Statistical Signal Processing 1 - Homework #09 and #10 📝

What to submit

- Hand-written or typed solutions (clearly readable). PDF version of MATLAB Live Script is also OK. No matter how you learn it, the final *explanations must be your own.*
- If you used AI, explain how it helped you to understand.
- You are allowed to work in groups and if so, ⚠️ **you must then use the MATLAB groups defined in our Moodle,** both group members submit the same report and list both group members with full names (no student numbers).
- Submit one PDF. Use filename: **HW0910_Group_xx.pdf** or **HW0910_(your name).pdf**
- Notice that due date is in Monday already (see Moodle for details).

Question 1

You are given a dataset D containing four pairs of features (x) and corresponding observations (y):

$$D = \{(1, 2), (3, 5), (4, 7), (9, 11)\}$$

Assume a simple linear model: $y = w_0 + w_1 x + \varepsilon$.

- We use the provided feature x directly (no fancy transformations like x^2).
- We include the intercept (w_0) by adding a "constant feature" (a column of ones).

1. **Full Design Matrix:** Write down the design matrix $\mathbf{X}$ and target vector $\mathbf{y}$ for the full dataset. *Question: How many unknowns (weights) and how many observations do we have?*
2. **Data Split:** Split the data into a **Training Set** (use the first 3 points as listed) and a **Test Set** (use the last point). Write these sets using proper set notation (e.g., $D_{\text{train}} = \{...\}$).
3. **Training Design Matrix:** Write the design matrix $\mathbf{X}_{\text{train}}$ and vector $\mathbf{y}_{\text{train}}$ for the training set only. *Question: How many observations are in this reduced set?*
4. **Model Fitting:** Solve for the unknown weights $\hat{\mathbf{w}} = [\hat{w}_0, \hat{w}_1]^\top$ using Ordinary Least Squares (OLS) on the **training set only**. *(Hint: You can use MATLAB/Python)*

5. **Evaluation (Training):** Calculate the **Mean Squared Error (MSE)** on the training set.

$$\text{MSE}_{\text{train}} = \frac{1}{N_{\text{train}}} \sum (y_i - \hat{y}_i)^2$$

Discussion: Why should we be careful about trusting this number too much?

6. **Evaluation (Test):** Calculate the **Squared Error** on the single test point.

$$\text{Error}^2 = (y_{\text{test}} - \hat{y}_{\text{test}})^2$$

Discussion: Is the model predicting well for this unseen point? Does it generalize?

7. **Visualization:** Plot the whole dataset (all 4 points). On the same figure, plot the line corresponding to your predicted model (use the weights $\hat{\mathbf{w}}$ you found in step 4). Visually compare how close the line is to the training points versus the test point.

Question 2 (Critical Reading & Synthesis)

Task: Read the following sections in Chapter 4 of *Hands-On Machine Learning with Scikit-Learn and PyTorch* by Aurélien Géron (Published October 2025, new book 🥰):

- "Linear Regression"
- "Polynomial Regression"
- "Regularized Linear Models"

Note: This book is for free viewing via O'Reilly (learning.oreilly.com) using your university email.

Essay: Write a short essay (approx. 0.5 pages, typed) discussing these sections. This is not a summary; it is a **critical reflection**. Compare what you read in the this book with what you learned in our "Statistical Signal Processing 1" course. *There is no single correct answer here. We are grading your ability to connect concepts and express a thoughtful engineering opinion.*