

DCSI 440W Implementation Assignment 1 (100 points)

Due 11:59 pm February 18, 2023

General instructions.

1. This is an individual assignment. You will only need to submit one copy of the source code and report. Please, type your report in LaTeX.
2. Your source code and report must be submitted through the Moodle site.
3. Be sure to answer all the questions in your report. You will be graded based on your code as well as the report. So please write your report in a clear and concise manner. Clearly label your figures, legends, and tables.

Linear regression, closed-form solution

Please implement the closed-form solution (i.e., normal equations) of the linear regression problem, which learns from a set of N training examples $\{\mathbf{x}^i, y^i\}_{i=1}^N$ the weight vector \mathbf{w} that optimize the following Sum of Squared Error (SSE) objective:

$$J(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (y^i - \mathbf{w}^T \mathbf{x}^i)^2 \quad (1)$$

Assume that $\forall i = \overline{1, N}$ x^i is a $(d+1)$ -dimensional vector, such that the first coordinate is 1 (for more details see slides on Linear Regression posted on Moodle). Vector \mathbf{w} is also $(d+1)$ -dimensional vector. In particular you need to do the following:

- (a) Implement a function that takes a matrix X and a vector y as arguments and returns vector \mathbf{w} , which is found using normal equations.
- (b) Implement a function that takes a matrix X , vector y , and vector \mathbf{w} as arguments and returns the SSE, i.e., the value of the objective function $J(\mathbf{w})$ (you can use compact formula for the SSE from the slides).
- (c) Write a piece of code that plots a line $f(x) = w_0 + w_1x$ for given parameters w_0 and w_1 .
- (d) Load the data from the data file (posted on Moodle), where even lines contain the values of temperature in Celsius (line numeration starts with zero, zero is an even number), and odd lines contain the noisy measurements of temperature in Fahrenheit.
- (e) Find optimal \mathbf{w}^* and SSE for the data provided. Plot the regression line and data points.
- (f) Predict the temperature in Fahrenheit, for the input (temperature in Celsius) $x = 33$.

Your report should have the following structure:

- (a) You full name and assignment number.
- (b) Introduction (Briefly state the problem you are solving).
- (c) Optimal \mathbf{w}^* .
- (d) The SSE.
- (e) The plot of all data points and regression line.
- (f) The predicted temperature in Fahrenheit, for the input (temperature in Celsius) $x = 33$.