Sports Betting using Neural Networks

Aaron Williams

Department of Electrical and Computer Engineering, Duke University

Introduction

Can we use neural networks (NN) to make informed betting decisions? Can we make money consistently doing so? I hope to answer these questions by using data to make predictions on College Football scoring spreads.

- Create an adequate estimate of the scoring differential between two teams.
- Guess the spread using statistical + recruiting data.
- Use that information to estimate which way we should bet against the spread.
- Estimate confidence in these guess.
- Start accruing a small fortune.

Data

There are limitations in data, but sources were determined for betting information, game by game statistics, and recruiting.

Statistics
- *Rush Yards per Game*
- *Rush Attempts per Game*
- *Pass Yards per Game*
- *Pass Attempts per Game*
- *Pass Comps per Game*
- *Fumbles per Game*
- *Interceptions per Game*
- *Final Score*

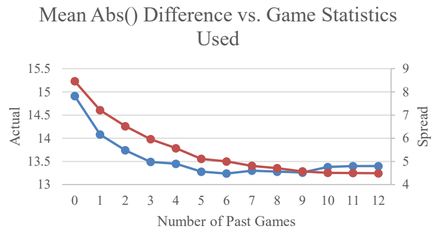
Recruiting
- *Number of Recruits*
- *Number of 5-stars*
- *Number of 4-stars*
- *Number of 3-stars*
- *Average Ranking*
- *Overall Score (by Rivals)*

Betting Information
- *Spread*
- *Over/Under*

- Training Data includes 2003 – 2014 (7684 datapoints)
- Test Data Includes 2015 – 2017 (1838 datapoints)
- Each value (besides spread) was 0-normed with respect to the training data.
- All data points include the opposing team's statistics, and the average statistics of the opponent's last 10 games.
- Also Game/(Total Games in Year) was included.

Hyper Parameterization

- Need correct amount of statistical and recruiting data per datapoint.
- Limited by amount of proceeding data.
- Tests done on two fully connected dense layers, no activation function.
- All subsequent tests were done with 10 past games and 4 past years.



Predictions and Convolutions

Actual Lower Bound: 16.83

Spread versus Actual: 12.55

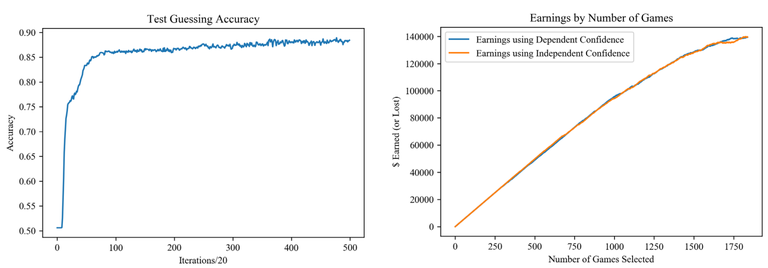
2-Layer FC NN	Betting Spread	Actual Difference
Prediction on Spread	Avg ± of 4.52	Avg ± of 13.25
Prediction on Actual	Avg ± of 5.24	Avg ± of 13.38

With Convolutions	Betting Spread	Actual Difference
Prediction on Spread	Avg ± of 4.55	Avg ± of 13.24
Prediction on Actual	Avg ± of 4.92	Avg ± of 13.25

- Lower Bound is the average absolute difference against mean.
- The spread performs better than both NNs by a significant margin. Both networks predict the spread better than Actual.
- No significant advantages for Convolution.
- Additional importance measures show convolution and flat vectors equally retain information.

False Results

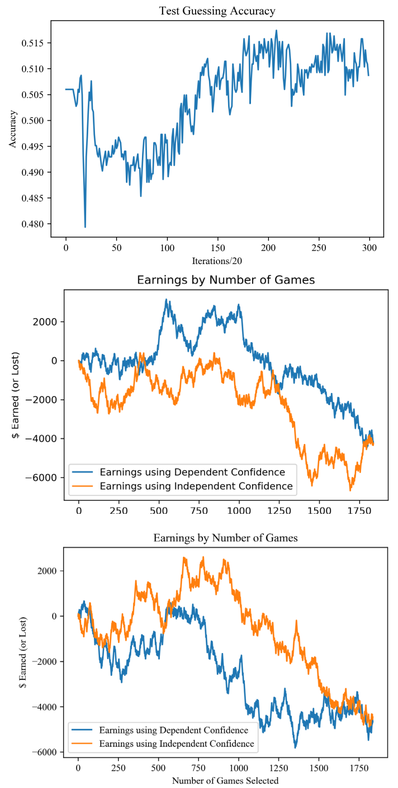
Included for fun. What if we could see the future and knew the game scores, but only needed to guess the betting spread?



We'd want to invest a lot more than $100 in each game.

Results

- If we bet in all games, accuracy needs to be just 52.4% to account for a $100 return on a pay of $110.
- Accuracy is typically around 51.5% and can be highly variable through training.
- Dependent is distance from 0.5 in determiner. Independent is a NN which takes determiner as input and all other data.
- Neither performs better consistently. There is some trend towards better performance earlier on, but not enough for reliable betting.
- Losses for betting on all games average around $4000.



Conclusions/Analysis

- **Not enough Data**: Could of used injury reports, off-field drama, coaching changes, or even weather to indicate the outcome of a game.
- **Training**: It's easy to find relevant information for an upcoming game, but how do you apply that to a trained NN? How do you train a NN with data not found historically?
- **RNN**: Applying an RNN to this problem might have merit. Retroactive Adjustments would be the main difficulty.
- **Over/Under**: It might have been better to try and solve over/under instead of the spread. Statistics are generally more relevant to total scoring.
- **Closing**: I was hoping this project would be a little more interesting, either as a result of the network architecture or the independent confidence metric. Images next time.

References:

Chernoff, A. (2017, August 30). How To Create And Use NFL Power Ratings To Beat The Point Spread. Retrieved from https://medium.com/@adamchernoff/how-to-create-and-use-nfl-power-ratings-to-beat-the-point-spread-3fa4c3ecdc22

A machine learning framework for sport result prediction. (2017, September 19). Retrieved from https://www.sciencedirect.com/science/article/pii/S2210832717301485

Thomas, Eduorard. "Beat the Bookmakers with Machine Learning (Tennis)." RSNA Pneumonia Detection Challenge | Kaggle, www.kaggle.com/edouardthomas/beat-the-bookmakers-with-machine-learning-tennis.

http://sports.snoozle.net/search/fbs/index.jsp

https://n.rivals.com/team_rankings/2019/all-teams/football

https://www.masseyratings.com/

https://github.com/mattp88/ncaa/blob/master/teams.cs