

Statistical Machine Learning – Week 3

H. Badertscher

HS 2022/23

1 Task

Implement a simple linear regression model for the “Advertising” data using NumPy. Your task is to predict the sales using only the budget for TV advertisement.

Use the template `week3_exercise.py` and complete all Todos. The data loading is already handled in the `main()` function.

Verify your results using the following table (taken from Table 3.1 on page 68 in the book):

	Coefficient	Std. error	t-statistic	p-value
Intercept	7.0325	0.4578	15.36	< 0.0001
TV	0.0475	0.0027	17.67	< 0.0001

Comments

- As we implement linear regression for only one input variable, we can use the equations (3.4) and (3.8).
- Use the function `t.sf()` from the `scipy.stats` library (which is already imported in the template) to calculate the p-value. `t.sf()` is defined as

$$1 - P(X \leq x)$$

but has a smaller numerical error than calling the `t.cdf()` function if $P(X \leq x) \approx 1$, which is the case here.

- Use a two-sided t-test to determine the p-value.