Quiz#1 High-dimensional Data Analysis, 2018 Spring +6

an

Q [+6] Consider a simple linear model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$.

+2 (1) [+2] Derive $\hat\beta_0$ and $\hat\beta_1$. Assume the sample size is n, and $n \geq 2$.

residual sum of square RSS : $\sum_{i=1}^{n} (Y_i - \hat\beta_0 - \hat\beta_1 x_i)^2$

$\hat\beta_0, \hat\beta_1$ is the estimator that minimize the RSS, thus that satisfy the following equation:

$$\begin{cases} \frac{\partial}{\partial\beta_0} RSS = \frac{\partial}{\partial\beta_0}\sum_{i=1}^{n}(Y_i - \hat\beta_0 - \hat\beta_1 x_i)^2 = -2\sum_{i=1}^{n}(Y_i - \hat\beta_0 - \hat\beta_1 x_i) \overset{set}{=} 0 \\ \frac{\partial}{\partial\beta_1} RSS = \frac{\partial}{\partial\beta_1}\sum_{i=1}^{n}(Y_i - \hat\beta_0 - \hat\beta_1 x_i)^2 = -2\sum_{i=1}^{n}x_i(Y_i - \hat\beta_0 - \hat\beta_1 x_i) \overset{set}{=} 0 \end{cases} \Rightarrow \begin{cases} \bar Y - \hat\beta_0 - \hat\beta_1 \bar x = 0 \\ \sum_{i=1}^{n} x_i Y_i - \hat\beta_0 \sum_{i=1}^{n} x_i - \hat\beta_1 \sum_{i=1}^{n} x_i^2 = 0 \end{cases}$$

(We should check the determinant of Hessian matrix to ensure that reach the global minimum, but here is ignored.)

$$\Rightarrow \begin{cases} \hat\beta_0 = \bar Y - \hat\beta_1 \bar x \\ \hat\beta_1 = \dfrac{\sum_{i=1}^{n}(x_i-\bar x)(Y_i - \bar Y)}{\sum_{i=1}^{n}(x_i - \bar x)^2} \end{cases}, \text{ where } \begin{cases} \bar x = \dfrac{\sum_{i=1}^{n} x_i}{n} \\ \bar Y = \dfrac{\sum_{i=1}^{n} Y_i}{n} \end{cases}$$

+1 (2) [+1] Derive $SE(\hat\beta_0)$

Here use the result of part (3), first we show that

$$Cov(\bar Y, \hat\beta_1) = Cov\left(\frac{1}{n}\sum_{i=1}^{n} Y_i, \sum_{j=1}^{n}(x_j - \bar x)Y_j \Big/ \sum_{i=1}^{n}(x_i-\bar x)^2\right) = \frac{1}{n\sum_{i=1}^{n}(x_i-\bar x)^2} Cov\left(\sum_{i=1}^{n} Y_i, \sum_{j=1}^{n}(x_j-\bar x)Y_j\right)$$

$$= \frac{1}{n\sum_{i=1}^{n}(x_i-\bar x)^2}\sum_{j=1}^{n}(x_j-\bar x)\sum_{i=1}^{n} Cov(Y_i, Y_j) \underset{Cov(Y_i,Y_j)=0 \text{ if } i\neq j}{=} \frac{1}{n\sum_{i=1}^{n}(x_i-\bar x)^2}\sum_{j=1}^{n}(x_j - \bar x)\cdot Var(Y_j)$$

$$= \frac{\sigma^2}{n\sum_{i=1}^{n}(x_i-\bar x)^2}\sum_{j=1}^{n}(x_j-\bar x) = 0$$

Thus, $Var(\hat\beta_0) = Var(\bar Y - \hat\beta_1 \bar x) = Var(\bar Y) - 2\bar x\, Cov(\bar Y, \hat\beta_1) + \bar x^2 Var(\hat\beta_1)$

$$= \frac{\sigma^2}{n} - 0 + \bar x^2 \cdot \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar x)^2} \Rightarrow SE(\hat\beta_0) = \sqrt{\hat\sigma^2\left(\frac{1}{n} + \frac{\bar x^2}{\sum_{i=1}^{n}(x_i-\bar x)^2}\right)}$$

+1 (3) [+1] Derive $SE(\hat\beta_1)$

$$\hat\beta_1 = \frac{\sum_{i=1}^{n}(x_i-\bar x)(Y_i-\bar Y)}{\sum_{i=1}^{n}(x_i-\bar x)^2} = \frac{\sum_{i=1}^{n}(x_i-\bar x)Y_i - \cancel{\bar Y\sum_{i=1}^{n}(x_i-\bar x)}^{\,0}}{\sum_{i=1}^{n}(x_i-\bar x)^2} = \frac{\sum_{i=1}^{n}(x_i-\bar x)Y_i}{\sum_{i=1}^{n}(x_i-\bar x)^2}$$

$$\Rightarrow Var(\hat\beta_1) = Var\left(\frac{\sum_{i=1}^{n}(x_i-\bar x)Y_i}{\sum_{i=1}^{n}(x_i-\bar x)^2}\right) = \left(\frac{1}{\sum_{i=1}^{n}(x_i-\bar x)^2}\right)^2 Var\left(\sum_{i=1}^{n}(x_i-\bar x)Y_i\right)$$

$$= \left(\frac{1}{\sum_{i=1}^{n}(x_i-\bar x)^2}\right)^2\cdot\sum_{i=1}^{n} Var((x_i-\bar x)Y_i) = \left(\frac{1}{\sum_{i=1}^{n}(x_i-\bar x)^2}\right)^2\cdot\sum_{i=1}^{n}(x_i-\bar x)^2 Var(Y_i)$$

$$= \left(\frac{1}{\sum_{i=1}^{n}(x_i-\bar x)^2}\right)^2\cdot\sum_{i=1}^{n}(x_i-\bar x)^2\cdot Var(\beta_0+\beta_1 x_i+\varepsilon) = \frac{1}{\sum_{i=1}^{n}(x_i-\bar x)^2}\cdot Var(\varepsilon_i) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar x)^2}$$

$\hat\sigma^2$

$$\Rightarrow SE(\hat\beta_1) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar x)^2}}$$

+2 (4) [+2] Let $Y = f(x) + \varepsilon, \quad \varepsilon \sim N(0,\sigma^2)$, be a future observation. Write the formula of

(prediction error) = (reducible error)+(irreducible error)

$$E[(Y - \hat f(x))^2] = E[(f(x) + \varepsilon - \hat f(x))^2]$$

(where the notation $\hat f(x)$ is our prediction given the future observation x)

$$= E[(f(x) - \hat f(x) + \varepsilon)^2]$$

$$= E[(f(x) - \hat f(x))^2] + 2E[(f(x)-\hat f(x))\varepsilon] + E(\varepsilon^2)$$

$$\overset{indep}{=} E[(f(x)-\hat f(x))^2] + 2E(f(x)-\hat f(x))\cdot E(\varepsilon) + (E(\varepsilon^2) - [E(\varepsilon)]^2)$$

$$= E[(f(x)-\hat f(x))^2] + Var(\varepsilon)$$

$$\underset{\text{our simple linear regression}}{=} E[(f(x) - \hat\beta_0 - \hat\beta_1 x)^2] + \sigma^2$$

(prediction error) = (reducible error) + (irreducible error)