Predicting Margin of Victory in NFL Games: Machine Learning vs. the Las Vegas Line

Jim Warner

December 17, 2010

Abstract

In this study we describe efforts to use machine learning to out-perform the expert Las Vegas line-makers at predicting the outcome of NFL football games. The statistical model we employ for inference is the Gaussian process, a powerful tool for supervised learning applications. With predictions for the margin of victory and associated confidence intervals from the Gaussian process model, we propose a simple framework which recommends a bet on a given game when it is deemed statistically favorable. The training dataset we consider in this study includes a wide variety of offensive and defensive NFL statistics from about 2,000 games between 2000 and 2009. We also explore the impact of including additional novel features previously unstudied: the temperature difference between competing team's cities and a team's computed strength according to [10]. We show that our predictions for margin of victory result in an error just 2% higher than that of the Las Vegas line and that we can successfully pick the game winner over 64% of the time. The bet-recommendation scheme we propose is shown to provide a win rate just under 51% but falls short of the mark of 52.4% needed to break even in the NFL gambling system.

1 Introduction

NFL football is arguably the most popular sport to bet on in the United States. It is said that gamblers bet nearly $\$1B$ per year on football games in Nevada alone [9]. Critical to the NFL gambling system is what is known as the *(Las Vegas) line* or *point spread*. The point spread is a handicap assigned to one of the teams, for betting purposes only, that is designed to give each team an equal chance of winning in the eyes of the bettor. For example, if team A is the clear favorite over team B, the bookkeepers (those in charge of the betting process) will create the point spread for the game to reflect this; say, team A is the ten point favorite. Now, a gambler will not win a bet simply in the event that team A is victorious, but only if they do so by a margin of victory larger than ten. Likewise, a bet on team B will pay off not only if they win but also if they lose by nine points or fewer.

The purpose of this study is to explore the use of a data-driven machine learning framework to predict the margin of victory in a matchup between any two given NFL football teams. With an accurate prediction of game winner and margin of victory, one could hypothetically compare this predicted value with the point spread designated to the game and proceed to make a statistically-favorable bet. The potential for success stems from the fact that the prediction supplied by a machine learning algorithm will be based solely off data and outcomes from previous games whereas the point spread is not necessarily an unbiased predictor of the game outcome. As stated by Vergin and Sosik, *... the line can be viewed as the best forecast of bettor behavior, rather than the best forecast of game outcome. Given this objective, it is conceivable that there might be biases in the line of sufficient size to make some technical betting strategies profitable*[8]. So although the point spread is designed by the Las Vegas bookkeepers to make consistently winning bets maximally difficult, the possibility of exploiting an existing bias in the line motivates the development of a predictive model for NFL games.

Despite being largely overshadowed by forecasting efforts in financial markets, there has been a modest amount of work done in the statistics & machine learning communities with regards to predictions for sports markets such as the NFL. A substantial amount of the work done relating to the NFL football betting market is theoretical in nature, debating the so-called *efficiency* of the market. A term that is commonly used in financial settings, an efficient market is one which is not predictable but random, such that no planned approach to betting or investing can be successful in the long term. These theoretical works report conflicting conclusions, with researchers in [2], [4], and [5] suggesting the NFL betting market is indeed an efficient one. While in [1], [3], [6], and [7] degrees of inefficiency in the market are shown.

Although there is a lack of agreement about the possible existence of successful betting strategies on NFL games, there have been several statistical models developed for this and similar purposes. One class of approaches deal with ranking or assigning relative strengths to each team. In [10], a number of general ranking schemes are introduced and subsequently applied to the problem of ranking the best college football teams. Rating methods are developed in [12] for soccer teams and in [13] for NFL teams, but predictions for game winners based on these ratings are not described.

There have been several statistical models developed to make predictions on NFL games that vary in both sophistication and success. In [15], the authors make predictions using simple probit regressions based on power scores published in *The New York Times*, but report that the official point spreads are in general more accurate. Mixed linear models based on home-field advantage and team performance are used in [16] resulting in an error that is only slightly higher than the bookmaker's predictions it is compared with. In [11], a state-space model is created in a fully-Bayesian context to model strengths of teams and make subsequent predictions of final scores to games. Accuracy of this model is reported to be about as high as the official point spread. Finally, a more successful approach is employed in [14] where an accuracy of up to 54% was reported for using a logistic regression classifier to predict the winner of NFL games while taking into account the point spread.

The related work in this area confirms the prowess of the bookmakers in Las Vegas in that it is hard to find evidence of a technical betting approach that consistently outperforms these expert line-makers. In this light, a baseline measure of success in this study is to provide predictions of the margin of victory in NFL games that is on average closer to the true score differential than the official point spread. However, the bettor is at a further disadvantage to the line-makers in the NFL betting system due to what is known as the *eleven for ten rule*; that is, one must put down $11 to win $10 on any given bet, providing the bookmakers with a commission known as the *vigorish.* Due to the vigorish, a bettor with a 50% success rate will actually lose money and instead needs to win 52.4% of their bets in order to break even. Therefore, a more geniune goal for this study is the development of predictive framework that when used to make informed decisions in the NFL betting market, results in a win rate of 52.4% or better.

In an effort to reach the accuracy goals mentioned above, this approach utilizes the Gaussian process model, which has emerged as a serious competitor for real supervised learning applications in the past decade [17]. Gaussian processes provide a powerful tool for inference with computational tractability as well as a principled manner in which to quantify uncertainty. The ability to generate confidence measures together with predictions seems to lend itself naturally to a betting scenario where one looks to balance the risk of placing a bet on a game with their relative certainty in its outcome. Indeed, we shall seek a scheme in which bets are placed on a game only when a specific confidence level is met in order to achieve an adequately high win rate. To the author's knowledge, this study is the first which utilizes the Gaussian process model in the area of sports forecasting.

The training dataset considered in this study includes a wide variety of offensive and defensive statistics for over 1,000 games from the 2000-2007 NFL seasons while 2008 and 2009 seasonal data is reserved for final testing. We also look to investigate the benefit of including novel features not considered in previous works. To this end, we explore the impact of home-field advantage on game outcomes by factoring in the temperature difference between the home and away cities with the belief that a visiting team's performance can be negatively affected by playing in a climate significantly different from their own[9]. Additionally, we seek improved accuracy in our approach by coupling it with a ranking system for sports teams. We compute the strength of the home and away teams going into each game according to [10] as an additional feature to the learning algorithms. It is expected that this data will be more benefitial than standard winning percentages in making predictions since a team's rank takes into account both the strength of its previous opponents as well as the outcomes of the previous games.

The remainder of the paper is laid out as follows: section 2 describes the process of data collection, provides a complete list of the features considered for training, and gives the formulation of the ranking system considered. Section 3 provides an overview of learning with Gaussian processes. Section 4 describes the process of feature selection, shows the impact of the novel features considered in this study, and compares results for each algorithm versus the Las Vegas lines for predictions in games from the 2009 and 2010 NFL seasons. Section 5 concludes the paper by discussing the effectiveness of this approach and suggestions for future work in this area.

2 Data Aquisition

In this section we provide an overview of the dataset utilized for margin of victory predictions and the process for collecting this data. NFL games from 8 seasons between 2000-2007 are used as training examples while games from the 2008 and 2009 seasons are reserved for final testing. We assume that individual seasons are mainly independent from one another and so a prediction for a particular game is based solely off data from the current season. In this light, games from the first four weeks of each season are excluded from training and testing due to lack of data. Note also that the dataset includes only matchups from the regular season (preseason and playoff games are excluded). For each game, we consider a variety of standard NFL statistics as well as two additional novel features to be described in the following subsections. In total, there are 1544 games in the training set and 390 games for final testing with 47 features considered. A list of the full set of features can be viewed in Figure 1.

Total Feature Set	
Winning %	
Points per game scored	
Points per game allowed	
Total yards per game gained	
Total yards per game allowed	
Rush yards per game gained	(H, A, H/S, A/S)
Rush yards per game allowed	
Pass yards per game gained	
Pass yards per game allowed	
Turnovers taken per game	
Turnovers given up per game	
Computed Strength	(H, A)
Temperature difference	

Figure 1: Full list of the features considered for training in this study. The "H" and "A" indicates there is data for both the home and away teams for that feature. The "S" indicates that there is streak (4-game moving average) data for that feature. Taking into account these variations, there are a total of 47 features.

2.1 NFL Seasonal Statistics

Although there exists plentiful sources of NFL statistics online, the availability of nicely formatted, downloadable data files is limited. Given the large amount of data needed for this project (scores & statistics for 32 teams playing 16 games per season from 10 seasons), custom MATLAB programs with the ability to go on the web and collect this data efficiently and accurately are employed. The website that is mainly utilized for collecting seasonal statistics is www.pro-football-reference.com, which is found to be the most extensive source of data.

Two main types of functions are used to obtain the NFL statistics for training: the first type collect raw NFL data online while the second process the raw data into individual files containing a particular statistic from each game in a given year. The MATLAB functions `urlread` and `urlwrite`, which input the url of a website and return a large character array containing the HTML source code for that site, form the basis of the raw data collection code. Additional code needed to be written to parse through the HTML, extract the useful data, and print it to a neatly csv-formatted text file. The second type of function to refine this raw data is necessary to simplify the process of loading input features to the machine learning algorithm. The function identifies the home and away team from every game in the raw data files and locates and organizes a particular statistic into two columns for each team, subsequently storing the columns individually in text files.

A total of eleven NFL statistics from each game are collected (see Figure 1) for both the home and away teams, producing 22 features for which to train the Gaussian process for predictions. While these features in general represent an average value over the course of the season up to that game, an additional moving average for every statistic is computed considering only the four previous games. This provides an additional 22 features that can take into account the known notion of "hot" or "cold" streaks in football where a team performs well above or below average in a particular stretch of games. Of course, a streak of any amount can be considered but a four game streak is used here as it represents about a month of game play or exactly one quarter of an NFL season which is deemed adequately long. Also, an average computed over a longer stretch would require omitting more games from the start of a season in the dataset.

2.2 Novel Features

A point of emphasis in this study is to expand the dataset beyond ordinary NFL statistics to include a couple of novel features and to explore their impact, if any, on making accurate predictions. One of these features arises from the desire to better account for the advantage a team has when playing in its home stadium. It is known that home teams generally win more often than visiting teams due to travel and crowd factors and it has been reported that this benefit can be larger than recognized by line-makers in some cases [8], [9]. In [9], the author focusses specifically on the effects of the difference in climate between the home and away team's cities, and shows statistically that this can have a significant negative impact on the visiting team. In this light, we shall consider the difference in average temperature between the home and away team's cities from the week of the year they played as an additional feature in the dataset. Note that since the output we are inferring with the Gaussian process is defined as the home team's score minus the away team's score, we are implicitly taking into account which team is playing in their home stadium. It is hoped that including the temperature difference in the dataset during training will have the effect of increasing or decreasing the impact of home-field advantage in making predictions.

Similar to the collection of NFL seasonal data, MATLAB is utilized to acquire the needed temperature data for every game in the ten seasons considered. Essentially, a program is written to loop over the raw NFL game data, take the date and two teams playing in the current game, find the cities for the respective teams, open a website containing temperature data for the given date and cities (using www.wunderground.com - a website with historical weather data), compute the difference in the average weekly temperature for the cities, and write this value to file. Since seven teams in the NFL play in climate-controlled dome stadiums, we assume games played in these cities lack the climate-induced home advantange. Therefore, the code is set up to identify such teams and enter a value of 0 for the temperature difference when they are playing at home.

As mentioned previously, some work has been done in the statistics community to develop ranking and strength rating systems for sports teams. A second novel feature to be investigated will entail adopting one of these rating systems [10], computing the current rank or strength for each team prior to every game, and using this as an additional feature for the Gaussian process regression algorithm. As the computed strength of a team depends on the outcomes of its previous games and strength of previous opponents, it is expected that including it in the dataset in addition to or in place of a team's winning percentage could yield higher accuracy. The formulation for the ranking scheme we adopt is provided in the following subsection.

2.2.1 Ranking System Formulation

With one of the ranking systems developed in [10], we look to assign a score to each team in the NFL based on their interactions with opponents throughout the course of the season. To begin the formulation, we suppose we have a rank vector $\vec{r}$ where each r_j represents a positive-valued strength of the j^{th} team in the NFL. Now we assume that a team's score s is a linear combination of the strengths of its opponents, where the weighting coefficients are dependent on the outcome of the games. In other words, we can write the score for team i as:

$$s_i = \frac{1}{n_i} \sum_{j=1}^{N} a_{ij} r_j \tag{1}$$

where a_{ij} is a nonnegative number depending on the outcome between team i and j, $N = 32$ is the number of teams in the NFL, and n_i is the number of games played by team i at the point in the season when the rank is computed. The ranking scheme transforms into an eigenvalue problem by proposing that a team's strength should be proportional to its score:

$$A\vec{r} = \lambda \vec{r} \tag{2}$$

where $A_{ij} = a_{ij}/n_i$. Hence, solving equation (2) for the eigenvector $\vec{r}$ provides us with the strengths for each team in the league.

To complete the formulation, we must specify the values for each a_{ij}. In the simplest version of this ranking scheme, one could let a_{ij} be 1 if team i won the game and zero if they lost. However, it makes more sense to distribute the 1 point between two competing teams based off the final score to the game. In one approach, if team i scores S_{ij} points while team j scores S_{ji} points in their matchup, we could let $a_{ij} = (S_{ij} + 1)/(S_{ij} + S_{ji} + 2)$, where the 1 and 2 in the numerator and denominator are present to prevent the winner from taking all the credit in a shutout. The specific approach we adopt from [10], however, makes one further improvement by distributing the point in a nonlinear fashion to prevent a team's rank from climbing from simply running up the score on an opponent. In this case, we assign the values of a_{ij} as:

$$a_{ij} = h\left(\frac{S_{ij} + 1}{S_{ij} + S_{ji} + 2}\right) \tag{3}$$

where

$$h(x) = \frac{1}{2} + \frac{1}{2} sgn(x - \frac{1}{2})\sqrt{|2x - 1|} \tag{4}$$

Viewing equation (4), the function $h(x)$ has the properties that $h(\frac{1}{2}) = \frac{1}{2}$, and away from $x = \frac{1}{2}$, h approaches 0 or 1 rapidly so that to improve a team's strength rating, it is important to win a given matchup but not as important to run up the score.

3 Gaussian Processes

The Gaussian process model provides an effective approach to supervised learning problems. We adopt such a probabilistic techique over deterministic methods as it allows for a straightforward means of quantifying the uncertainty in predictions. This will enable a betting framework to be developed in which a bet will only be recommended when the confidence in a predicted game outcome reaches a certain threshold. This study will also be the first to employ a Gaussian process for a learning application in the area of sports forecasting, so it will be interesting to see its performance relative to previous approaches.

Roughly speaking, a Gaussian process describes a distribution over functions and is the generalization of the typical multivariate Gaussian distribution to infinite dimensions [17]. Just as a Gaussian distribution is fully specified by a mean vector and covariance matrix, the Gaussian process is fully specified by a mean function and covariance function. If we define the mean function $m(\vec{x})$ and covariance function $k(\vec{x}, \vec{x}')$ of a real process $f(\vec{x})$ as

$$m(\vec{x}) = \mathbb{E}[f(\vec{x})] \tag{5}$$

$$k(\vec{x}, \vec{x}') = \mathbb{E}[(f(\vec{x}) - m(\vec{x}))(f(\vec{x}') - m(\vec{x}'))] \tag{6}$$

we can then write the Gaussian process as:

$$f(\vec{x}) \sim GP(m(\vec{x}), k(\vec{x}, \vec{x}')) \tag{7}$$

Without loss of generality we shall consider zero mean Gaussian processes for notational simplicity, which in practice is equivalent to subtracting the mean from training outputs prior to learning. The covariance function we use is the squared exponential, given as:

$$k(\vec{x}, \vec{x}') = \sigma_f^2 \exp\left(-\frac{1}{2}(\vec{x} - \vec{x}')^T M (\vec{x} - \vec{x}')\right) \tag{8}$$

where $M \in \mathbb{R}^{dxd} = diag(\vec{l})$, d is the dimension of the input vector, and $\vec{l} = [l_1, ..., l_d]$. The parameters $l_1, ..., l_d$ represent the characteristic length scales in the problem or how far you need to move along a particular axis in input space for function values to become uncorrelated.

In our approach we seek to fit a Gaussian process to the unknown underlying function that maps from the NFL dataset to the margin of victory of a given matchup. Hence, we suppose that our margin of victory output values are noisy versions of this unknown function, $y = f(x) + \epsilon$, where we assume $\epsilon \sim N(0, \sigma_n^2)$ is i.i.d. Given this model, we would like to consider new test data x^* and make predictions on the corresponding value of the unknown output function $f^*(\vec{x}^*)$. The key to inference with the infinite dimensional Gaussian process is that any finite set of points from the process can be described by a multivariate Gaussian distribution. Hence, we can write the joint distribution of the observed output values y and the unknown function values f^* as

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix}\right) \tag{9}$$

Here, if we are considering n training points and n^* test points, then $K(X^*, X)$ represents an $n^* \times n$ matrix of covariances evaluated according to equation (8) and similarly for the other matrices above. Now, we can condition the joint distribution on the observations to yield the key predictive equations for Gaussian process regression:

$$p(f^*|X, y, X^*) \sim N(\bar{f}^*, \bar{K}) \tag{10}$$

where

$$\bar{f}^* = K(X^*, X)[K(X,X) + \sigma_n^2 I]^{-1} y \tag{11}$$

$$\bar{K} = K(X^*, X^*) - K(X^*, X)[K(X,X) + \sigma_n^2 I]^{-1} K(X, X^*) \tag{12}$$

Predictions for a series of new data points X^* can now be made by evaluating equation (11) and note also that the diagonal entries of the matrix in equation (12) represent the variance corresponding to each prediction. Therefore, we can easily express an interval of 95% confidence for a given prediction $\bar{f}_i^*$ as:

$$(95\%\text{confidence})_i = \bar{f}_i^* \pm 2\sqrt{\bar{K}_{ii}} \tag{13}$$

To complete the formulation, we now need to consider fitting the Gaussian process to the unknown underlying function using the training data. As it turns out, learning with a Gaussian process model is achieved by choosing appropriate values for the parameters in the covariance function. We will denote the set of parameters we are seeking by $\theta = [l_1, ..., l_d, \sigma_f, \sigma_n]$. The optimum set of parameters is given by those that maximize the log-likelihood function corresponding to this model. Since $y \sim N(0, K + \sigma_n^2 I)$, we can write the log-likelihood as

$$\log p(y|X) = -\frac{1}{2} y^T (K + \sigma_n^2)^{-1} y - \frac{1}{2} \log |K + \sigma_n^2 I| - \frac{n}{2} \log 2\pi \tag{14}$$

And so the optimum set of covariance function parameters is given as

$$\theta^* = \arg\max_{\theta} \log p(y|X) \tag{15}$$

A custom implementation of Gaussian process regression proved to be too inefficient given the computational demands of this project and sometimes failed to successfully optimize the log likelihood function given above. Therefore, an implementation found at `http://www.GaussianProcess.org/gpml/code` is utilized for this study. Since we assume the use of a zero mean Gaussian process, the mean of the margin of victory outputs is removed prior to training. Also, each data feature is normalized to have zero mean and unit variance prior to training in efforts to improve performance.

4 Approach

In this section we explain the approach we take to produce accurate predictions of NFL game outcomes using the training dataset and Gaussian process predictive model described previously. Recall that there are two measures for success in this study. The first is to predict the margin of victory in NFL games more accurately than the Las Vegas line-makers on average. Specifically, our measure for accuracy will be the average absolute error between the forecasted and actual margins:

$$e_{avg} = \frac{1}{N_{games}} \sum_{i=1}^{N_{games}} \left(|M_{pred}^i - M_{actual}^i| \right) \tag{16}$$

Note here that our convention for a margin (similar for the point spread) is the home team's score minus the away team's score. Once predictions are made as accurately as possible with the Gaussian process model, the second goal is to take these outputs and create a bet-recommendation scheme that results in a winning rate of higher than 52.4%. The first step we take is finding an optimum set of features from which to train the Gaussian process, which is described next.

4.1 Feature Selection

Given the large amount of features and training examples, performing an exhaustive feature selection algorithm is computationally infeasible and so we adopt a more efficient approach. The feature selection scheme we perform is outlined below:

1. Choose two features for a "base-set"
2. Perform cross validation on feature sets containing the base-set plus one additional feature (for all remaining features).
3. Choose the top 20 individual features that yield the lowest CV error to form a "search-set".
4. Perform a standard forward search feature selection over the search-set starting with the base-set.

	Feature	Error
1	Computed Strength (A)	10.913
2	Pass yards per game allowed (H/S)	10.937
3	Points per game scored (A)	10.949
4	Total yards per game gained (H/S)	10.950
5	Computed Strength (H)	10.951
6	Total yards per game allowed (H)	10.969
7	Points per game scored (A/S)	10.979
8	Pass yards per game gained (A/S)	10.982
9	Pass yards per game gained (H/S)	11.000
10	Rush yards per game allowed (A/S)	11.006
...	...	...
39	Temperature Difference	11.161

Figure 2: A list of the top ten performing features that are selected for the search-set of the forward search algorithm. The error listed is the average cross-validation error when that feature is combined with home and away team's winning percentage.

This procedure allows us to eliminate more than half of the total number of features prior to performing a forward search routine while keeping only those features that seem to perform the best.

For step 1 above, we choose the base-set to be the home and away team's winning percentage as we assume these features are the most indicative of the outcome a game. In step 2 and 3, cross validation is performed by season for the 2000-2007 years to find which individual features result in the lowest error when combined with the base-set for inference. Ten of the top twenty features from these steps that make up the search-set are shown in Figure 2 along with the cross-validation error they incur. Figure 2 also indicates the performance of the novel features we are considering, showing the temperature difference feature ranked rather low at 36th while the home and away team's computed strength are 1st and 5th best, respectively.

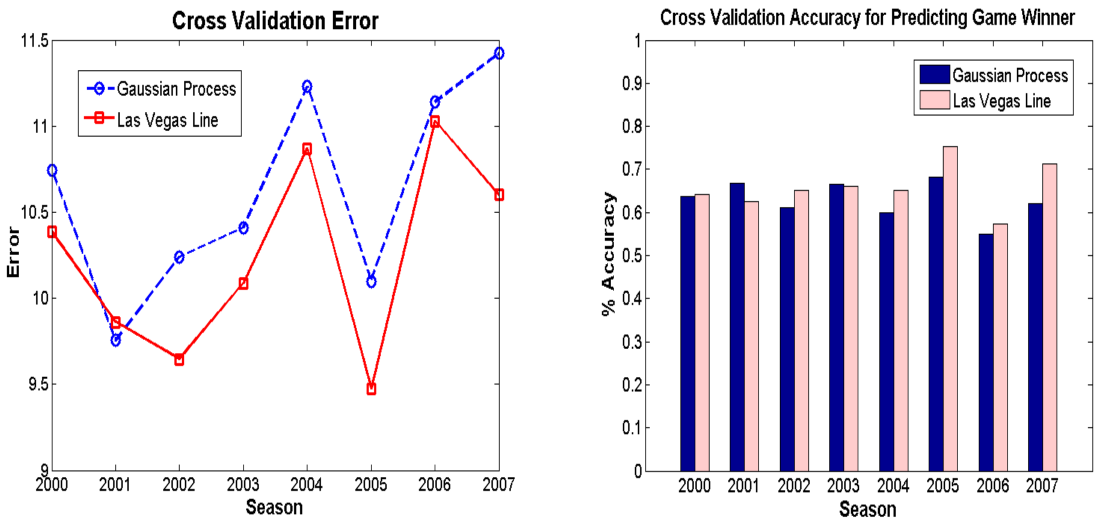
Final Feature Set #1
Winning Percentage (H)
Winning Percentage (A)
Computed Strength (A)
Pass yards per game gained (A/S)

Figure 3: A list of the final four features obtained after performing the forward search algorithm over the search-set.

Once the search-set is selected, a forward search feature selection routine is performed. Surprisingly, the search concluded with only four features in the final set, as adding any additional features from the search-set increased the cross-validation error. This final feature set is shown in Figure 3. The cross-validation error for the Gaussian process trained with this feature set is compared with the Las Vegas line error in the left graph in Figure 4. Also, the accuracy in predicting game winners is compared in the right graph. It can be seen that the Las Vegas line is slightly more accurate on average. (NOTE: several additional feature sets were constructed both randomly and by starting with different base-sets, but in the end the final set shown in Figure 3 had the lowest CV error)

4.2 Betting Framework

Once we have made predictions of the margin of victory using the Gaussian process model trained on our final feature set, we would like to be able to make a statistically-informed decision upon which games in



(a) (b)

Figure 4: A comparison of the cross-validation performance of the Gaussian process margin of victory predictions with the official point spreads produced by Las Vegas line-makers. The left plot shows the average absolute error for each season considered in cross-validation. The right plot shows the average accuracy of predicting the game winners in each game.

a season to bet on. This can be done rather conveniently by making use of confidence intervals we can construct from the output of the Gaussian process given by equation (13). Using this confidence interval, we can employ a simple bet-recommendation scheme as depicted in Figure 5. If we check the official point spread for a given NFL matchup and it is greater than upper bound of our interval, we can be relatively confident that the true margin of victory lies below the point spread. In this case, a bet in favor of the away team would win. Similarly, if the point spread lies below the prediction's confidence interval, our prediction is telling us that betting on the home team is the statistically favorable choice in this case. In the event that the point spread value lies within the confidence interval, no bet shall be placed.

5 Results

As mentioned previously, we have reserved all games from the 2008 and 2009 seasons for final testing. We now train the Gaussian process model on the complete set of training data using the final feature set given in Figure 3 and subsequently make margin of victory predictions. The resulting margin of victory errors and accuracy in predicting game winners are shown in Figure 6 and compared with the Las Vegas point spread errors. Although the Gaussian process model is seen to be effective in that it predicted game winners over 64% of the time with an average error of about 11.5 for the margin of victory, it is still slightly outperformed by the Las Vegas line-makers. The difference in performance between the two is similar to the case of the cross-validation testing. We also note than in general the Gaussian process predictions are close to the official point spreads, with an average difference of 3.38 between them for the 2008 and 2009 seasons.

We now look at the performance of the bet-recommendation scheme described earlier when applied to games in the 2008 and 2009 seasons. Specifically, we construct 95% confidence intervals according to equation (13) using the predictions on the final testing set, choose which games to bet on according to Figure 5, and tally the number of winning bets. The results are shown in Figure 7. We see that while the scheme results in more than half (50.90%) of the bets being successful, the performance falls slightly short of the goal of 52.4%, or the percentage of bets one would need to break even in the NFL gambling system. Betting in the

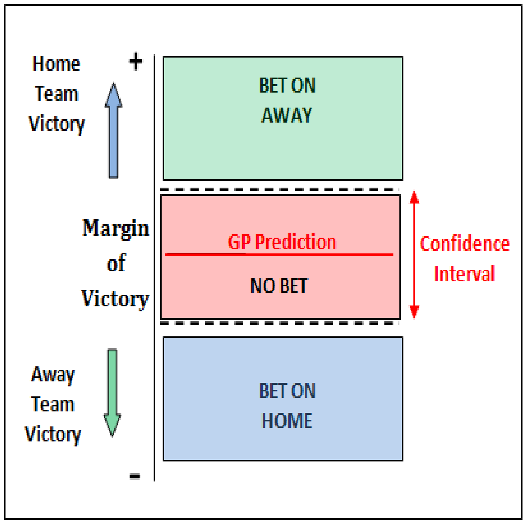


Figure 5: Diagram illustrating the bet-recommendation scheme employed. If the official point spread lies above (is greater than) the confidence interval associated with the Gaussian process prediction of the margin of victory, one should bet on the away team. Likewise, if the point spread lies below the confidence interval, one should bet on the home team. No bet should be placed if the point spread lies within the confidence interval.

	Margin of Victory Error			**Game Winner Accuracy**		
	2008 Season	2009 Season	Mean	2008 Season	2009 Season	Mean
Gaussian Process	11.291	11.752	**11.522**	66.15%	62.56%	**64.36%**
Las Vegas Line	11.167	11.439	**11.303**	68.21%	68.72%	**68.47%**

Figure 6: The performance of the Gaussian process model on the final testing dataset (2008-2009 season games) compared with that of the Las Vegas lines.

2008 season is successful with a winning percentage of 55.22%, but the overall percentage is brought down since more bets were placed in 2009 with a much lower success rate (47.96%).

Year	# Games	# Bets Placed	# Bets Won	Win %
2008	195	67	37	55.22%
2009	195	98	47	47.96%
	390	165	84	50.90%

Figure 7: The results from implementing the bet-recommendation scheme based on the Gaussian process predictions for the 2008 and 2009 NFL seasons.

6 Conclusion

In the end, this study confirms what many already know: the Las Vegas line-makers are indeed very good at what they do. As is the case in just about all of the related work in this field, our approach fell slightly short of making more accurate predictions than the Las Vegas line on the outcomes of NFL games. On average, the margin of victory predictions using the Gaussian process model for regression were about 2% less accurate than the official point spread. However, a respectable accuracy in predicting game winners (64.36%) was achieved and a win rate of 50.90% on bets for the 2008-2009 seasons was obtained using the proposed bet-recommendation scheme. Recall though that a win rate of 52.4% is required to make money on NFL bets due to the vigorish.

We were also able to explore the use of novel training features for NFL game outcome forecasting. Namely, we included the temperature difference between opposing cities in our dataset as a result of analysis done in [9] and also included team "strengths" computed according to a ranking system for sports teams described in [10]. The temperature data was found to have little bearing on predictions made, producing high errors when added to the base-set during feature selection. The computed strengths, however, performed well during feature selection and ended up in the final optimum feature set for testing. When used in place of winning percentages in the base-set, the final feature set produced an error comparable to that of the set shown in Figure 3. We conclude that incorporating such rating schemes in a predictive framework has potential for success and is worthy of further study.

Although the ultimate accuracy goals were not quite met here, this topic warrants continued research. The Las Vegas line-makers have the advantage of heuristics which are hard to quantify in a statistical model. A primary example of this would be the impact of injuries on the outcome of NFL matchups. Injuries are more prevalent in football than in just about any other sport and the loss of key players on a team can drastically alter there chances of winning. One could certainly come up with a method to quantify the impact of injuries, but the task of finding a complete set of historical injury data is difficult. It goes with out saying that this is one possible (but challenging) avenue for further research in order to outperform the point spread. Despite the fact that the temperature data utilized in the study had little impact, working to quantify the notion of home-field advantage is also worth further exploration. Data collected on factors such as stadium capacity, crowd noise, number of miles (or even time-zones) traveled over for away games, etc. could potentially improve accuracy of a predictive model.

References

[1] Zuber, R. A., Gandar, J. M., and Benny, D. B. (1985) Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games. *Journal of Political Economy*, **93**, 800-806.

[2] Boulier, B. L., Stekler, H. O., and Amundson, S. (2006) Testing the efficiency of the National Football League betting market, *Applied Economics*, **38**, 279-284.

[3] Gray, P. and Gray, S. F. (1997) Testing market efficiency: evidence from the NFL sports betting market, *Journal of Finance*, **52**, 1725-37.

[4] Sauer, R. D., Brajer, V., Ferris, S. P. and Marr, M. W. (1998) Hold your bets: another look at the effiency of the gambling market for National Football League games, *Journal of Political Economy*, **96**, 206-13.

[5] Levitt, S. D. (2004) How do markets function? An empirical analysis of gambling on the National Football League, *Economic Journal*, **114**, 2043-66.

[6] Golec, J. and Tamarkin, M. (1991), The degree of inefficiency in the football betting markets, *Journal of Financial Economics*, **30**, 311-323.

[7] Vergin, R. C. and Scriabin, M. (1978) Winning strategies for wagering on National Football League games, *Management Science*, **24**, 809-18.

[8] Vergin, R. C. and Sosik J. J. (1999) No place like home: An examination of the home field advantage in gambling strategies in NFL football, *Journal of Economics and Business*, **51**, 21-31.

[9] Borghesi, R. (2007) The home team weather advantage and biases in the NFL betting market, *Journal of Economics and Business*, **59**, 340-354.

[10] J. P. Keener (1993) The Perron-Frobenius Theorem and the ranking of Football Teams, *SIAM Review*, **35**, 80-93.

[11] Glickman, M. E. and Stern, H. S. (1998) A state-space model for National Football League scores, *Journal of the American Statistical Association*, **93**, 25-35.

[12] Knorr-Held, L (2000) Dynamic rating of sports teams, *The Statistician*, **49**, 261-76.

[13] Bassett, G.W. (1997) Robust sports ratings based on least absolute errors, *The American Statistician*, **51**, 99-105.

[14] K. Gimbel, Beating the NFL Football Point Spread. Carnegie Mellon University (Unpublished)

[15] Boulier, B. L. and Stekler, H. O. (2003) Predicting the outcomes of National Football League games, *International Journal of Forecasting*, **19**, 252-70.

[16] Harville, D. (1980) Predictions for National Football League games via linear-model methodology, *Journal of American Statistical Association*, **75**, 516-24.

[17] C. E. Rasmussen and C. K. I. Williams. Gaussian Processes for Machine Learning, The MIT Press, Cambridge, MA, 2006.ISBN 026218253X.