Selection Bias

Nate Wells

Math 243: Stat Learning

October 6th, 2021

Outline

In today's class, we will. . .

- Discuss data from homework 3 (Ames House Prices)

Section 1

Ames House Price Data

Overview

- Students fit models of varying complexity based on data on 66 predictors for 1808 houses.

Overview

- Students fit models of varying complexity based on data on 66 predictors for 1808 houses.

- Models were assesses by computing rMSE on a test set of 597 houses.

- Additionally, to assess variability, rMSE was computed on 20 bootstrap samples from the test data.

Overview

- Students fit models of varying complexity based on data on 66 predictors for 1808 houses.

- Models were assesses by computing rMSE on a test set of 597 houses.

- Additionally, to assess variability, rMSE was computed on 20 bootstrap samples from the test data.

- The median model rMSE was \$33, 249.21.

- The median model standard deviation in rMSE on bootstrap samples was \$\$1,890 \$.

Overview

- Students fit models of varying complexity based on data on 66 predictors for 1808 houses.

- Models were assesses by computing rMSE on a test set of 597 houses.

- Additionally, to assess variability, rMSE was computed on 20 bootstrap samples from the test data.

- The median model rMSE was $33,249.21.

- The median model standard deviation in rMSE on bootstrap samples was $$1,890$.

- The lowest three model rMSE were

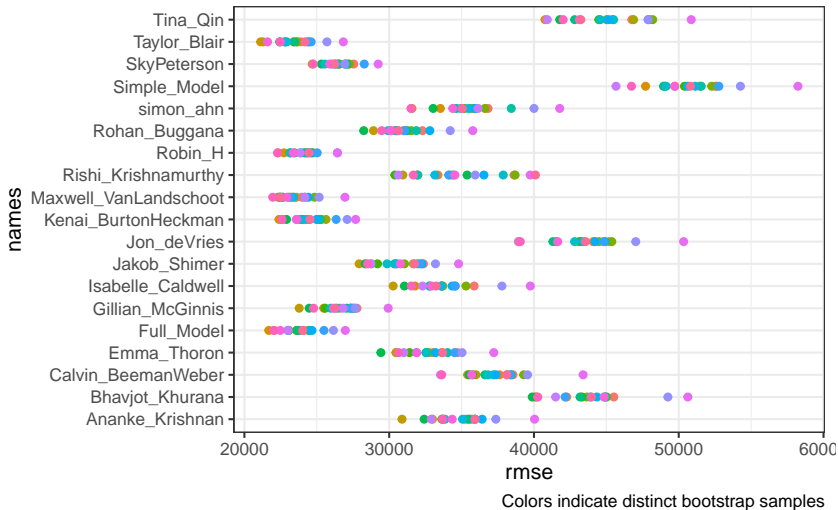| Name	Taylor	Maxwell	Robin
rMSE	$23,405	$23,507	$24,084
SD	$1,415	$1,202	$1,123

Results

Model rMSE, Based on 20 Bootstraps from Test Data
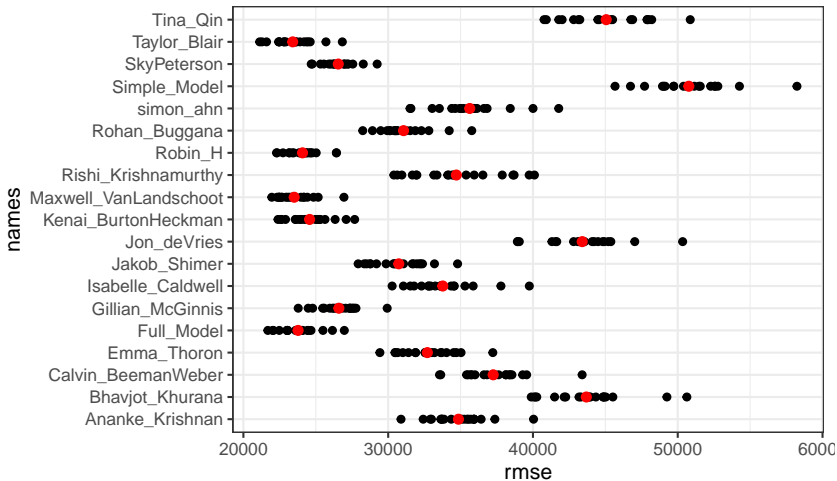
Red dot indicates mean rMSE

Results

Model rMSE, Based on 20 Bootstraps from Test Data

Colors indicate distinct bootstrap samples

Retrospective

Trends:

Retrospective

Trends:

- Models with more predictors tended to do better than models with fewer predictors

Retrospective

Trends:

- Models with more predictors tended to do better than models with fewer predictors
- Models with 0 interaction terms tended to do better than those with 1 interaction

Retrospective

Trends:

- Models with more predictors tended to do better than models with fewer predictors
- Models with 0 interaction terms tended to do better than those with 1 interaction
- Models that transformed key predictors tended to do better than those that did not

Retrospective

Trends:

- Models with more predictors tended to do better than models with fewer predictors

- Models with 0 interaction terms tended to do better than those with 1 interaction

- Models that transformed key predictors tended to do better than those that did not

- Performing log or root transformation moderately reduced test MSE

Retrospective

Trends:

- Models with more predictors tended to do better than models with fewer predictors

- Models with 0 interaction terms tended to do better than those with 1 interaction

- Models that transformed key predictors tended to do better than those that did not

- Performing log or root transformation moderately reduced test MSE

- The full model was near the front of the pack, while the simple model using just 1 predictor (`Gr_Liv_Area`) was at the back.

Retrospective

Trends:

- Models with more predictors tended to do better than models with fewer predictors

- Models with 0 interaction terms tended to do better than those with 1 interaction

- Models that transformed key predictors tended to do better than those that did not

- Performing log or root transformation moderately reduced test MSE

- The full model was near the front of the pack, while the simple model using just 1 predictor (`Gr_Liv_Area`) was at the back.

Further Investigation (Homework 5):

- Use `regsubsets` to assist with feature selection

- Use a cross-validation to assess and compare model performance