Data Science Final Project

Brandon Feder, Atticus Wang

May 30, 2022

Introduction

—OUR PLAN—

Response variable: soldPrice (price at which the house was sold)

Possible predictors:

- nbhd (neighborhood)
- bed (number of bedrooms)
- fullBath (number of full baths)
- halfBath (number of half baths)
- style (style)
- age (yearSold minus yearBuilt)
- originalPrice (original price)
- lastPrice (price at which the house was last sold)
- marketDays (days the house was on market)
- yearSold, daySold, monthSold (date at which the house was sold)

Regression methods we will use:

- full linear model (lm)
- subset selection (forward, backward)
- ridge (ridge)
- lasso (lasso)
- elastic net (ela)
- principal components regression (pcr)

Regression Analysis

Our final models are Lasso and PCR. Lasso has the lowest variance, while PCR has the highest R2.

Discussion and Limitations

Assumptions possibly not met, curse of dimensionality, did not incorporate census data

Conclusion

Additional Work

Introduce the other models we tried