EEL 6814

HMW # 1
Due September 14, 2023

The goal is to design a LINEAR regressor using a set of features from a house sales database to determine its estimated market value. You are going to use a well-known data set, the Boston Housing data set located at

https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html

This data set has 14 attributes (features) and you can regress to price (14^{th} feature), or any other. Therefore, you should not use the selected target variable as an input.

The data set is small, so please divide it 2/3 for training and 1/3 for testing.

Summarize the results by presenting the training error and its variance across different initializations (or shuffle the order of presentations), and show also the test set error and its variance across different runs in a table.

Please use the Least Square (LS) solution, also called statistical regression, and the LMS algorithm to train the regressor. You must develop the code for this homework.

In the LS, show the effect of different levels of regularizations. When training with LMS you have to properly select the learning rate. Show the effect of the stepsize (or learning rate) by plotting the learning curve. Also show the weight tracks for the LMS. Compare the accuracy of the LMS with the analytic solution for the best initialization.

If you use the target variable as an input, what happens? Explain the result by visualizing the learned parameters of the model.

Does a bias term in the model improve performance for this problem?