Predictive Modeling on 2017 NBA Playoffs

Qingyuan, Zhang Jihan, Wei Ruxue, Peng Zixuan, Guan Yue, Jin

April 26, 2017

1 Introduction

The Playoffs champion goes to...

2 Data

The data used for this project is mainly scrapped from basketball-reference.com and stats.nba.com. Among all kinds of measurement for a player, for a game, or for a season, we chose to use these key features as measurement for the performance of a team during one game played.

EFF Feature

The NBA publishes online all the basic statistics recorded officially by the league. Individual player efficiency is expressed there by a stat referred to as 'efficiency' and abbreviated EFF. It is derived by a simple formula below: NBA Efficiency recap = ((Points + Rebounds + Assists + Steals + Blocks) - ((Field goals attempts - Field goals made) + (Free throws attempts - Free throws made) + Turnovers)) To construct the feature, we found out for each game, the player lineup of the two teams and their minutes played in that game, then query their efficiency score of the corresponding game season, at last we average over players' EFF score for the two team.

By incorporating the players' efficiency score, we took into consideration the influence of change of player at each game within our model.

3 XGboost

We take the following steps to train and fit XGBoost model to our data:

1. Read in 2015 Regular Season data as training data and 2016 Regular Season data as test data
2. Cumulatively add all the data available before the date of each test data point as test features
3. Transform the data to the desired form for xgboost model
4. Select the best parameters (depth, eta) of the XGBoost model through cross validaation
5. Train the XGBoost model using the best parameters to the entire dataset and obtain predictions for this season

References