Honor code. This assignment is individual work. The goal of this assignment is for you to put in practice the concepts we learned in the video recordings, as well as explore complementary concepts. As mentioned in the synchronous lecture, academic integrity will be strictly enforced. If for any reason you are tempted to cheat (i.e., because you are facing personal hardship), contact the instructor immediately by email.

To facilitate grading, please continue to follow the following guidelines when uploading your assignment to Crowdmark:

- On Crowdmark, you will upload a *separate* image for each question.
- You can use any tool you want to generate these answers (e.g., word, LaTeX, scan your handwriting), but each image should be easy to read and oriented properly.
- If you decide to handwrite your assignment rather than typeset it, ensure your handwriting is readable otherwise TAs will have the discretion to not grade your answer.
- Graphs produced should be clearly interpretable. Include labels on axes and a legend.

Assignment structure. The assignment contains 16 questions worth a total of 39 points.

Problem 1 Assume we collected a dataset $D = \{(x^{(i)}, t^{(i)})\}_{i\in 1..7}$ of $N = 7$ points (i.e., observations) with inputs $\{x^{(i)}\}_{i\in 1..7} = (1, 2, 3, 4, 5, 6, 7)$ and outputs $\{t^{(i)}\}_{i\in 1..7} = (6, 4, 2, 1, 3, 6, 10)$ for a regression problem with both scalar inputs and outputs.

1. (1 point) Draw a scatter plot of the dataset on a spreadsheet software (e.g., Excel).
2. (6 points) Let us use a linear regression model $g_{w,b}(x) = wx + b$ to model this data. Write down the analytical expression of the least squares loss (covered in Video 6) of this model on dataset D. Your loss should take the form of

$$\frac{1}{2N} \sum_{i\in 1..N} A_i w^2 + B_i b^2 + C_i wb + D_i w + E_i b + F_i$$

where A_i, B_i, C_i, D_i, E_i, and F_i are expressed only as a function of $x^{(i)}$ and $t^{(i)}$ or constants. Do not fill-in any numerical values yet.

3. (4 points) Derive the analytical expressions of w and b by minimizing the mean squared loss from the previous question. Your expressions for parameters w and b should only depend on $A = \sum_i A_i, B = \sum_i B_i, C = \sum_i C_i, D = \sum_i D_i$ and $E = \sum_i E_i$. Do not fill-in any numerical values yet.
4. (2 points) Give approximate numerical values for w and b by plugging in numerical values from the dataset D.

5. (0 points) Double-check your solution with the scatter plot from the question earlier: e.g., you can use Excel to find numerical values of w and b. You do not need to hand in anything here, this is just for you to verify you obtained the correct solution in the previous questions.

Problem 2 The goal of this problem is to revisit Problem 1, but solving it with a different technique known as the method of least squares. This will serve as a "warm-up" to Problem 3. In the rest of this problem, any reference to a dataset refers to the dataset described in Problem 1.

1. (1 point) Verify that one can rewrite the linear regression model $g_{w,b}(x) = wx + b$ in the simpler form of

$$g_w(\vec{x}) = \vec{x}\vec{w}$$

if one assumes each input $\vec{x}$ is a two-dimensional row vector such that a point in our dataset is now $\vec{x^{(i)}} = (x^{(i)}, 1)$ where $x^{(i)}$ is the scalar input described in Problem 1. Write the components of the new column vector $\vec{w}$ as a function of w and b from Problem 1.

2. (4 points) Derive analytically $\nabla_{\vec{w}} \|X\vec{w} - \vec{t}\|^2$ where X is a $N \times 2$ matrix such that each row of X is a vector $\vec{x^{(i)}}$ described in the previous question, and $\vec{t} = \{t^{(i)}\}_{i \in 1..7}$.

3. (1 point) Conclude that the model's weight value $\vec{w}^*$ which minimizes the least squares loss (covered in Video 6) must satisfy

$$2X^\top X\vec{w}^* - 2X^\top \vec{t} = 0$$

4. (1 point) Assuming that $X^\top X$ is invertible, derive analytically the value of $\vec{w}^*$.

5. (0 points) Using numPy, implement the solution you found in the previous question and verify that you obtain the same results for w and b than in Problem 1. You do not need to hand in anything here, this is just a way for you to verify you obtained the correct solution in the previous questions.

Problem 3 Let us now assume that D is a dataset with d features per input and $N > 0$ inputs. We have $D = \{((\vec{x_j^{(i)}})_{j \in 1..d}, t_i)\}_{i \in 1..N}$. In other words, each $\vec{x^{(i)}}$ is a column vector with d components indexed by j such that $x_j^{(i)}$ is the jth component of $\vec{x^{(i)}}$. The output $\vec{t^{(i)}}$ remains a scalar (real value).

Let us assume for simplicity that we have a simplified linear regression model, as presented in the Question 1 of Problem 2. We would like to train a regularized linear regression model, where the mean squared loss is augmented with an ℓ_2 regularization penalty $\frac{1}{2}\|\vec{w}\|_2^2$ on the weight parameter $\vec{w}$:

$$\varepsilon(\vec{w}, D) = \frac{1}{2N} \sum_{i \in 1..N} (g_{\vec{w}}(\vec{x^{(i)}}) - t^{(i)})^2 + \frac{\lambda}{2}\|\vec{w}\|_2^2$$

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

1. (3 points) Let $A = \sum_{i \in 1..N} \vec{x^{(i)}} \vec{x^{(i)}}^\top$. Give a simple analytical expression for the components of A.

2. (6 points) Let us write $\vec{b} = \sum_{i \in 1..N} t^{(i)} \vec{x^{(i)}}$, prove that the following holds:

$$\nabla \varepsilon(\vec{w}, D) = \frac{1}{N}\left(A\vec{w} - \vec{b}\right) + \lambda \vec{w}$$

3. (2 points) Write down the matrix equation that $\vec{w}^*$ should satisfy, where:

$$\vec{w}^* = \arg\min_{\vec{w}} \varepsilon(\vec{w}, D)$$

   Your equation should only involve $A, \vec{b}, \lambda, N$, and $\vec{w}^*$.

4. (3 points) Prove that all eigenvalues of A are non-negative.

5. (3 points) Demonstrate that matrix $A + \lambda N I_d$ is invertible by proving that none of its eigenvalues are zero. Here, I_d is the identity matrix of dimension d.

6. (2 points) Using the invertibility of matrix $A + \lambda N I_d$, solve the equation stated in question 3 and deduce an analytical solution for $\vec{w}^*$. You've obtained a linear regression model regularized with an ℓ_2 penalty.

*

* *