Math 342W/642/742W

Recitation – Day #11 (3.18.25)

Problem #1

Consider the following R code from the class demos and labs. The numbers on the right are line numbers that will be referred to later. They are not part of the code.

```
1 y = MASS::Cars93$Price
2 var(y)
3 [1] 93.30458
4 X = as.matrix(cbind(1, MASS::Cars93[,c(7,8,12,13,14,15,17,18)]))
5 n = nrow(X)
6 n
7 [1] 93
8 Xt = t(X)
9 XtX = Xt %*% X
10 XtXinv = solve(XtX)
11 XtXinvXt = XtXinv %*% Xt
12 b = XtXinvXt %*% y
13 H = X %*% XtXinvXt
14 I_minus_H = diag(n) - H
15 yhat = H %*% y
16 e = I_minus_H %*% y
17 var(e)
18          [,1]
19 [1,] 32.70302
```

(*This problem is adapted from Problem #2 of Professor Kapelner's Spring 2024 Midterm I.*)

- What is returned by R when evaluating `length(b)`? 9

- What is returned by R when evaluating `ncol(XtX)`? 9

- What is returned by R when evaluating `ncol(H)`? 93

- What is returned by R when evaluating `Matrix::rankMatrix(I_minus_H)`? 84

- Compute SST, SSE, SSR, and R^2 to the nearest two decimal places.

  ⋆ $\text{SST} = (n-1)s_y^2 = (93-1) \times 93.30458 = 8584.02$

  ⋆ $\text{SSE} = (n-1)s_e^2 = (93-1) \times 32.70302 = 3008.68$

  ⋆ $\text{SSR} = \text{SST} - \text{SSE} = 5575.34$

  ⋆ $R^2 = 1 - \dfrac{\text{SSE}}{\text{SST}} = 1 - \dfrac{3008.68}{8584.02} = 0.65$

- What is the number of random predictors that can be added before the code throws an error and halts? 84

Problem #2

(a) Assume that $\mathbf{X} \in \mathbb{R}^{n\times(p+1)}$ with $p > 1$ composed of random realizations from iid standard normal random variables, full rank. Let $\mathbf{Q}$, $\mathbf{R}$ be the matrix results of the QR-decomposition procedure run on $\mathbf{X}$. Let $\boldsymbol{y} \in \mathbb{R}^n$ which represents a vector of measurements of a phenomenon of interest.

Prove that

$$\frac{\|\mathbf{Q}\mathbf{Q}^T\boldsymbol{y}\| + \|(I - \mathbf{Q}\mathbf{Q}^T)\boldsymbol{y}\|}{\|\boldsymbol{y}\|} \geq 1$$

Proof:

Recall that $H = \mathbf{Q}\mathbf{Q}^T$. Hence, $\mathbf{Q}\mathbf{Q}^T\boldsymbol{y} = H\boldsymbol{y} = \hat{\boldsymbol{y}}$ and $(I - \mathbf{Q}\mathbf{Q}^T)\boldsymbol{y} = (I - H)\boldsymbol{y} = \boldsymbol{e}$.

We know that $\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{e}$. By the triangle inequality,

$$\|\boldsymbol{y}\| \leq \|\hat{\boldsymbol{y}}\| + \|\boldsymbol{e}\|$$

which leads us to our desired result,

$$\frac{\|\hat{\boldsymbol{y}}\| + \|\boldsymbol{e}\|}{\|\boldsymbol{y}\|} \geq 1 \quad \blacksquare$$

(b) Claim: $\exists c \neq 0$ such that $\mathbf{X}_{\cdot 3} = c\,\mathbf{Q}_{\cdot 3}$.

Show whether the claim is true or not.

This claim is not necessarily true. According to the $QR-$decomposition,

$$\mathbf{X}_{\cdot 3} = r_{13}\,\mathbf{Q}_{\cdot 1} + r_{23}\,\mathbf{Q}_{\cdot 2} + r_{33}\,\mathbf{Q}_{\cdot 3}.$$

(*This problem is adapted from Problem #4 of Professor Kapelner's Spring 2024 Midterm I.*)