Predicting the Total Number of Points Scored in NFL Games

Max Flores (mflores7@stanford.edu), Ajay Sohmshetty (ajay14@stanford.edu)

CS 229 Fall 2014

1 Introduction

Predicting the outcome of National Football League games is a topic of great interest to both fans and gamblers. The vast majority of efforts into this topic attempt to predict the winner of each game or the margin of victory. As a result of these efforts, the Las Vegas "odds" and "spread" for each game are very tough to beat. We believe that the "over/under," a betting line on the total number of points, is more beatable in part because it is less tried. Furthermore, a prediction of the total points scored is more relevant than a prediction of the winner to the fantasy football industry (which has actually eclipsed the NFL in revenue[3]) because it is directly correlated with the performance of fantasy players. Our target success rate versus the Vegas "over/under" line is 52.4%, the break-even point versus typical handicapped odds.[2]

2 Dataset[1]

Our data is statistics on NFL games from the 2003 to the 2013 season. Each row in our data represents a game played, and each column a statistic of that game. Our data has various statistics for each game, only a subset of which are of importance for us. Namely, we are interested in the following: date of game, home team, away team, home team score, away team score, home team rushing yards, away team rushing yards, home team third down percentage, away team third down percentage, home team passing yards attempted, home team passing yards completed, away team passing yards attempted, away team passing yards completed, home team fumbles, and away team fumbles.

We preprocessed the data (computed pass completion percentage, which is given by dividing the number of pass completions by the number of pass attempts; and total score, by taking the sum of the home team score and the away team score.

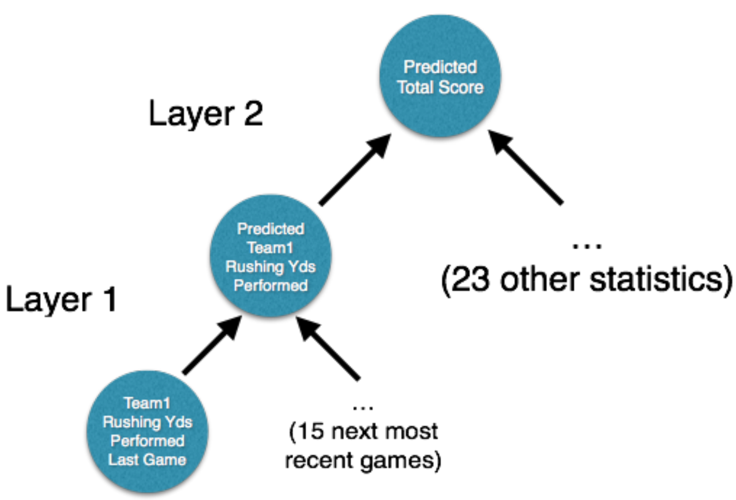
3 Features

We chose 6 key stat categories per game that are not highly correlated: 3rd down percentage, rushing yards, passing completion percentage, passing yards, pass interceptions, and fumbles. Our data contains the past 11 seasons worth of games (2789 games). For each game and for each stat category, we take 4 features: home team performed, home team allowed, away team performed, away team allowed.

For example, for the "rushing yards" stat category, we have rushing yards the home team performed, rushing yards the home team allowed, rushing yards the away team performed, and the rushing yards the away team allowed. For interceptions and fumbles, the terms "performed" and "allowed" are counterintuitive. To clarify, performed refers to interceptions thrown or fumbles made and allowed refers to interceptions thrown or fumbles made by the other team.

6 (stat categories) x 2 (each team) x 2 (allowed/performed) + 1 (intercept term) = 25 total features per game

4 Models



We ran a two-layer neural network.

For our first layer, in order to predict a particular statistic for any team in any game, we look at the most recent 16 games by that team (one NFL season). We run a linear regression on the same statistic for that team looking at the past 16 values of that statistic in order to predict the next one. We ran 12 linear regressions on our train data to produce 12 weights vectors (of size 16) for each of the 12 feature types (6 stat categories x 2 allowed/performed).

We used the aforementioned weights to compute predicted statistics of each game of both our train and test data. In our second layer, we then used these predicted statistics for each game as our input, and output whether the total points scored would be "over" or "under" the Las Vegas line for the game.

We made this "over" or "under" prediction using three different methods. Our first was least squares linear regression with stochastic gradient descent. We computed the predicted score for each game, and then appropriately chose "over" or "under" relative to the Vegas line.

Our second method was logistic regression with stochastic gradient descent. In order to fit this model, we appended the Vegas line as an extra feature (26 features). We then directly predicted the "over" or "under."

Similarly, we also used an SVM to directly predict "over" or "under" for each game.

In addition to these methods, we attempted an alternative strategy using linear regression to further improve our results: we only made a prediction on the data only if we could ensure a certain degree of confidence on the result. If the magnitude of the difference between the Vegas line and our predicted score exceeded our confidence, we made a prediction. We ran this model for a confidence of 3 points and a confidence of 5 points.

5 Results

Average Weights For Each Past Week For Each Stat Type (Layer 1)

	1	2	3	4	5	6	7	8
3rd Down % Performed	0.071	0.058	0.066	0.064	0.069	0.070	0.055	0.067
3rd Down % Allowed	0.055	0.082	0.067	0.067	0.060	0.062	0.055	0.060
Rushing Yards Performed	0.072	0.094	0.092	0.065	0.085	0.072	0.047	0.061
Rushing Yards Allowed	0.070	0.094	0.075	0.079	0.077	0.049	0.048	0.047
Pass Completion % Performed	0.070	0.074	0.073	0.070	0.068	0.059	0.061	0.061
Pass Completion % Allowed	0.059	0.071	0.066	0.069	0.061	0.065	0.061	0.063
Passing Yards Performed	0.106	0.088	0.080	0.077	0.077	0.077	0.076	0.035
Passing Yards Allowed	0.075	0.073	0.051	0.082	0.045	0.078	0.038	0.047
Passing Interceptions Performed	0.058	0.029	0.079	0.093	0.053	0.046	0.075	0.042
Passing Interceptions Allowed	0.044	0.057	0.084	0.080	0.065	0.061	0.054	0.070
Fumbles Performed	0.059	0.070	0.058	0.060	0.069	0.065	0.060	0.060
Fumbles Allowed	0.059	0.061	0.067	0.064	0.065	0.057	0.067	0.062
Average	*0.067*	*0.071*	*0.071*	*0.072*	*0.066*	*0.063*	*0.058*	*0.056*
	9	10	11	12	13	14	15	16
3rd Down % Performed	0.062	0.070	0.059	0.065	0.065	0.054	0.052	0.055
3rd Down % Allowed	0.062	0.054	0.068	0.059	0.069	0.056	0.059	0.067
Rushing Yards Performed	0.045	0.077	0.054	0.034	0.033	0.068	0.053	0.047
Rushing Yards Allowed	0.073	0.057	0.032	0.058	0.070	0.061	0.073	0.037
Pass Completion % Performed	0.058	0.061	0.056	0.056	0.061	0.053	0.063	0.057
Pass Completion % Allowed	0.067	0.062	0.059	0.058	0.058	0.064	0.057	0.061
Passing Yards Performed	0.047	0.056	0.058	0.048	0.033	0.056	0.039	0.047
Passing Yards Allowed	0.058	0.072	0.074	0.068	0.061	0.072	0.058	0.047
Passing Interceptions Performed	0.082	0.042	0.062	0.077	0.091	0.058	0.055	0.059
Passing Interceptions Allowed	0.083	0.089	0.079	0.021	0.047	0.066	0.040	0.060
Fumbles Performed	0.067	0.059	0.067	0.060	0.056	0.061	0.062	0.067
Fumbles Allowed	0.062	0.065	0.054	0.060	0.063	0.068	0.065	0.061
Average	*0.064*	*0.064*	*0.060*	*0.055*	*0.059*	*0.061*	*0.056*	*0.055*

Average Success Rates For Each Model*

	Train Success Rate (# examples)	Test Success Rate (# examples)
Linear	52.14% (1952)	51.53% (837)
Logistic	52.05% (1952)	49.35% (837)
SVM	50.26% (1952)	49.56% (837)
Linear (confidence > 3)	52.84% (avg of 754)	52.08% (avg of 320)
Linear (confidence > 5)	53.46% (avg of 304)	52.78% (avg of 132)

* As explained in the next section, any success rate > 50% is a good result

6 Discussion

We ran the entire neural network repeatedly using a random 70% of the data for training each time to get the average results in the previous section.

First we will discuss the results from layer 1 of the neural network. The average weights for the past 16 games tended to decrease as the number of games ago increases. This makes intuitive sense that when making a prediction it matters more what happened on the most recent game than 16 games ago. This decreasing trend was most present for the stat types rushing yards performed, rushing yards allowed, passing yards performed, and passing yards allowed. It was least present for the stat types interceptions performed, interceptions allowed, fumbles performed, and fumbles allowed all of which had more random weights.

Layer 2 of the neural network predicted “over” or “under” on each game versus the line using several different models. Before discussing these results it is necessary to state that beating the Las Vegas line is very difficult. While random guessing will result in a 50% success rate, very sophisticated models will struggle to do much better. A model needs a reliable success rate of over 52.4% to be profitable against typical handicapped odds offered.

It is immediately clear from the results that neither logistic regression nor the SVM were effective. Both had a test success rate slightly worse than random guessing (50%). Linear regression, which doesn’t predict “over” or “under” directly, had much better results.

The linear regression had a decent test success rate of 51.53%. Despite being an improvement from random guessing, this rate is below our target of 52.4%. However, if we run linear regression and only make a prediction if our output differs from the line by at least 3 or at least 5 we get better results (for the fraction of examples predicted on). A confidence of 3 yields a test success rate of 52.08% and a confidence of 5 a test success rate of

52.78%. The latter exceeds the target success rate.

7 Conclusions

Passing yards and rushing yards are best predicted weighting recent data more heavily whereas interceptions and fumbles require more evenly weighted data. Furthermore, linear regression using predicted stats is much more effective at beating the "over/under" than both logistic regression and an SVM using the same stats. By predicting on only the ~15% of games we are most confident on, we found that we can achieve a profitable model versus the Vegas "over/under" line with handicapped odds.

8 Future Work

For further work, the first thing we would do is relook at our feature set. A lot of NFL stat categories are very correlated and because of this we chose the least correlated and most relevant ones using just our football knowledge. Various feature selection algorithms could help choose better stat categories, such as forward search.

While feature selection algorithms on team stats may make an impact, we believe that to get much better results data on individual players needs to be used. Looking at team statistics only works so well when the players in the games are constantly changing. Specifically, in any given game it is critical to know whether certain key players are playing. Incorporating player data will likely require an entirely new model.

9 References

[1] http://www.repole.com/sun4cast/data.html
[2] http://www.thesportsgeek.com/sports-betting/math/
[3] http://www.forbes.com/sites/briangoff/2013/08/20/the-70-billion-fantasy-football-market/