Project 1 Report

Dreycen Foiles, Olek Yardas

October 13, 2022

1 Introduction

We report on our efforts to use different regression techniques to predict the price of a house given various different parameters. The two techniques that we present in this report is the Least Absolute Shrinkage and Selection Operator (LASSO) and the Random Forest regression algorithm.

2 Data Preprocessing

For this project, we are given a data set made up numerous data points. These include basement area, zoning type, heating type, year built, etc. Our goal is to use these values to predict the price of the corresponding house. Before we can build a model, we must convert non-numerical data into a form that can be used with our regression techniques. To do this, we use a very simple conversion where every unique string in a column corresponds to a number. We had two different

3 Linear Model

3.1 Results

4 Random Forest

Our random forest model appeared to work will with a simple factorize command. This is because the random forest model is nonlinear and category values of 1,2,3, etc. are not going to introduce artifacts into the prediction that it would for linear models. We used the Scikit-Learn implementation of the random forest regressor and we found that even with the default parameters, we were able to get the RMSE test error below the benchmark.

4.1 Results

Split #	RMSE	Runtime (sec.)
1	0.086	2.361
2	0.087	2.328
3	0.097	2.335
4	0.093	2.434
5	0.090	2.451
6	0.087	2.432
7	0.085	2.363
8	0.097	2.342
9	0.092	2.435
10	0.088	2.422

5 Conclusion