

Notes - Theoretical exercises 1

November 8, 2021

Homework 1.3

b)

$$\mathbb{E}(e) = \mathbb{E}(y - \hat{y}) = \mathbb{E}(X\beta + \varepsilon - X\beta) = \mathbb{E}(\varepsilon) = 0.$$

Then $\text{Var}(e_i) = \mathbb{E}((e_i - \mathbb{E}(e_i))(e_i - \mathbb{E}(e_i))) = \mathbb{E}(e_i^2)$ and thus

$$\text{trace}(\text{Cov}(e)) = \sum_{i=1}^n \text{Var}(e_i) = \sum_{i=1}^n \mathbb{E}(e_i^2).$$

Exercise 2.1

- **Question:** Why it is desirable to prove that matrix

$$\text{Cov}(b^*) - \text{Cov}(b)$$

is positive semidefinite?

- **Answer:** Because in this case we have

$$\text{Var}(b_i) \leq \text{Var}(b_i^*) \quad \forall i.$$

- **Justification:** Choose vector $a = (1 \ 0 \ 0 \ \dots 0)^T$. Then by positive semidefiniteness we have

$$\begin{aligned} a^T (\text{Cov}(b^*) - \text{Cov}(b)) a &\geq 0 \\ \Rightarrow \text{Var}(b_1^*) - \text{Var}(b_1) &\geq 0. \end{aligned}$$

The trick in the proof is to write the matrix of the arbitrary linear estimator $b^* = Cy$ in the form $C = D + (X^T X)^{-1} X^T$.

Steps of the proof:

1. From the unbiasedness of b^* it follows that $DX = 0$.
2. Calculate $\text{Cov}(b^*) = \sigma^2 DD^T + \text{Cov}(b)$ by using following facts:
 - $DX = 0$ and
 - $\text{Cov}(y) = \sigma^2 I$.

Then

$$a^T(\text{Cov}(b^*) - \text{Cov}(b))a = a^T(\sigma^2 DD^T)a = \sigma^2(D^T a)^T(D^T a) = \sigma^2\|Da\|^2 \geq 0.$$

Exercise 2.2

See also LASSO and ridge regression, for example from the book "The Elements of Statistical Learning".