Notes - Theoretical exercises 1

November 9, 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 $\mathrm{Var}(e_i) = \mathbb{E}((e_i - \mathbb{E}(e_i))(e_i - \mathbb{E}(e_i))) = \mathbb{E}(e_i^2)$ and thus

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

Exercise 2.1

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

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

  is positive semidefinite?

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

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

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

  $$a^T(\mathrm{Cov}(b^*) - \mathrm{Cov}(b))a \geq 0$$
  $$\Rightarrow \mathrm{Var}(b_1^*) - \mathrm{Var}(b_1) \geq 0.$$

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

Steps of the proof:

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

Then

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

Exercise 2.2

- Note that in the proof only zero points of derivatives are solved. However, one should still check that solution is the minimum. This step is omitted from the proof.
- See also LASSO and ridge regression, for example from the book "The Elements of Statistical Learning".

This time there are no hints for the homework.