1 Project 1 (Linear Regression)

Due date: February 3

Consider the problem of predicting the price of a home in Bothell, WA. We will use a linear regrwession model for the given problem. The data will be sale records in Bothell, WA from the last 3 months and it can be downloaded from

https://www.redfin.com/city/29439/WA/Bothell/filter/include=sold-3mo

using the (Download All) link.

1. After downloading the data you have to select the parameters that you want to use for your prediction. You need to select at least three parameters as input values. The value that we would like to predict is the price of the home. You will need to solve a few problems before you can implement the method itself, such as what will be your policy for rmissing data, etc. Another problem that needs to be addressed is whether you scale some of the values. For linear regression scaling is not absolutely necessary (mathematically speaking), however it might be needed as you encounter overflows from makeing matrix operations with large matrices with large numbers. To test your algorithm please validate it by using only 70% of the available data for training and the remaining 30% for testing your algorithm. You may also encounter singular matrices, in which case you can use the algorithm that we have learned in class.

2. Implement the linear regression algorithm.

3. Validate the linear regression algorithm. On the testing data, compare your prediction with the sale price of the house.

4. In order to better understand the results please plot some graphs. For example you can plot the data in a 2D plane (price vs. square footage), and if you keep all parameters constant (at an average value) you can graph the regression line (projection of the regression hyperplane). Your program should plot the regression line as described here.

You can use the numpy library and its linalg module for matrix operations, however you are not allowed to use the already implemented linear regression methods of python.

1