Ames Iowa House Prediction Project

Jordan Kassof and Jose Torres

2/11/2018

Contents

Introduction

Buying a house is one of the largest financial decisions many people will make. So many factors go into someone's decision, but it can be hard to really explain why one house "felt right" and another didn't. We want to quantify the factors that add up to someone making the decision to purchase a house.

Data Description

We will be using the Ames, Iowa individual residential property sales data set freely available on Kaggle.com. The dataset contains 2,930 observations with 79 explanatory variables. All observations occur between 2006 and 2010. Given the geography dependent nature of home value, the results of below analyses can't be applied nationally. For more information on the data, or to download it yourself, visit https://www.kaggle.com/c/house-prices-advanced-regression-techniques.

See the codebook.txt file in the github repo for complete information about all variables.

Exploratory Analysis

Comments on various things noticed as exloring the data and general idea of what the cleaning was. Reference appendix for entire cleaninf functions and scripts.

Questions of Interest

We focused on two approaches, one geared towards finding the most performant model we could and one towards finding the most interpretable model we could.

Interpretable Models

Restatement of problem here

Model Selection

Type of Selection Assumptions Optional Comparing Competing Models

Parameter Interpretation

Interpretation Confidence Intervals

Predictive Models

Restatement of problem here

Model Selection

Type of Selection Assumptions Comparing Competing Models AIC, BIC, adj R2 Interval CVPress External Cross Validation Kaggle Score

Conclusion

Appendix