Assignment 1

Problem 1

Plot θ vs. $L(\theta)$, where $L(\theta) = \theta^2$. θ varies from -10 to +10 with a step size of 0.1. Now locate the minimum value of $L(\theta)$ with corresponding θ value from the plot.

Problem 2

Plot θ vs. $L(\theta)$, where $L(\theta) = \theta_1^2 + \theta_2^2$. θ_1 and θ_2 vary from -10 to +10 with a step size of 0.1. Locate the minimum value of $L(\theta_1, \theta_2)$ with corresponding θ_1, θ_2 values from the plot.

Problem 3

a). Plot for $L(\theta) = \sum_{i=1}^{m}(y(i) - (\theta_0 + \theta_1 x(i)))^2/m$, where m is the number of input examples and $x(i), y(i)$ are the values taken from the given dataset. Obtain the minimum value of $L(\theta)$ with corresponding θ_0, θ_1 values from the plot.

b). Apply Pseudo Inverse (Least Squares (LS)) approach to get θ vector for the cost function (objective function) $L(\theta)$ given in example 3(a). Verify whether θ_1, θ_2 obtained are the same as that found in Q3(a).

Problem 4

Calculate the value of $L(\theta)$ using the θ vector obtained by Pseudo Inverse (as done in Q3(b)). Now assume any θ vector (other than the one obtained in Q3(b)) and compute the new $L(\theta)$ value. Comment on why the Pseudo Inverse is also called LS method.

Problem 5

a). For the following X and Y, use scikit-learn to learn a linear model.

b). Solve the problem using normal equations. You may find that one of the matrices in the normal equation is non-invertible. Why does the matrix turn out to be non-invertible? Why can scikit-learn implementation still correctly solve this regression problem?

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \\ 4 & 8 \end{bmatrix} \quad Y = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$$

Problem 6

a). Show the usage of scikit-learn's linear regression module for the real estate price prediction regression problem. What is the RMS error on the test set?

b). Based on the regression coefficients, what can you comment about the importance of different features? Is it correct to assume that larger coefficients mean more important features?

c). Now, standardize the dataset to have all features on a scale of 0 to 1. Re-learn the regression coefficients and now comment on the importance of different features.

d). What is the distribution of the residuals?

e). Use cross-correlation to find the optimal set of features to use for regression. Using all possible feature sets of length 1, 2, 3, or 4, what is the optimal feature set as per the validation set and how does this set of features perform on the test set wrt the model learned on the entire feature set?