Quiz#2 High-dimensional Data Analysis, 2018 Spring

Use **fraction** to answer all questions, e.g., use "1/3" instead of "0.333".

Name: Huang, Xinwei

Q [+5] Data is given below:

y_i	x_i	h_i	$\hat{y}_i$
3	1		
4	-1		
2	1		
1	-1		
0	1		
2	-1		

(1) [+1] Compute the leverage h_i in the table above. Write the fraction (e.g., 1/2, not 0.5)

(2) [+1] Compute the predictor $\hat{y}_i$ in the above table

(3) [+1] Compute the training MSE

(4) [+2] Compute the testing MSE by CV

Solution:

We have

$$\mathbf{y}=\begin{bmatrix}3\\4\\2\\1\\0\\2\end{bmatrix} \text{ and } \mathbf{X}=\begin{bmatrix}1&1\\1&-1\\1&1\\1&-1\\1&1\\1&-1\end{bmatrix},$$

then $\mathbf{X}^{\mathbf{T}}=\begin{bmatrix}1&1&1&1&1&1\\1&-1&1&-1&1&-1\end{bmatrix}$ $\mathbf{X}^{\mathbf{T}}\mathbf{X}=\begin{bmatrix}6&0\\0&6\end{bmatrix}$ $\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}=\begin{bmatrix}\frac{1}{6}&0\\0&\frac{1}{6}\end{bmatrix}$.

(1)

The leverage of $\mathbf{X}$ can be expressed as

$$h_i=\mathbf{x}^{\mathbf{T}}\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{x},$$

and there are only two cases

$$\begin{pmatrix}1&-1\end{pmatrix}\begin{bmatrix}\frac{1}{6}&0\\0&\frac{1}{6}\end{bmatrix}\begin{pmatrix}1\\-1\end{pmatrix}=\frac{1}{3} \text{ and } \begin{pmatrix}1&1\end{pmatrix}\begin{bmatrix}\frac{1}{6}&0\\0&\frac{1}{6}\end{bmatrix}\begin{pmatrix}1\\1\end{pmatrix}=\frac{1}{3}.$$

(2)

The LSE estimators are

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}=\begin{bmatrix}2\\-\frac{1}{3}\end{bmatrix},$$

so $\hat{y}_i=2-\frac{1}{3}x_i$.

y_i	x_i	h_i	$\hat{y}_i$
3	1	1/3	5/3
4	-1	1/3	7/3
2	1	1/3	5/3
1	-1	1/3	7/3
0	1	1/3	5/3
2	-1	1/3	7/3

(3)

The training MSE is

$$\begin{aligned}
\mathrm{MSE} &= \frac{1}{6}\sum_{i=1}^{6}\left(y_i - \hat{y}_i\right)^2 \\
&= \frac{1}{6}\left\{\left(3-\frac{5}{3}\right)^2 + \left(4-\frac{7}{3}\right)^2 + \left(2-\frac{5}{3}\right)^2 + \left(1-\frac{7}{3}\right)^2 + \left(0-\frac{5}{3}\right)^2 + \left(2-\frac{7}{3}\right)^2\right\} \\
&= \frac{1}{6}\left\{\left(\frac{4}{3}\right)^2 + \left(\frac{5}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + \left(-\frac{4}{3}\right)^2 + \left(-\frac{5}{3}\right)^2 + \left(-\frac{1}{3}\right)^2\right\} \\
&= \frac{1}{6}\left\{\frac{16}{9} + \frac{25}{9} + \frac{1}{9} + \frac{16}{9} + \frac{25}{9} + \frac{1}{9}\right\} \\
&= \frac{14}{9}
\end{aligned}$$

(4)

The testing MSE by LOOCV is

$$\begin{aligned}
\mathrm{CV} &= \frac{1}{6}\sum_{i=1}^{6}\left(\frac{y_i - \hat{y}_i}{1-h_i}\right)^2 \\
&= \frac{1}{6}\left\{\left(3-\frac{5}{3}\right)^2 + \left(4-\frac{7}{3}\right)^2 + \left(2-\frac{5}{3}\right)^2 + \left(1-\frac{7}{3}\right)^2 + \left(0-\frac{5}{3}\right)^2 + \left(2-\frac{7}{3}\right)^2\right\} \Big/ \left(1-\frac{1}{3}\right)^2 \\
&= \frac{14}{9}\Big/\frac{4}{9} \\
&= \frac{7}{2}
\end{aligned}$$