Linear Regression and Interpretability
A Fully Worked Tutorial (Questions and Answers)

Dataset and Problem

We study a dataset of $n = 20$ students. The goal is to understand how the exam score y depends on three features:

- x_1: Study hours
- x_2: Number of practice exercises
- x_3: Sleep quality
  - $x_3 = 0$: poor sleep
  - $x_3 = 1$: good sleep

The dataset is given below:

Student	x_1	x_2	x_3	y
1	1	1	0	3
2	1	2	1	6
3	1	3	0	4
4	1	4	1	7
5	2	1	1	5
6	2	2	0	6
7	2	3	1	9
8	2	4	0	7
9	3	1	0	8
10	3	2	1	9
11	3	3	0	8
12	3	4	1	12
13	4	1	1	10
14	4	2	0	9
15	4	3	1	13
16	4	4	0	12
17	5	1	0	10
18	5	2	1	12
19	5	3	0	14
20	5	4	1	14

We want to learn a a Linear model that links y to x.
We assume a linear regression model without biais:

$$\hat{y} = w_1 x_1 + w_2 x_2 + w_3 x_3.$$

In matrix form, this can be written as:

$$y = Xw + \varepsilon,$$

where:

$$X \in \mathbb{R}^{20\times 3}, \quad w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}, \quad y \in \mathbb{R}^{20}.$$

Question 1: Please compute the least-squares estimator of the weights for the dataset written up?

Answer.

Question 2: How do we interpret the coefficients?

Answer.

Question 3: How do we compute the residual variance?

Answer.

Question 4: Can you compute the standard errors of the weights?

Answer.

Question 5: Can you compute the t-statistics?

Answer.

Question 6: How do we define feature importance?

Answer.

Question 7: How do we compute Partial Dependence Plots (PDP)?

Answer.

Question 8: How do we compute Shapley values?

Answer.

Question 9: How do we construct box plots of feature effects?

Answer.