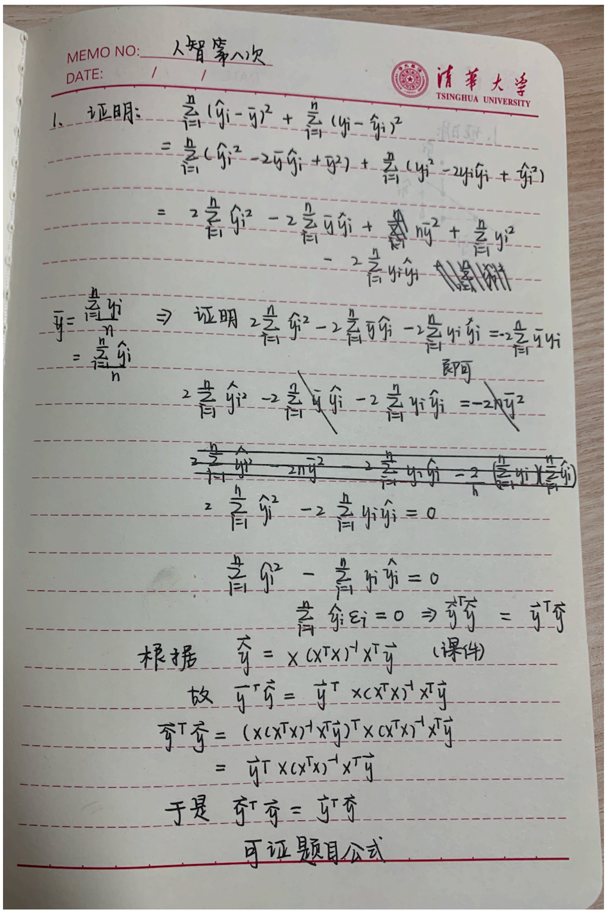
1. 

MEMO NO: 人智第八次

DATE: / /

清華大学 TSINGHUA UNIVERSITY

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

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

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

$\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{\sum_{i=1}^{n}\hat{y}_i}{n}$ $\Rightarrow$ 证明 $2\sum_{i=1}^{n}\hat{y}_i^2-2\sum_{i=1}^{n}\bar{y}\hat{y}_i-2\sum_{i=1}^{n}y_i\hat{y}_i=-2\sum_{i=1}^{n}\bar{y}y_i$ 即可

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

$\cancel{2\sum_{i=1}^{n}\hat{y}_i^2-2n\bar{y}^2-2\sum_{i=1}^{n}y_i\hat{y}_i=-\frac{2}{n}\left(\sum_{i=1}^{n}y_i\right)\left(\sum_{i=1}^{n}\hat{y}_i\right)}$

$2\sum_{i=1}^{n}\hat{y}_i^2-2\sum_{i=1}^{n}y_i\hat{y}_i=0$

$\sum_{i=1}^{n}\hat{y}_i^2-\sum_{i=1}^{n}y_i\hat{y}_i=0$

$\sum_{i=1}^{n}\hat{y}_i\varepsilon_i=0\Rightarrow\vec{\hat{y}}^T\vec{\hat{y}}=\vec{y}^T\vec{\hat{y}}$

根据 $\vec{\hat{y}}=X(X^TX)^{-1}X^T\vec{y}$ (课件)

故 $\vec{y}^T\vec{\hat{y}}=\vec{y}^TX(X^TX)^{-1}X^T\vec{y}$

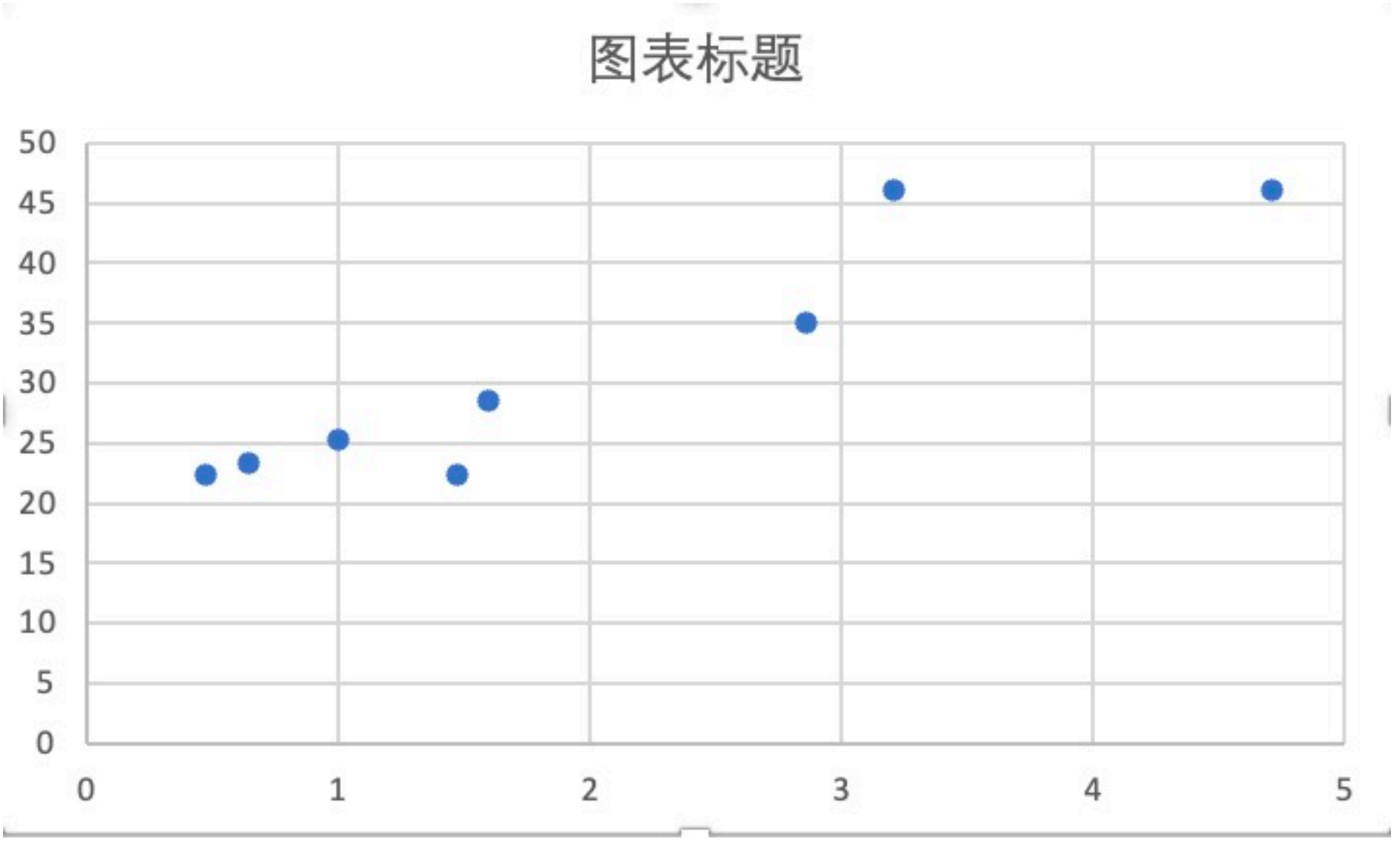
$\vec{\hat{y}}^T\vec{\hat{y}}=(X(X^TX)^{-1}X^T\vec{y})^TX(X^TX)^{-1}X^T\vec{y}$

$=\vec{y}^TX(X^TX)^{-1}X^T\vec{y}$

于是 $\vec{\hat{y}}^T\vec{\hat{y}}=\vec{y}^T\vec{\hat{y}}$

可证题目公式

2. (a)



(b)

$$\hat{w}=\frac{s_{xy}}{s_{xx}}=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(x_i-\bar{x}\right)}=6.422$$

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

$$r^2=0.872$$

3. (a): 普通最小二乘不能使用，因为此时x也存在观测误差，通过方法优化的参数不是最优参数，需要使用正交最小二乘法

   (b): 根据向量法求距离可得如下距离平方和公式

$$\frac{1}{n}\sum_{i=1}^{n}(\vec{n}(\vec{x_0}-\vec{x}_i))^2$$

   (c): 由于直线参数固定，故单位法向量$\vec{n}$固定，假设为(a,b)，则原式可化简为:

$$\frac{1}{n}\sum_{i=1}^{n}((a(x_0-x_i)+b(y_0-y_i)))^2$$

   求出x_0和y_0使得此式最小，对x_0求导，其中y_0是x_0的函数值：

$$\frac{dMSE}{dx_0}=\frac{1}{n}\sum_{i=1}^{n}2a(a(x_0-x_i)+b(y_0-y_i))$$

$$\frac{dMSE}{dy_0}=\frac{1}{n}\sum_{i=1}^{n}2b(a(x_0-x_i)+b(y_0-y_i))$$

   所以此时可得:

$$\text{满足}ax_0+by_0=a\bar{x}+b\bar{y}\text{时，也过}x_0=\bar{x},y_0=\bar{y}\text{处}$$

$$\frac{dMSE}{dx_0}=0$$