Exercise 3 : expected value of empirical risk for ols

June 27, 2024

1 Question 1: Expected Value over ϵ

We need to show:

$$E\left[R_n(\hat{\theta})\right] = E_\epsilon\left[\frac{1}{n}\|I_n - X(X^TX)^{-1}X^T\epsilon\|^2\right]$$

We know that:

$$R_n(\hat{\theta}) = \frac{1}{n}\|y - X\hat{\theta}\|^2 = \frac{1}{n}\|(I_n - X(X^TX)^{-1}X^T)\epsilon\|^2$$

So we can do:

$$E\left[R_n(\hat{\theta})\right] = E_\epsilon\left[\frac{1}{n}\|(I_n - X(X^TX)^{-1}X^T)\epsilon\|^2\right]$$

2 Question 2: Trace Identity

We need to show:

$$\sum_{(i,j)\in[1,n]^2} A_{ij}^2 = tr(A^TA)$$

By definition:

$$\sum_{(i,j)\in[1,n]^2} A_{ij}^2 = \sum_{i=1}^n\sum_{j=1}^n A_{ij}^2$$

The trace of A^TA is:

$$tr(A^TA) = \sum_{i=1}^n (A^TA)_{ii}$$

Each diagonal element of A^TA is the dot product of the i-th row of A^T and the i-th column of A:

$$(A^TA)_{ii} = \sum_{j=1}^n (A^T)_{ij}A_{ji} = \sum_{j=1}^n A_{ji}A_{ji} = \sum_{j=1}^n A_{ij}^2$$

Adding all i, we get:

$$tr(A^T A) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}^2$$

So:

$$\sum_{(i,j)\in[1,n]^2} A_{ij}^2 = tr(A^T A)$$

3 Question 3: Expected Norm of $A\epsilon$

We need to show:

$$E_\epsilon \left[\frac{1}{n} \|A\epsilon\|^2 \right] = \frac{\sigma^2}{n} tr(A^T A)$$

First, let's do $E_\epsilon \left[\|A\epsilon\|^2 \right]$. Since $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$, we have:

$$E_\epsilon \left[\epsilon^T A^T A \epsilon \right] = \sigma^2 tr(A^T A)$$

So:

$$E_\epsilon \left[\|A\epsilon\|^2 \right] = E_\epsilon \left[\epsilon^T A^T A \epsilon \right] = \sigma^2 tr(A^T A)$$

Dividing by n,

$$E_\epsilon \left[\frac{1}{n} \|A\epsilon\|^2 \right] = \frac{\sigma^2}{n} tr(A^T A)$$

4 Question 4: Idempotency of A

We need to show:

$$A^T A = A$$

where $A = I_n - X(X^T X)^{-1} X^T$.

First, note that A is symmetric:

$$A^T = (I_n - X(X^T X)^{-1} X^T)^T = I_n - X(X^T X)^{-1} X^T = A$$

To show idempotency:

$$A^2 = A$$

$$A^2 = (I_n - X(X^T X)^{-1} X^T)(I_n - X(X^T X)^{-1} X^T)$$

$$A^2 = I_n - 2X(X^T X)^{-1} X^T + X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T$$

Using $X^T X (X^T X)^{-1} = I_d$,

$$A^2 = I_n - 2X(X^T X)^{-1} X^T + X(X^T X)^{-1} I_d (X^T X)^{-1} X^T$$

$$A^2 = I_n - 2X(X^T X)^{-1} X^T + X(X^T X)^{-1} X^T$$

$$A^2 = I_n - X(X^T X)^{-1} X^T$$

$$A^2 = A$$

5 Question 5: Conclusion

Using results from Questions 1-4:

$$E\left[R_n(\hat{\theta})\right] = E_\epsilon\left[\frac{1}{n}\|I_n - X(X^TX)^{-1}X^T\epsilon\|^2\right]$$

From Question 3:

$$E_\epsilon\left[\frac{1}{n}\|A\epsilon\|^2\right] = \frac{\sigma^2}{n}tr(A^TA)$$

From Question 4, $A^TA = A$. So:

$$E_\epsilon\left[\frac{1}{n}\|A\epsilon\|^2\right] = \frac{\sigma^2}{n}tr(A)$$

The trace of $A = I_n - X(X^TX)^{-1}X^T$ is:

$$tr(A) = tr(I_n) - tr(X(X^TX)^{-1}X^T) = n - d$$

So:

$$E_\epsilon\left[\frac{1}{n}\|A\epsilon\|^2\right] = \frac{\sigma^2}{n}(n-d)$$

So:

$$E\left[R_n(\hat{\theta})\right] = \frac{n-d}{n}\sigma^2$$

6 Question 6: Expected Value of $\frac{\|y-X\hat{\theta}\|_2^2}{n-d}$

In the context of linear regression with the OLS estimator $\hat{\theta} = (X^TX)^{-1}X^Ty$, we want to find the expected value of $\frac{\|y-X\hat{\theta}\|_2^2}{n-d}$.

Residual Calculation:

$$y - X\hat{\theta} = y - X(X^TX)^{-1}X^Ty = (I_n - X(X^TX)^{-1}X^T)y$$

Norm Squared of Residuals:

$$\|y - X\hat{\theta}\|_2^2 = \|(I_n - X(X^TX)^{-1}X^T)y\|_2^2$$

Expected Value: By the properties of OLS and assuming $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$, we know:

$$E[\|y - X\hat{\theta}\|_2^2] = \sigma^2 tr(I_n - X(X^TX)^{-1}X^T)$$
$$E[\|y - X\hat{\theta}\|_2^2] = \sigma^2(n - tr(X(X^TX)^{-1}X^T))$$

Expected Value of $\frac{\|y-X\hat{\theta}\|_2^2}{n-d}$:

$$E\left[\frac{\|y - X\hat{\theta}\|_2^2}{n-d}\right] = \frac{\sigma^2(n - tr(X(X^TX)^{-1}X^T))}{n-d}$$