求需求量(y)关于价格(x)的线性回归方程和回归系数$r^2$(请写出详细的计算过程，计算结果保留 4 位有效数字)；
(3) 试计算平均绝对误差(MAE)和均方误差(MSE)来评估(2)中拟合的线性函数的好坏。

(1)




(2) 由表中数据求得：$\bar{y} \approx 7.667$ $\bar{x} \approx 1.900$

$S_{xx} = \sum_{i=1}^{6}(x_i-\bar{x})^2 = 0.7000$

$S_{xy} = \sum_{i=1}^{6}(x_i-\bar{x})(y_i-\bar{y}) \approx -6.400$

$S_{yy} = \sum_{i=1}^{6}(y_i-\bar{y})^2 \approx 69.33$

$\therefore \hat{w} = \frac{S_{xy}}{S_{xx}} \approx -9.143$ $\hat{b} = \bar{y} - \hat{w}\cdot\bar{x} \approx 25.04$

$\therefore$ 回归方程：$\hat{y} = -9.143x + 25.04$

$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \approx 0.8440$

(3) $MAE = \frac{1}{6}\sum_{i=1}^{6}|y_i - (\hat{w}x+\hat{b})| \approx 1.109$

$MSE = \frac{1}{6}\sum_{i=1}^{6}(y_i - (\hat{w}x+\hat{b}))^2 \approx 1.803$

可见，MAE，MSE均较小，说明拟合较好。

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。

由题： $P(Y=k) = e^{\vec{\beta}_k\cdot\vec{x} - \log Z}$.

$\therefore \sum_{i=1}^{K}(Y=i) = \sum_{i=1}^{K} e^{\vec{\beta}_i\cdot\vec{x}-\log Z} = 1$

$\therefore \frac{\sum_{i=1}^{K} e^{\vec{\beta}_i\cdot\vec{x}}}{Z} = 1$ $\therefore Z = \sum_{i=1}^{K} e^{\vec{\beta}_i\cdot\vec{x}}$

$\therefore P(Y=k) = \frac{e^{\vec{\beta}_k\cdot\vec{x}}}{\sum_{i=1}^{K} e^{\vec{\beta}_i\cdot\vec{x}}}$

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$(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$，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：

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

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

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

即证 $\sum_{i=0}^{n}(\hat{y}_i-\bar{y})(y_i-\hat{y}_i) = 0$

$\because \hat{y}_i = \hat{w}x_i + \hat{b}$

$\therefore \sum_{i=1}^{n}\hat{y}_i = \sum_{i=0}^{n}(\hat{w}x_i + \bar{y} - \hat{w}\bar{x}) = \sum_{i=0}^{n}\bar{y} = \sum_{i=0}^{n} y_i$

左边 $= \sum_{i=0}^{n}\hat{y}_i y_i - \sum_{i=0}^{n}\bar{y}y_i - \sum_{i=0}^{n}\hat{y}_i^2 + \sum_{i=0}^{n}\bar{y}\cdot\hat{y}_i$

$= \sum_{i=0}^{n}\hat{y}_i(y_i-\hat{y}_i) + \bar{y}(\sum_{i=0}^{n}\hat{y}_i - \sum_{i=0}^{n}y_i)$

$= \sum_{i=0}^{n}\hat{y}_i(y_i-\hat{y}_i) = \sum_{i=0}^{n}[w(x_i-\bar{x})+(y_i-\bar{y})](w(x_i-\bar{x})+\bar{y})$

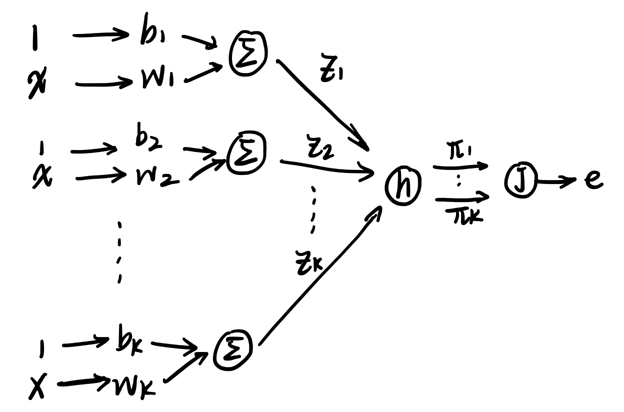
$= w(S_{xy} - wS_{xx})$

$= 0$

$\therefore$ 得证

4. 对于一个 K 分类任务，即输入数据的标签$y \in \{1,2,\ldots,K\}$。构建一个 Softmax 回归模型，设该模型对应的 K 个类别的权重分别为$w_1,w_2,\ldots,w_K$，偏置为$b_1,b_2,\ldots,b_K$，损失函数为交叉熵损失，简单起见，输入数据$x$、每一类的权重$w_i$和偏置$b_i$都是标量。给定单个输入数据$(x,y)$，试参考课件中的 Logistic 回归的计算方式，给出使用梯度下降求解该模型时由输入计算输出和由输出计算梯度的过程。

(1) 由输入计算输出：




对给定的单个数据$(x,y)$：

$z_i = w_i x + b_i$

$\pi_i = \frac{e^{z_i}}{\sum_{i=1}^{K} e^{z_i}}$

$e = J(\pi) = -\sum_{i=1}^{K} y_i \log\pi_i$

(2) 由输出计算梯度：

$\frac{\partial e}{\partial w_i} = -\sum_{m=1}^{K}\frac{\partial e_m}{\partial \pi_m}\cdot\frac{\partial \pi_m}{\partial z_i}\cdot\frac{\partial z_i}{\partial w_i}$

$\frac{\partial e}{\partial b_i} = -\sum_{m=1}^{K}\frac{\partial e_m}{\partial \pi_m}\frac{\partial \pi_m}{\partial z_i}\frac{\partial z_i}{\partial b_i}$

$\because \frac{\partial \pi_m}{\partial z_i} = \begin{cases} \pi_i(1-\pi_i) & (m=i) \\ -\pi_i\pi_m & (m\neq i) \end{cases}$

$\frac{\partial e_m}{\partial \pi_m} = \frac{y_m}{\pi_m}$

$\frac{\partial z_i}{\partial b_i} = 1$, $\frac{\partial z_i}{\partial w_i} = x$

$\therefore \frac{\partial e}{\partial w_i} = (\pi_i - y_i)x$

$\frac{\partial e}{\partial b_i} = \pi_i - y_i$