ECE/CS/ME 539 Introduction to Artificial Neural Networks

Project Progress Report

NBA Predictions

Team #4: Oliver Hammond (Graduate), Tyler Richards (Undergraduate),

Russel Spence (Undergraduate)

Date of submission:

**ABSTRACT**:

The goal of this project is to use machine learning models to predict the win / loss outcome of NBA games using teams’ past statistics. We developed logistic regression and decision tree models for this purpose; with our best accuracy score being 64% for logistic regression. Various parameters required tuning in our models, some of which were deciding which features to use, and deciding how to pull stats from a team’s previous games. We moved away from averaging stats from a team’s entire previous season to taking rolling averages for games leading up to the game we wanted to predict.

**Introduction**

The national basketball association (NBA) is comprised of 30 teams divided into the eastern and western conference. Each team roster has a max of 15 players. Each team plays 82 5v5 games during the regular season which typically spans from October to April. Following the regular season, the eight best teams from the eastern and western conference will play in the playoffs (also called the postseason). Playoff matches between two teams are best of seven, with single elimination. The teams who win the eastern and western conference finals play in the NBA championship.

The motivation for using machine learning algorithms to predict the result of NBA games stems from a few areas. Foremost, sports betting platforms such as DraftKings and individual sports betters use such algorithms for economic gain. The teams themselves may also gain an advantage from these algorithms, as variations of these algorithms can be used to determine which players from a team to play in a given game.

**Method**

Various features can be pulled from any given NBA game. For the sake of this project, we are utilizing a public basketball dataset hosted on kaggle.com by user Wyatt Walsh[1]. This particular dataset has over 60,000 basketball games from 1946-2020. Of the features contained in a given game, we are mostly concerned with field goals made, field goals attempted, 3-point field goals made, free throws made, free throws attempted, offensive rebounds, defensive rebounds, assists, steals, turnovers, personal fouls, and points for both the home and away team.

We’ve chosen to train both logistic regression and decision tree models to see which method produces a higher accuracy score. //TODO: FINISH METHOD

1. Method: Data, algorithm and program, platform, how to conduct experiments, what are the results, how to evaluate results (performance metrics). << copy from the proposal with additional new content if appropriate>>
2. Results: results should be presented in tabular/graph format. Screen dump of is not acceptable. Experiment conditions, parameters and related details should be provided in the text. Quality of the results (according to the performance metrics used) should be explained and comparison with the baseline results if available should be included here.
3. Discussions: Based on the results report in section 4 above and the objectives in section 2,
4. References: you should use standard reference format.