Assignment 2: *Linear Regression*

Date: *Jan 30, 2025*
Due : *Feb 13, 2025*

Acknowledgment This assignment has been adapted in part from the materials of the courses ECE421 by N. Papernot and ECE1513 by S. Emara.

Code of Honor Assignments are designed to enhance your understanding and advance your skills, constituting a significant portion of your final assessment. They must be completed individually, as engaging in any form of academic dishonesty violates the principles of the Code of Honor. If you encounter any challenges while solving the assignments, please contact the instructional team for guidance.

How to Submit This assignment has 3 questions. For each question, you need to upload one file on Crowdmark. If the question asks for your code, you need to copy and paste your code in the file that you submit.

Grading The grades add up to 100 and comprise roughly 10% of the final mark.

Deadline The deadline for your submission is on **Feb 13, 2025 at 11:59 PM**. Please note that this deadline is strict and **no late submission will be accepted.**

Questions

Question 1 [40 Points] **(Scalar Linear Regression)** We have collected the dataset of $N = 7$ points

$$\mathbb{D} = \left\{ (1,6), (2,4), (3,2), (4,1), (5,3), (6,6), (7,10) \right\}.$$

In the sequel, we denote the n-th input and label by x_n and v_n, e.g., $(x_1, v_1) = (1, 6)$. We intend to fit this data to a linear model via linear regression.

1. Draw a scatter plot of the dataset using `matplotlib` in Python.
2. We now use the affine model
$$f(x) = wx + b$$
to model this data. Write down the analytical expression of the empirical risk determined on dataset $\mathbb{D}$. The derived expression should be on the form
$$\hat{R}(w, b) = \frac{1}{2N} \sum_{n=1}^{N} A_n w^2 + B_n b^2 + C_n wb + D_n w + E_n b + F_n$$
where A_n, B_n, C_n, D_n, E_n, and F_n are expressed only as a function of x_n and v_n or constants. Do not fill-in any numerical values yet.

3. Derive the analytical expressions of w and b by minimizing the empirical risk computed in the previous question. Your expressions for parameters w and b should only depend on

$$A = \sum_{n=1}^{N} A_n \qquad B = \sum_{n=1}^{N} B_n \qquad C = \sum_{n=1}^{N} C_n \qquad D = \sum_{n=1}^{N} D_n \qquad E = \sum_{n=1}^{N} E_n$$

   Do not fill-in any numerical values yet.

4. Give approximate numerical values for w and b by plugging in values from the dataset $\mathbb{D}$.

5. Use numpy `polyfit` to double-check your solution with the scatter plot from the question earlier. This would yield the values of w and b. Paste your lines of code for this question and show you obtained the correct solution in the previous questions.

QUESTION 2 [25 Points] **(Linear Regression with Weight Vector)** Consider the dataset $\mathbb{D}$ in Question 1. We now want to solve the linear regression with the alternative approach we discusses in the lecture, i.e., looking at the bias as another weight.

1. Write the affine function $f(x) = wx + b$ as

$$f(\mathbf{x}) = \mathbf{w}^{\mathsf{T}}\mathbf{x},$$

   where we define $\mathbf{x} = [1, x]^{\mathsf{T}}$. Specify the vector $\mathbf{w}$ in terms of w and b.

2. Construct the dataset matrix $\mathbf{X}$ as $\mathbf{X} = [\mathbf{x}_, \dots, \mathbf{x}_7]$ and the label vector $\mathbf{v} = [v_1, \dots, v_7]^{\mathsf{T}}$, and derive

$$\nabla_{\mathbf{w}} \hat{R} = \frac{1}{7} \nabla_{\mathbf{w}} \|\mathbf{X}^{\mathsf{T}}\mathbf{w} - \mathbf{v}\|^2$$

   analytically. Compare your result with what you get when we use the expression derived in the lecture.

3. Find optimal model $\mathbf{w}^{\star}$ by solving the system of equations given by $\nabla_{\mathbf{w}} \hat{R} = 0$, and compare it to what we derived in the lecture in terms of the pseudo-inverse of $\mathbf{X}$.

4. Use `numPy` to implement this solution and compare it to the solution in Question 1. You may find `dot`, `matmul`, `transpose` and `linalg.inv` helpful. Paste your lines of code for this question and show you obtained the correct solution in the previous questions.

QUESTION 3 [35 Points] **(Penalizing with ℓ_2-norm)** Let us now switch to a more general case: we have a dataset

$$\mathbb{D} = \{(\mathbf{x}_n, v_n) : n = 1, \dots, N\}$$

where $\mathbf{x}_n \in \mathbb{R}^d$ is a d-dimensional vector. We focus on the linear hypothesis, i.e.,

$$f(\mathbf{x}) = \mathbf{w}^{\mathsf{T}}\mathbf{x}$$

for some $\mathbf{w} \in \mathbb{R}^d$. In this problem, we consider a *penalized* empirical risk given by

$$\tilde{R}(\mathbf{w}) = \frac{1}{2N} \sum_{i=1}^{N} (f(\mathbf{x}_i) - v_i)^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2,$$

where $\lambda > 0$ is a *hyperparameter* that controls how much importance is given to the penalty.

1. Use the gradient derivation in the lecture, i.e., the fact that

$$\nabla_{\mathbf{w}} \|\mathbf{X}^{\mathsf{T}}\mathbf{w} - \boldsymbol{v}\|^2 = 2\left(\mathbf{X}\mathbf{X}^{\mathsf{T}} - \mathbf{X}\boldsymbol{v}\right),$$

   and derive analytically $\nabla_{\mathbf{w}} \tilde{R}$.

2. Find the system of equations the optimal model $\mathbf{w}^\star$ should satisfy.
   Hint: $\mathbf{w}^\star$ *is a stationary point of* $\tilde{R}(\mathbf{w})$.

3. Solve the system of equations for $\mathbf{w}^\star$ and derive the optimal model explicitly. Compare the result with what we derived in the course. What benefits can this penalty give us?