PREDICTING NBA GAME OUTCOMES

Jaak Uudmae juudmae@stanford.edu
CS229 Fall 2017-18

Overview:

Trying to predict game winners and scores of upcoming NBA games based on previous game scores.

Used three models:
1. SVM Classifier – predicting the winner only (trained per team)
2. Linear Regression – predicting the game score (general model)
3. Neural Network Regression – predicting the game score (general model)

All models achieved at around 62-65% accuracy of predicting game winners on the test set.

The last two models were both off by around 10-12 points for each game.

Data:

From stats.nba.com where NBA keeps track of all the stats related to games played in the league.

Using statistics associated with games – teams playing, home/away, scores.

Train Set: games from 2013-2016 (March).
Test Set: games for the rest of the 2016 season.

Future:

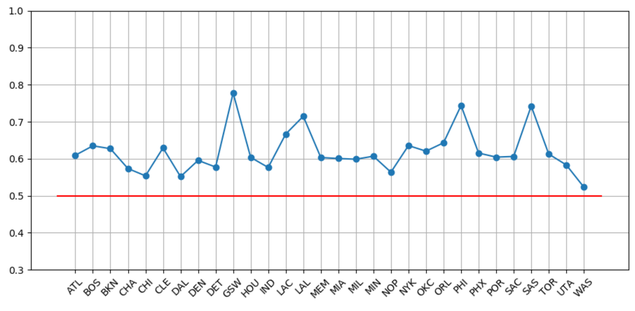
Given more time to train the neural network, the results could be improved.

Trying out different neural networks (width, number of hidden layers, number of neurons each layer) could yield better results.

References:

Keras Neural Network: https://keras.io
SKLearn: http://scikit-learn.org/stable/
Neural Network based on: https://machinelearningmastery.com/regression-tutorial-keras-deep-learning-library-python/
2013-2014 Team Scoring Analysis: https://www.sportingcharts.com/articles/nba/detailed-look-at-nba-team-by-team-consistency-during-the-2013-14-season.aspx

MODEL 1: SVM (sklearn.svm.SVC)



Features

[1,0,0,1,0…,1,20,10] size = 33
- 30 "1"s and "0"s indicating teams
- 1 indicator for home/away
- Wins/losses in the season so far

Description and Results

Predictor $\text{sgn}(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + \rho)$
K = radial basis function
Average test accuracy 62% of predicting the correct winner. Predicting teams that have been consistent since 2013 is more accurate

MODELS 2 and 3: Linear Regression (sklearn linear_model) and Neural Network Regression (Keras)

Results

The Neural Network Regression (NNR) predicted scores less accurately than Linear Regression (LR) – averaging about 2p worse for home and away teams. Interestingly, based on those scores, the NNR did better on predicting the winner of the game.
NOTE: 2013-2014 season the average standard dev for teams scoring was around 12.

	Model 1:SVM	Model 2: LR	Model 3: NNR
TEST: Average Accuracy	0.62	0.64	**0.65**
TEST: Home Score Distance	N/A	**9.82**	11.78
TEST: Away Score Distance	N/A	**9.95**	12.19
TRAIN: Average Accuracy	**0.71**	0.67	0.70
TRAIN: Home Score Distance	N/A	8.77	8.89
TRAIN: Away Score Distance	N/A	8.84	8.86

Features

[1,0,0,0,0…0,1,0,0] size = 60
2 "1"s and 58 "0"s indicating teams playing and home/away

Description

Linear regression solves ordinary least squares: $\min_w ||Xw - y||_2^2$
Neural Network is a 4-layer network.
Layers are defined as:
Input -> 60n -> 30n ->16n -> 2n Output
First two layers have 0.1 dropout to stop overfitting.
They all have "relu" activation.
Output layer has linear activation.
Using mean squared error.
Trained for 64000 epochs.

Discussion:

The goal was to see if it was possible to predict game scores based on just previous game scores. Both models achieved good results given that the team scores vary at around 12p on average (based on 2013-14 season data).

The results are affected by various factors:
1. Data size is too small. Right now there are around 4500 games from 2013-2016.
2. Team rosters/coaches change between seasons so the team's scoring is not very consistent from season to season.

Overall, I think the neural network has the potential to overtake LR predictions.