July 2020

TEXAS
The University of Texas at Austin

NBA - Score Prediction

Predict from dataset with all NBA games from 2004 Season to Feb. 2020

India Lindsay, Zeba Pathan, Suchit Das, Sujay Chebbi, Adam Hoard, Yiming Jin
MSBA, The University of Texas at Austin

Project Introduction

- Basketball
  - a game between two teams who score by tossing an inflated ball through a raised goal. The winner of the game scores more points.
- Our Task
  - Predict how many **points** a team will score based on other offensive metrics:
  - Assists
  - Rebounds
  - Field Goal Percentage
  - 3PT Percentage
  - Free Throw Percentage

Model Selection Process

- Split data into three groups:
  - Training (50%)
  - Validation (25%)
  - Test (25%)
- Fit models using training set, obtain test RMSE using validation set and compare to select optimal model
- Fit optimal model using train and validation sets combined
- Obtain out of sample error for optimal model by predicting y values using test set

Univariate Analysis

Implementing a linear model with one variable at a time for Points (PTS_home)

Model	Predictor	Variable Name	Model RMSE
Linear Model	Field Goal Percentage	FG_PCT_home	9.623471
Linear Model	Free Throw Percentage	FT_PCT_home	12.77778
Linear Model	3 point Percentage	FG3_PCT_home	11.74781
Linear Model	Assists	AST_home	10.4955
Linear Model	Rebounds	REB_home	12.79566

Multivariate Model 1: Stepwise Regression

- We try to get the response by implementing step regression model in 3 directions: Forwards, Backwards and Both
- We start by adding each variable until we get a optimal mix or sequentially delete each until reach an optimal mix of variables
- We calculate the error (RMSE) for each of the direction to find the best fit

Step Direction	Final Predictors	RMSE
Forwards	FG_PCT_home FT_PCT_home FG3_PCT_home AST_home REB_home	8.025279
Backwards		8.025279
Both		8.025279

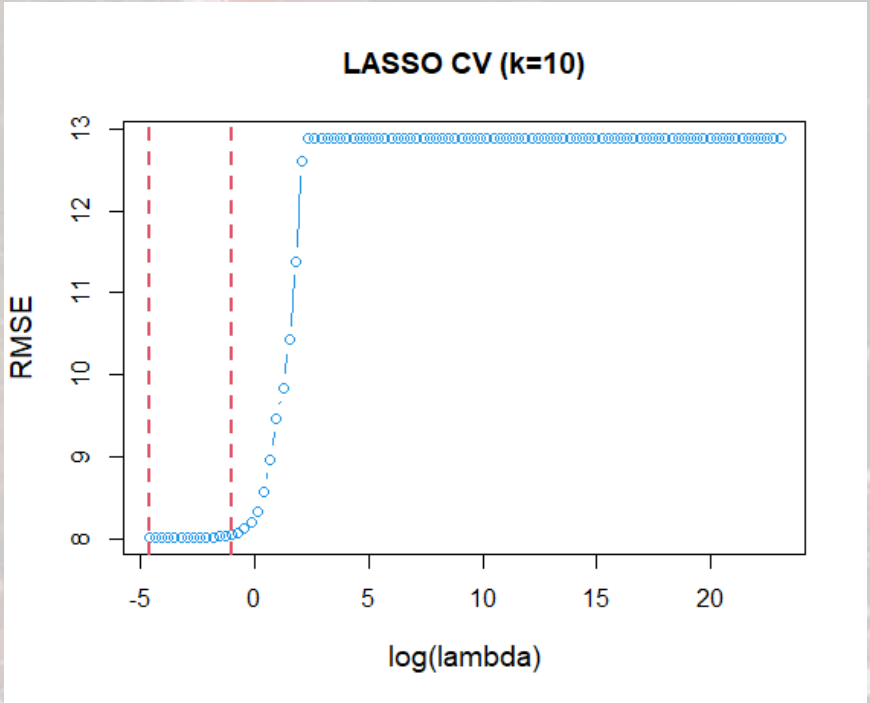
Model Selection

Model	Validation set RMSE
Univariate Regression (Field Goal %)	9.623471 (Lowest RMSE)
Stepwise Regression (All 3 Directions)	8.025279
Ridge Regression	8.025234
Lasso Regression	8.025041
PCR	8.025279
PLS	8.032477
Gradient Boosting Method (GBM)	16.14957

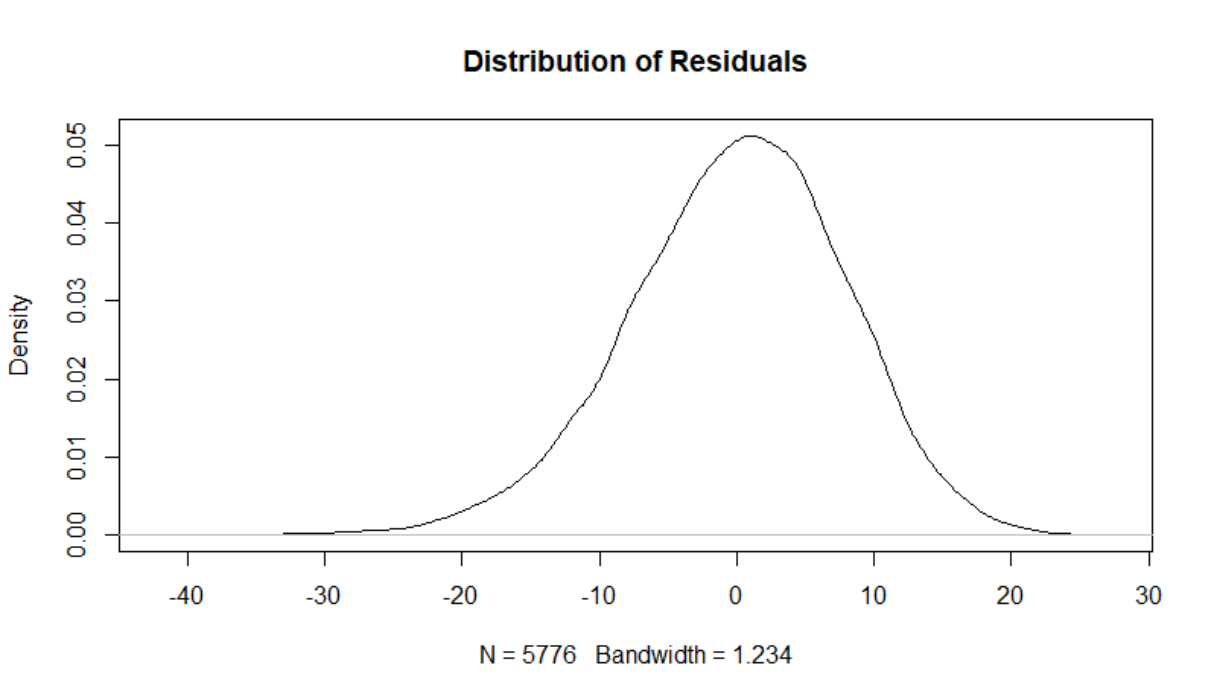
Lasso: Identifying Lambda

- Min. Lambda: 0.01
- Test Error: 7.831

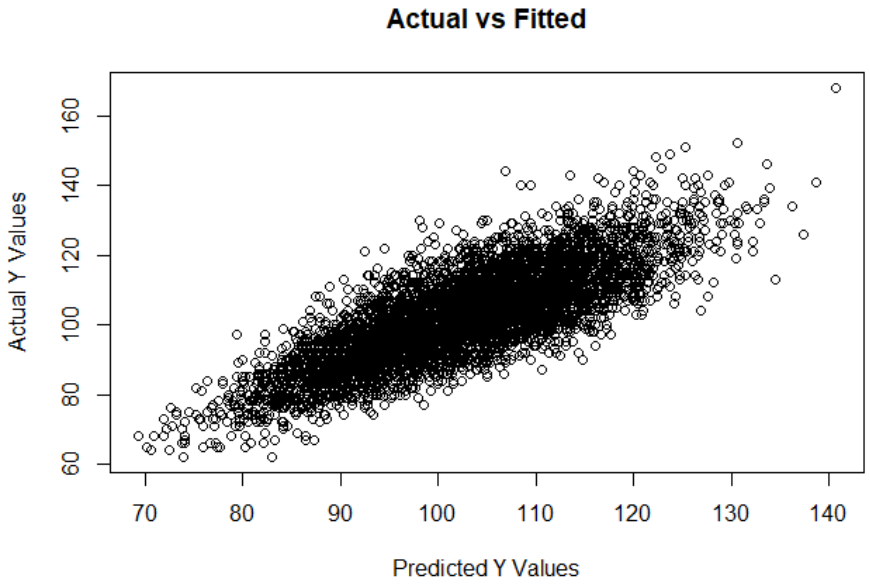
- Lambda 1 SE: 0.376
- Test Error: 7.877



Lasso Residuals Plots



Optimal Multilinear Formula

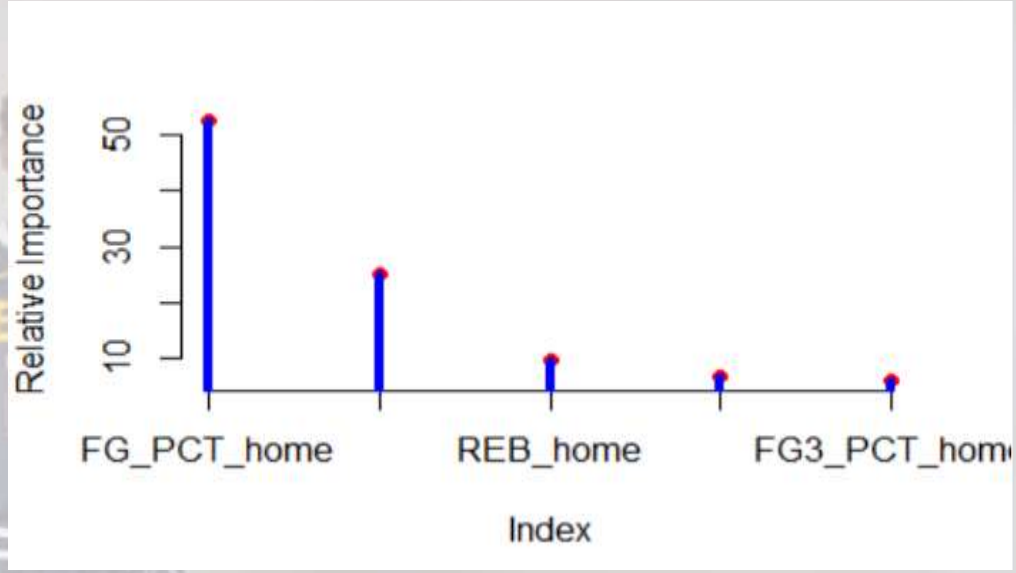


$$\hat{Y} = -10.702 + 120.552(FG\_PCT) + 22.557(FT\_PCT) + 13.551(FG3\_PCT) + 0.597(AST) + 0.5(REB)$$

Adjusted R-squared: 0.633

Boosting

Predictor	Relative Importance
FG_PCT_home	52.62
AST_home	25.12
REB_home	9.45
FT_PCT_home	6.75
FG3_PCT_home	6.03



Interpreting the Results

- Field Goal Percentage (FG_PCT) was determined to exert the **most** influence
- Stats (from last season):

	LAL	ORL	MIA	MIL
Field Goal Percentage	46%	44.2%	47%	44.7%
Free Throw Percentage	58.3%	77%	77.8%	72.3%
3 point Percentage	37.5%	34.1%	38.3%	37.2%
Assists	24	24	26	27.8
Rebounds	46	44.5	44.5	52.7
Predicted Score	**107**	**111**	**117**	**107**
Actual Score	**114**	**106**	**112**	**118**

Interpreting the Results

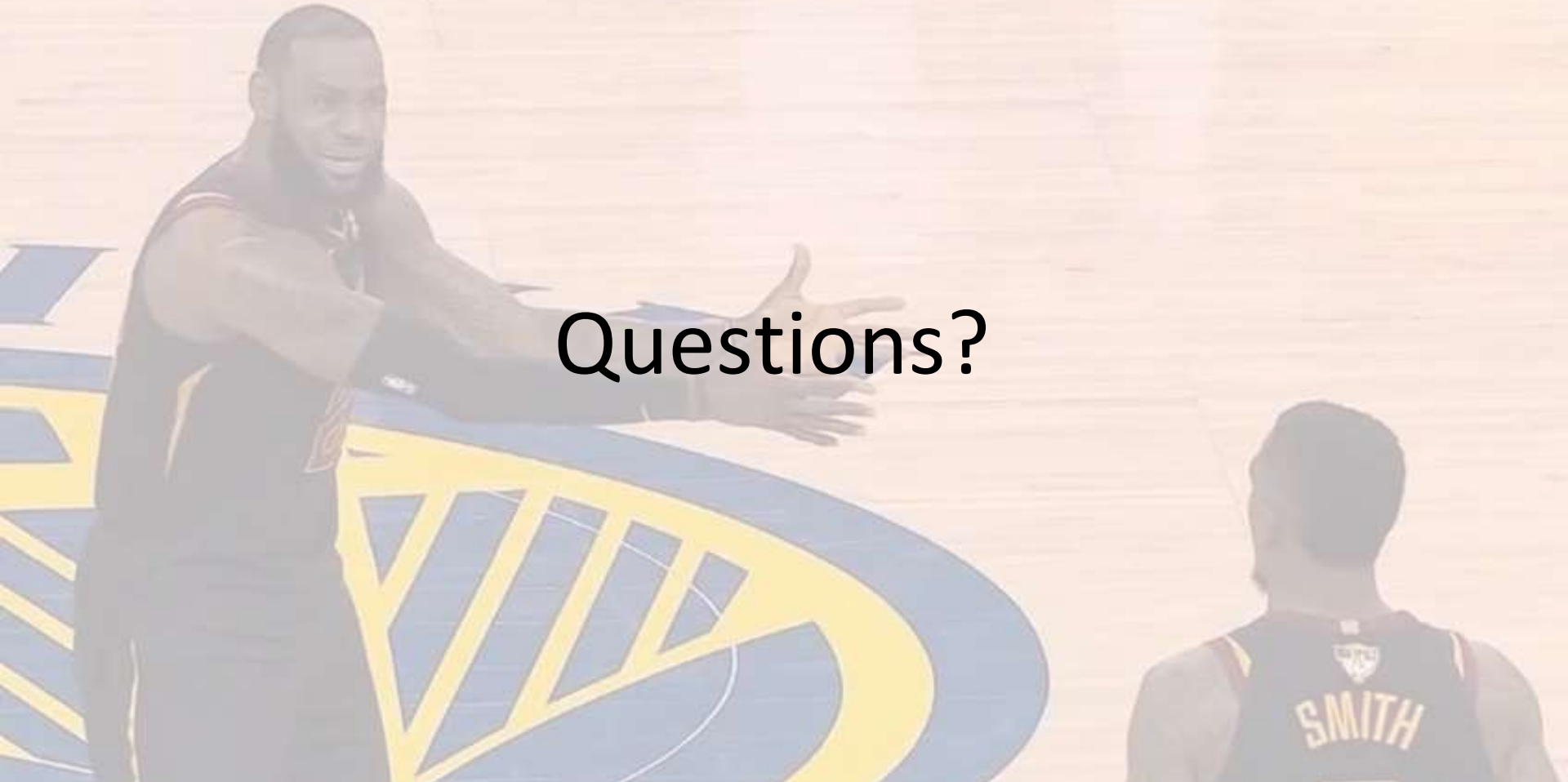
- Stats (from the games on July 25th):

	LAL	ORL	MIA	MIL
Field Goal Percentage	56.9%	34.3%	41.4%	44.4%
Free Throw Percentage	77.8%	84.1%	79.5%	78.9%
3 point Percentage	45%	17.9%	44.4%	35%
Assists	24	21	29	25
Rebounds	49	45	54	47
Predicted Score	**120**	**87**	**107**	**103**
Actual Score	**119**	**112**	**99**	**131**

Data Set Shortcomings

- Does not account for **volume of shots taken**
  - 50% FG_PCT could be 3/6 or 20/40
- Does not account for **style of offense played**
  - Assists directly correlate to points
  - Houston Rockets are near the bottom of the league in **assists** but one of the top in **points per game**

Questions?



Appendix: Univariate Analysis Fits