STATS 7022 – Data Science PG Assignment 2

Dang Thinh Nguyen

2024-07-11

Question 2: Leave-One-Out Cross-Validation and Linear Regression

(a).

We have: $d(A) = (p \times p)$; $d(v) = (p \times 1)$

So: $d(v^T A^{-1} v) = (1 \times p)(p \times p)(p \times 1) = (1 \times 1)$

To prove:

$$(A - vv^T)^{-1} = A^{-1} + \frac{1}{1 - v^T A^{-1} v} A^{-1} v v^T A^{-1} \qquad (1)$$

We prove: $(A - vv^T)[A^{-1} + (1 - v^T A^{-1} v)^{-1} A^{-1} v v^T A^{-1}] = 1 \qquad (2)$

Let $X = (1 - v^T A^{-1} v)^{-1} \Longrightarrow d(X) = (1 \times 1)$ so X can be considered as a scalar

The equation (2) becomes:

$(A - vv^T)[A^{-1} + XA^{-1} v v^T A^{-1}] = (A - vv^T)[A^{-1} + A^{-1} v v^T A^{-1} X]$

$= [I + vv^T A^{-1} X] + [-vv^T A^{-1} - vv^T A^{-1} v v^T A^{-1} X]$

$= [I - vv^T A^{-1}] + [vv^T A^{-1} X - vv^T A^{-1} v v^T A^{-1} X]$

$= I - vv^T A^{-1} + (v - vv^T A^{-1} v) v^T A^{-1} X$

$= I - vv^T A^{-1} + v(1 - v^T A^{-1} v) v^T A^{-1} X$

$= I - vv^T A^{-1} + vX^{-1} v^T A^{-1} X$

$= I - vv^T A^{-1} + vX^{-1} X v^T A^{-1} X$

$= I - vv^T A^{-1} + vv^T A^{-1}$

$= I$

Thus:

$$(A - vv^T)^{-1} = A^{-1} + \frac{1}{1 - v^T A^{-1} v} A^{-1} v v^T A^{-1}$$

(b).

Let $A = X^T X$ and $v = x_i$. Then:

$$X_{(i)}^T X_{(i)} = X^T X - x_i x_i^T = A - vv^T \quad \left(\text{since } X^T X = \sum_{j=1}^{n} x_j x_j^T\right)$$

The equation (1):

$$(A - vv^T)^{-1} = A^{-1} + \frac{1}{1 - v^T A^{-1} v} A^{-1} v v^T A^{-1}$$

becomes:

$$\left(X_{(i)}^T X_{(i)}\right)^{-1} = (X^T X)^{-1} + \frac{1}{1 - x_i^T (X^T X)^{-1} x_i} (X^T X)^{-1} x_i x_i^T (X^T X)^{-1}$$

$$= (X^T X)^{-1} + \frac{1}{1 - h_{i,i}} (X^T X)^{-1} x_i x_i^T (X^T X)^{-1}$$

where $h_{i,i} = [X(X^T X)^{-1} X^T]_{i,i} = x_i^T (X^T X)^{-1} x_i$

(c).

We have:

$$\hat{\beta} = (X^T X)^{-1} X^T y \Longrightarrow \hat{\beta}_{(i)} = \left(X_{(i)}^T X_{(i)}\right)^{-1} X_{(i)}^T y_{(i)} \qquad (3)$$

$$X_{(i)}^T y_{(i)} = X^T y - y_i x_i$$

$$h_{i,i} = [X(X^T X)^{-1} X^T]_{i,i} = x_i^T (X^T X)^{-1} x_i$$

From part (b), equation (3) becomes:

$$\hat{\beta}_{(i)} = \left(X_{(i)}^T X_{(i)}\right)^{-1} X_{(i)}^T y_{(i)}$$

$$= \left[(X^T X)^{-1} + \frac{1}{1 - h_{i,i}} (X^T X)^{-1} x_i x_i^T (X^T X)^{-1}\right] X_{(i)}^T y_{(i)} \qquad (4)$$

$$= \left[(X^T X)^{-1} + \frac{1}{1 - h_{i,i}} (X^T X)^{-1} x_i x_i^T (X^T X)^{-1}\right] (X^T y - y_i x_i)$$

Because (y_i, x_i) is an observation, $y_i x_i = x_i y_i$.

$$\hat{\beta}_{(i)} = (X^T X)^{-1} X^T y - (X^T X)^{-1} x_i y_i + \frac{1}{1 - h_{i,i}} (X^T X)^{-1} x_i x_i^T (X^T X)^{-1} X^T y$$

$$- \frac{1}{1 - h_{i,i}} (X^T X)^{-1} x_i x_i^T (X^T X)^{-1} x_i y_i$$

$$= \hat{\beta} - (X^T X)^{-1} x_i y_i + \frac{1}{1 - h_{i,i}} (X^T X)^{-1} x_i x_i^T \hat{\beta} - \frac{1}{1 - h_{i,i}} (X^T X)^{-1} x_i h_{i,i} y_i$$

Multiply both side with x_i^T:

$$x_i^T \hat{\beta}_{(i)} = x_i^T \hat{\beta} - x_i^T (X^T X)^{-1} x_i y_i + \frac{1}{1 - h_{i,i}} x_i^T (X^T X)^{-1} x_i x_i^T \hat{\beta} - \frac{1}{1 - h_{i,i}} x_i^T (X^T X)^{-1} x_i h_{i,i} y_i$$

$$= x_i^T \hat{\beta} - h_{i,i} y_i + \frac{h_{i,i}}{1 - h_{i,i}} x_i^T \hat{\beta} - \frac{h_{i,i}}{1 - h_{i,i}} h_{i,i} y_i$$

$$= x_i^T \hat{\beta} \left(1 + \frac{h_{i,i}}{1 - h_{i,i}}\right) - h_{i,i} y_i \left(1 + \frac{h_{i,i}}{1 - h_{i,i}}\right)$$

$$= \left(x_i^T \hat{\beta} - h_{i,i} y_i\right) \left(1 + \frac{h_{i,i}}{1 - h_{i,i}}\right)$$

$$= \left(x_i^T \hat{\beta} - h_{i,i} y_i\right) \left(\frac{1 - h_{i,i} + h_{i,i}}{1 - h_{i,i}}\right)$$

$$= \left(x_i^T \hat{\beta} - h_{i,i} y_i\right) \left(\frac{1}{1 - h_{i,i}}\right)$$

Thus:

$$x_i^T \hat{\beta}_{(i)} = \frac{1}{1 - h_{i,i}} \left(x_i^T \hat{\beta} - h_{i,i} y_i\right)$$

(d).

We have: $y = X\beta + \epsilon \Longrightarrow \epsilon = y - X\beta$

The i^{th} deletion residual is: $\tilde{\epsilon}_i = y_i - x_i^T \hat{\beta}_{(i)}$

From part (c), the equation becomes:

$$\tilde{\epsilon}_i = y_i - \frac{1}{1 - h_{i,i}} \left(x_i^T \hat{\beta} - h_{i,i} y_i\right)$$

$$= \frac{y_i\left(1 - h_{i,i}\right) - \left(x_i^T \hat{\beta} - h_{i,i} y_i\right)}{1 - h_{i,i}}$$

$$= \frac{y_i - y_i h_{i,i} - x_i^T \hat{\beta} + h_{i,i} y_i}{1 - h_{i,i}}$$

$$= \frac{y_i - x_i^T \hat{\beta}}{1 - h_{i,i}}$$

Thus:

$$\tilde{\epsilon}_i = \frac{y_i - x_i^T \hat{\beta}}{1 - h_{i,i}}$$

The leave-one-out cross validated estimate of the MSE ($CV_{(n)}$) for the linear regression model $y = X\beta + \epsilon$ is:

$$CV_{(n)} = \frac{1}{n} \sum_{i=1}^{n} MSE_i = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \frac{1}{n} \sum_{i=1}^{n} \tilde{\epsilon}_i{}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - x_i^T \hat{\beta}}{1 - h_{i,i}}\right)^2$$