ECE1513H - Winter 2020 | **Assignment 1**
University of Toronto - LEC0101 (Prof. Papernot) | **Due Jan 20th**

This assignment contains 11 questions worth a total of 22 points. Because assignments will help you learn things that are complementary to what we covered in class, they should be completed on your own. Otherwise, you will not learn from taking this course and you are harming yourself.

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

1. (0 points) 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 mean squared loss of this model on dataset D. Your loss should take the form of

$$\frac{1}{2N} \sum_{i \in 1..N} Aw^2 + Bb^2 + Cwb + Dw + Eb + F$$

where A, B, C, D, E, and F are expressed only as a function of X_i and Y_i and constants. Do not fill-in any numerical values yet.

3. (3 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, B, C, D and E. Do not fill-in any numerical values yet.

4. (1 point) 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.

Problem 2 Let us now assume that D is a dataset with d features per input and $N > 0$ inputs. We have $D = \{((X_{ij})_{j \in 1..d}, Y_i)\}_{i \in 1..N}$. In other words, each X_i is a column vector with d components indexed by j such that X_{ij} is the jth component of X_i. The output Y_i remains a scalar (real value).

Let us assume for simplicity that $b = 0$ so we have a simplified linear regression model:

$$g_{\vec{w}}(X) = \cancel{\vec{w}X} = w^T X = X^T w$$

where $\vec{w}$ is now a vector of dimensionality d. Each component of $\vec{w}$ multiplies the corresponding feature of X, which gives the following: $g_{\vec{w}}(X_i) = \sum_{j \in 1..d} w_j X_{ij}$.

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 $\vec{w}$:

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

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

1. (2 points) Let $A = \sum_{i \in 1..N} X_i X_i^\top$. Give a simple analytical expression for the components of A.

2. (3 points) Let us write $B = \sum_{i \in 1..N} Y_i X_i$, prove that the following holds:

$$\nabla L(\vec{w}, D) = \frac{1}{N}(A\vec{w} - B) + \lambda \vec{w}$$

3. (1 point) Write down the matrix equation that $\vec{w}^*$ should satisfy, where:

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

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

4. (2 points) Prove that all eigenvalues of A are positive.

5. (2 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.

*
* *