Homework 6

Tuesday, July 11, 2023 4:14 PM

Question 1

(a) $R = \frac{1}{n}\sum_i (y-\hat{y})^2 = \frac{1}{n}\sum_i (y-\hat{\theta}_0)^2$

$\frac{dR}{d\theta_0} = -\frac{2}{n}\sum_i (y-\hat{\theta}_0) = 0$

$\Rightarrow \sum_i (y-\theta_0) = 0 = \sum_i y - \sum_i \hat{\theta}_0 = \sum_i y - n\hat{\theta}_0$

$\Rightarrow \hat{\theta}_0 = \frac{1}{n}\sum_i y = \bar{y}$

The second derivatie gives that

$\frac{d^2R}{d\theta_0^2} = -\frac{2}{n}\sum(-1) = 2 > 0$ for all θ

Then by second derivate test, this is a minimum

(b) $R = \frac{1}{n}\sum_i |y-\hat{y}| = \frac{1}{n}\sum_i |y-\hat{\theta}_0|$

$\frac{dR}{d\theta_0} = 0 = \frac{1}{n}\sum_i \frac{d}{d\theta}|y-\hat{\theta}_0| = \sum_i \frac{d}{d\theta}|y-\hat{\theta}_0|$

If $y-\hat{\theta}_0 > 0$, $\frac{d}{d\theta}|y-\hat{\theta}_0| = 1$, If $y-\hat{\theta}_0 < 0$, $\frac{d}{d\theta}|y-\hat{\theta}_0| = -1$

$\Rightarrow \sum_{y>\theta_0} 1 - \sum_{y<\theta_0} 1 = 0$

$\Rightarrow \sum_{y>\hat{\theta}_0} 1 = \sum_{y<\hat{\theta}_0} 1$

Then there's equal amount of points greater and less than $\hat{\theta}_0$, which means $\hat{\theta}_0$ is the sample median.

Question 2

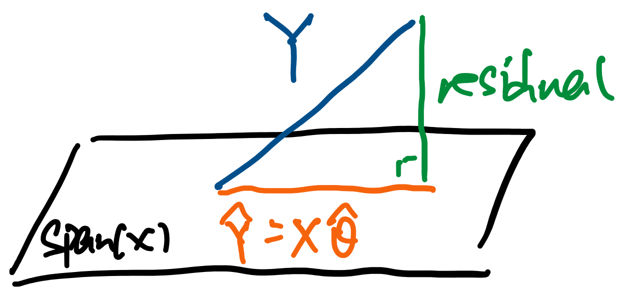
Tuesday, July 11, 2023 4:27 PM

(a) $\hat{Y}^T \cdot \vec{e} = (\hat{\theta}_0 \mathbb{X}_{:,0} + \hat{\theta}_1 \mathbb{X}_{:,1})^T \cdot \vec{e}$

$= \hat{\theta}_0 (\mathbb{X}_{:,0})^T \vec{e} + (\hat{\theta}_1 \mathbb{X}_{:,1})^T \vec{e}$

$= 0$ since residual is orthogonal to span($\mathbb{X}$)

Also



In geometry, because the error between Y and $\hat{Y}$ is the distance, so minimizing error means minimizing the distance. The smallest distance is the one connecting $\hat{Y}$ and Y and perpendicular to $\hat{Y}$. Since $\hat{Y}$ is in span(x), residual also perpendicular to $\mathbb{X}$

(b) We know that

$\hat{\theta}_0 \cdot (\mathbb{X}_{:,0})^T \vec{e} = 0$

$\Rightarrow \hat{\theta}_0 (1, 1, \ldots 1) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$

$= \hat{\theta}_0 \sum e_i = 0$

Because $\hat{\theta}_0$ is a constant, then

$\sum e_i = 0$

(c) $\hat{\theta}_1 (\mathbb{X}_{:,1})^T \vec{e} = 0$

$\Rightarrow \hat{\theta}_1 (x_1, x_2, \ldots x_n) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$

$= \hat{\theta}_1 \sum x_i e_i = 0$

Because $\hat{\theta}_1$ is a constant

$\sum x_i e_i = 0$

Question 3

Tuesday, July 11, 2023 9:23 PM

(a) we know

$$\mathbb{X}^T \vec{e} = 0$$

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{n1} \\ \vdots & & & & \\ x_{1p} & x_{2p} & x_{3p} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

$$= \begin{pmatrix} e_1 + e_2 + \cdots e_n \\ x_{11} e_1 + x_{21} e_2 + \cdots x_{n1} e_n \\ \vdots \\ x_{1p} e_1 + x_{2p} e_2 + x_{3p} e_3 + \cdots \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

For the first row, we have

$$e_1 + e_2 + \cdots e_n = \sum_i e_i = 0$$

Therefore, we have $\frac{1}{n} \sum_i e_i = 0$

For the rest of the row, we have

$$x_{1j} e_1 + x_{2j} e_2 + x_{3j} e_3 + \cdots = \sum_i x_{i,j} e_i = 0$$

(b) $$\frac{\partial MSE}{\partial \theta_0} = 0 = -\frac{2}{n} \sum_i (y_i - \theta_0 - \theta_1 x_{i,1} - \cdots \theta_p x_{i,p})$$

$$\Rightarrow \frac{1}{n} \sum_i (y_i - \theta_0 - \theta_1 x_{i,1} - \cdots \theta_p x_{i,p}) = \frac{1}{n} \sum_i e_i = 0$$

$$\frac{\partial MSE}{\partial \theta_j} = 0 = -\frac{2}{n} \sum_i (y_i - \theta_0 - \theta_1 x_{i,1} - \cdots \theta_p x_{i,p}) x_{i,j}$$

$$\Rightarrow \sum_i x_{i,j} (y_i - \theta_0 - \theta_1 x_{i,1} - \cdots \theta_p x_{i,p}) = \sum_i x_{i,j} e_i = 0$$

Question 4

Tuesday, July 11, 2023 11:12 PM

(a)

$$X = \begin{pmatrix} 1 & 1 & -1 \\ 1 & -2 & 0 \\ 1 & 1 & 1 \end{pmatrix} \Rightarrow X^T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & 0 & 1 \end{pmatrix} \qquad Y = \begin{pmatrix} -1 \\ 3 \\ 4 \end{pmatrix}$$

Then

$$X^T X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & -1 \\ 1 & -2 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$[X^T X]^{-1} = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{6} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Thus

$$\hat{\theta} = (X^T X)^{-1} X^T Y$$

$$= \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{6} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 3 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} -1 \\ 3 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} 2 \\ -\frac{1}{2} \\ \frac{5}{2} \end{pmatrix}$$

(b) MSE:

$$\frac{1}{n}\sum_i (Y - \hat{Y})^2 = \frac{1}{n}\sum_i (y_i - \theta_0 - \theta_1 X_{i,1} - \theta_2 X_{i,2})^2$$

$$= \frac{1}{3}\Big[(-1 - 2 - (-\tfrac{1}{2}) - \tfrac{5}{2}(-1))^2 +$$

$$(3 - 2 - (-\tfrac{1}{2})(-2))^2 +$$

$$(4 - 2 - (-\tfrac{1}{2}) - \tfrac{5}{2})^2\Big]$$

$$= \frac{1}{3}(0 + 0 + 0) = 0$$

Since every residule is 0, so Y should on the same plane as span(X), which means is a linear combination of the column