P. Dolan, H. Karaouni, A. Powell
Fall 2015

Machine Learning for Daily Fantasy Football Quarterback Selection

Introduction

Fantasy football has ballooned during the last 20 years from a tiny niche hobby into a massive, 15-billion-dollar per year industry. The appeal (at least in part) is due to the game's elegant simplicity: participants pretend to run a football team, picking and choosing real players from the NFL. The better these players perform in a season, the better the participant's team scores, and the more likely he or she is to beat their friends.

But the traditional fantasy system above seems to have been replaced by a new kind of fantasy football: daily fantasy football, or DFS. In DFS, participants choose players not once per season, but once per week. Further, rather than a traditional turn based draft, DFS has an auction system, where every player has a 'cost'. In this game, the optimal strategy changes from trying to rank players best-worst to instead trying to determine a 'true value' for each player, and selecting the players that are undervalued.

The following paper details an exploration of the usage of a linear regression model to project the scores of quarterbacks for DFS. Quarterbacks receive fantasy points based on five statistics: rushing yards, rushing touchdowns, passing yards, passing touchdowns, and interceptions thrown. Actual DFS points are calculated by the sum of these statistics, each weighted by known scalar values. Linear regression is thus a very suitable projection model. The input to the model changed over the course of the project, but the output remained constant: a quarterback's DFS score for a given week.

$$\text{DFS QB score} = \mathbf{.04} \times (\text{Pass Yds}) + \mathbf{4} \times (\text{Pass Tds}) + \mathbf{-1} \times (\text{Interceptions}) + \mathbf{.1} \times (\text{Rush Yds}) + \mathbf{6} \times (\text{Rush Tds})$$

The above represents the equation for computing a fantasy score as well as the format of the equation the linear regression algorithms were attempting to model

Related work

We first investigated if others had attempted to predict fantasy scores using machine learning. Matt Bookman completed a CS229 project about fantasy points prediction using both linear regression and a SVM, but found the SVM implementation in practice to be not as effective because he was constrained to a linear kernel. Nitin Kapania also conducted a similar study with k-means clustering, but reported that the algorithm got stuck in local minima/maxima and as a result become more erroneous compared to the linear regression model. Outside of Stanford, Evan Boyd found a method for ranking quarterback fantasy performance using distances between rankings. For these reasons, the project focused on tailoring a linear regression model.

Dataset and features

Data was drawn from two sources. Statistical data on quarterbacks came from a DFS site: http://www.fftoday.com/. The data came as a table, with a team and statistics on each line. An example of a line in the table is shown below:

			Passing					Rushing	
Player Sort First: Last:	Team	G	Comp	Att	Yard	TD	INT	Att	Yard
1. Russell Wilson	SEA	1	23	36	313	2	0	7	106
2. Peyton Manning	DEN	1	22	26	318	4	0	1	-1
3. Aaron Rodgers	GB	1	19	22	255	3	0	3	21

Data was collected in this format for every quarterback in the 2013-14 and 2014-15 seasons. Statistics for defenses opposing a quarterback were obtained in a similar manner. Given that there are 16 games per NFL season, the data thus comprised 1024 quarterback and defensive performances each to draw from. However, in order to focus on statistically significant data, only quarterbacks that started at least 15 games each season were considered. Roughly 60% of quarterbacks met this requirement, leaving a total of approximately 600 training examples. Cross validation was implemented with 70% of data used for testing and 30% of data used for testing.

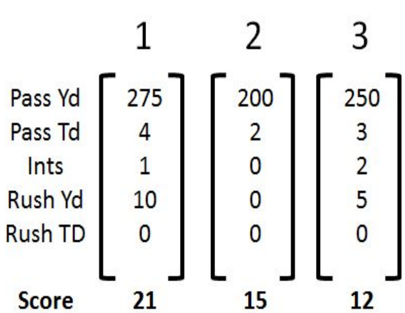
The model was trained using two primary groups of features. The first group of features were features that fell into the category of quarterback statistics. Initially, these were the raw quarterback fantasy scores for the past several weeks but later this was changed to the five individual statistics discussed earlier. The second feature group comprised of statistics pertaining to the defense. Initially, this feature was simply a ranking for the defense, but was later updated to be the average DFS score allowed by the defense.

Methods

The aforementioned linear regression model was the only model explored for DFS quarterback score projections. This decision was motivated by two factors: first, linear regression is a natural model for the problem given the calculation of DFS scores. Second, prior attempts to solve a similar problem in Stanford's machine learning course and beyond found linear regression models to perform better than or only negligibly worse than other models. The following methods thus focused on two approaches to linear regression explained below.

The first approach was focused on creating a generalized regression model for all quarterbacks and was purposefully simple. The only features used were a quarterback's prior fantasy scores. Intuitively, the model aimed to learn a general way to weight a quarterback's score for the previous few games (henceforth referred to as 'lookback') in attempt to project a score for a given week. The first attempt to improve the model was to consider varying numbers of lookback games. A lookback of three was eventually chosen.

The next iteration of the model was to learn each quarterback statistic individually and weight the scores based on DFS standards to project a score.



3 example games

Using the three previous games as shown to the left in, the initial model would only have the quarterback's overall scores (21, 15, 12) as features. In the second model each individual statistic is predicted. Pass yardage is predicted based off 275, 200, 250, Pass Tds off of 4, 2 and 3, and so on.

The second model performed noticeably better than the first model in the above example, which intuitively makes sense: what appears to be highly variant games by a quarterback may be variant in only a single statistic. This level of granularity allows the model to make accurate predictions on the statistics that stay constant, and limits the significance of variant ones.

The next step in creating a generalized regression model for all quarterbacks was to implement defensive ranking as a feature. This was challenging from an implementation perspective due to the difficulty in acquiring the data. However, what made it most difficult was determining how to use it as a statistic. Because of the approach of projecting each statistic individually, it was impossible to simply add it as another feature. Instead, an initial score was projected based on projected statistics and subsequently adjusted based on the defensive ranking.

The second approach was to craft a linear regression model for each quarterback individually, as opposed to for quarterbacks as a whole. To clarify, that means that in order to predict Quarterback A's score, we learned exclusively based on Quarterback A's prior performances and projected forward without any knowledge of other quarterbacks. The approach proved to be quite helpful and was the last step in what became the optimal quarterback-specific model.

Results

The metric used for the evaluation of methods was mean absolute error, which is appropriate given the nature of the problem. For each model, and surely for DFS participants, it is most important to understand the degree to which predicted and actual scores differ. As mentioned previously, 70/30 cross validation was used.

$$\text{Mean Absolute Error} = \left| \frac{\text{(Every Predicted Score)} - \text{(Every Actual Score)}}{\text{Total Predictions Made}} \right|$$

An important parameter for model selection was the lookback mentioned earlier. Qualitatively, looking back to only a certain number of performances was promising because it accounts for the fact that a quarterback does not perform consistently across a career. Quarterbacks instead tend to have streaks, varying between high and low performances. Quantitatively, the degree of lookback did have an effect on the efficacy of the models. In summary, the ideal lookback was found to be three games, with an average test error of 7.2. A quantitative summary is shown in Figure 1. The graphs represent the generalized regression model using quarterback statistics but not yet accounting for defensive features.

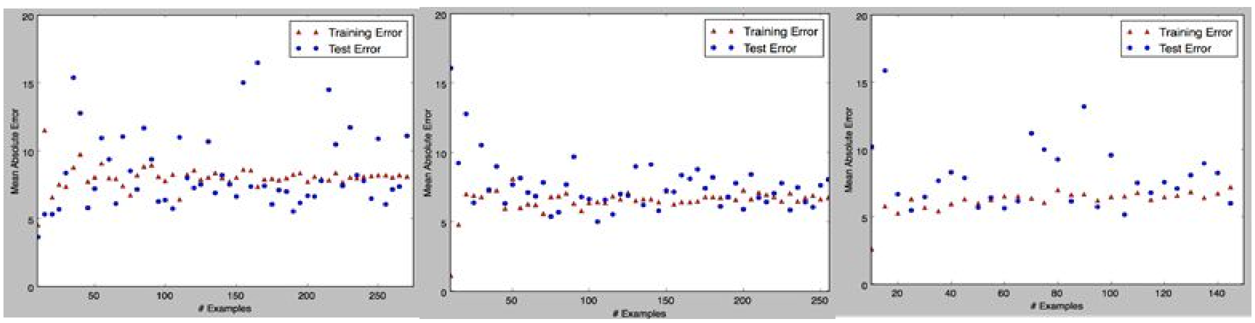


Figure 1: Training and test error using a lookback of one, three, and six respectively. To summarize, a lookback of 3 was found to have the least variability. The x-axis shows the size of the training set used.

Figure 1 demonstrates that plotting the training and test error for multiple data set sizes was a useful diagnostic for improving the models. Specifically, Figure 1 displays that the model exhibited high bias. Train and test error approximately converged, thus hinting that more data for the model would not improve performance. Such a conclusion was qualitatively disheartening - an average test error of 7.2 seemed unacceptably erroneous. To quantify the error, however, the model was compared to a human expert as shown in Figure 2. The expert has an average error of 8.5 so, although model test error seemed high, the model still performed well compared to alternatives. An example week is shown below.

Predicting Fantasy Scores for QBs, 2014 Week 16					
Linear Regression	Pred. Pts.	**Human Expert**	Pred. Pts.	**Actual Week 16**	Actual Pts.
Matt Ryan	23.8	Drew Brees	28	Russell Wilson	36.4
Ryan Tannehill	21.1	Andrew Luck	24.8	Ryan Tannehill	30.1
Derek Carr	20.8	Aaron Rodgers	23.8	Colin Kaepernick	29.6
Eli Manning	20.8	Matt Ryan	23.3	Eli Manning	27.3
Drew Brees	19.8	Matthew Stafford	23.3	Philip Rivers	27.2
Ben Roethlisberger	18.1	Tom Brady	23	Matt Ryan	17.6
Russell Wilson	17.3	Ben Roethlisberger	22.8	Aaron Rodgers	17.5
Philip Rivers	17.2	Peyton Manning	21.5	Derek Carr	16.7
Aaron Rodgers	16.6	Ryan Tannehill	18.8	Peyton Manning	16.4
Andrew Luck	14.2	Russell Wilson	18.5	Andy Dalton	15.3
Tom Brady	14.2	Andy Dalton	16.5	Drew Brees	15.3
Colin Kaepernick	14	Colin Kaepernick	16	Joe Flacco	13.3
Joe Flacco	13.4	Eli Manning	16	Ben Roethlisberger	12.4
Andy Dalton	13	Joe Flacco	16	Tom Brady	11.2
Peyton Manning	12.2	Philip Rivers	16	Matthew Stafford	7.9
Matthew Stafford	11.3	Derek Carr	15.3	Andrew Luck	2.4

Figure 2: Scores for quarterbacks in Week 16 of 2014. Green cells indicate projections within 5 points of the actual score. Red cells indicate projections off by over 10 points. The linear regression model outperformed human expert, Mike Krueger from http://www.fftoday.com/.

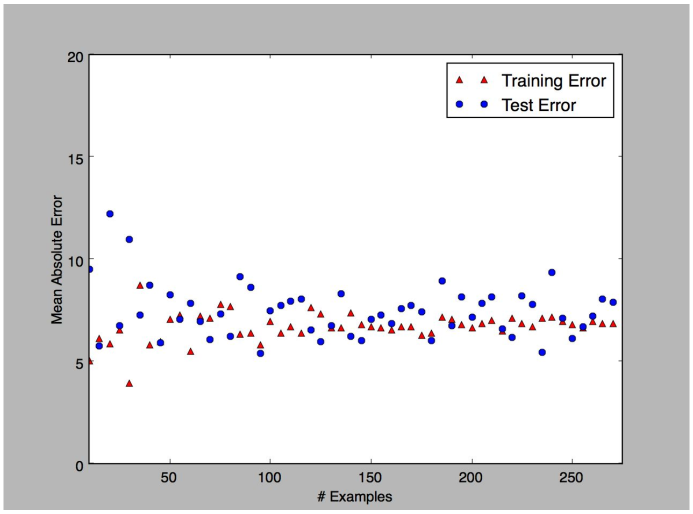


Figure 3: The graph displays the model using defensive features and no discernible improvement in effectiveness.

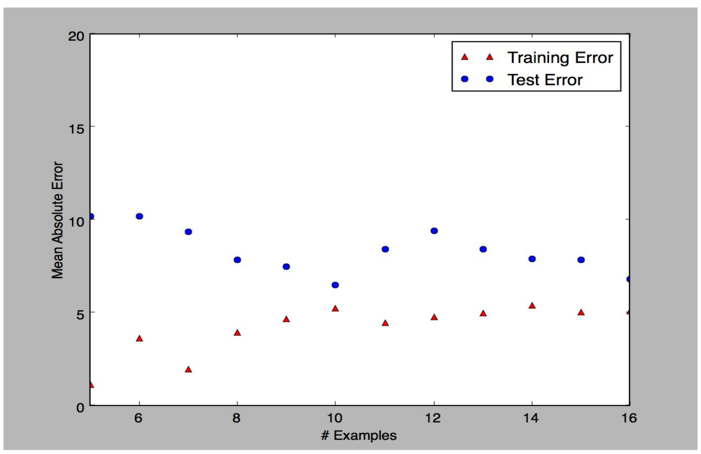


Figure 4: Training and test error for the quarterback-specific model.

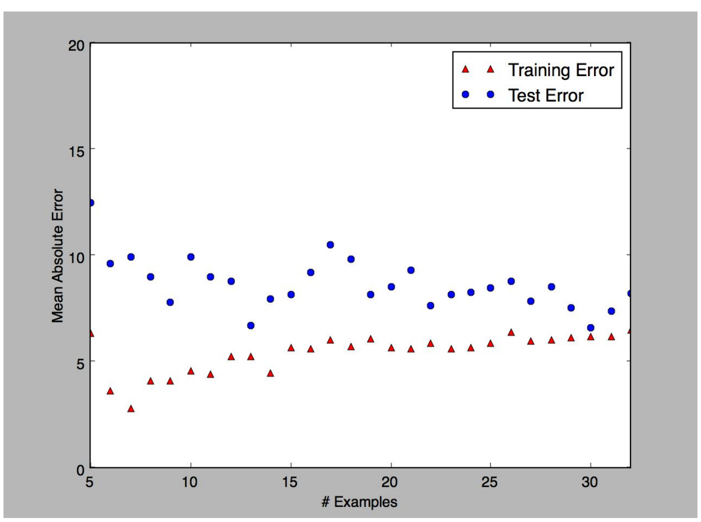


Figure 5: Training and test error for the quarterback-specific model. Two seasons of data are used.

While the model performed suitably compared to a human expert, efforts were made to improve performance. As the model was diagnosed to have high bias, two types of improvements were evaluated: adding more features and changing to the quarterback-specific model.

When considering features to add, defensive features were added first. A new model integrated the quality of the opposing defense but did not seem to have a discernible effect on the projection accuracy.

After the addition of defensive features was considered, it was determined that the most viable next step would be to consider the quarterback-specific model. The motivation was that, beyond defensive features, few features seemed to have the promise of dramatic improvements. Consider the types of features which could be added: player injuries, weather, stadium-type, and perhaps others of similar ilk. Qualitative knowledge of the NFL hints that featuring player injuries would bear the most potential, but would still be impactful for only a handful of games and entail obtaining very complex data.

From a qualitative standpoint, the quarterback-specific model seemed more promising than the generalized regression model. The generalized model relies on the assumption that most quarterbacks have a similar playing style. This is a non-issue for most quarterbacks. For instance, the NFL's top quarterbacks typically score a large number of fantasy points in the same way: passing. However, there are also quarterbacks with a dominant rushing presence such as Colin Kaepernick and Russell Wilson. The generalized model would understandably not generalize well to these quarterbacks and this is shown quantitatively in Figure 2. Russell Wilson and Colin Kaepernick are two of the most innaccurate predictions alongside Andrew Luck who is also well-known for rushing ability.

The new model was first evaluated within a giving season, meaning that there were 16 examples per model. The training and test errors were averaged for all quarterbacks and plotted in Figure 4.

The change resulted in a 23% improvement in training error as the average training error was reduced from 6.5 to 5.0. Figure 4 also suggested that, with more data, test error may decrease. Data was thus incorporated from the 2013-14 season for a total of 32 performances per quarterback.

The result, shown in Figure 5, is reminiscent of the generalized quarterback approach and less performant than the single season quarterback-specific model. Although at first surprising, such a result is sensible upon further consideration. One may correctly expect the performance of elite quarterbacks such as Tom Brady and Aaron Rodgers to remain consistent across seasons. However, consider quarterbacks such as Eli Manning and such an expectation is no longer valid. For instance, Eli Manning's average DFS score per game in the 2013-14 season was 14.3 compared to 19.1 in the 2014-15 season. Statistically speaking, Eli Manning was an entirely different quarterback in each season and thus trying to model him across the two seasons was similar to modeling him using the generalized linear regression model.

Conclusion

Predicting quarterback success on a game-by-game basis is an extremely challenging problem due to the high degree of variance that a typical NFL game provides. However, linear regression on a per quarterback basis using defensive ratings and looking back three games generates a predictive model better than an expert NFL fantasy analyst.

While the algorithm in question performed better than an expert, it still suffers from high bias and would likely be bettered with additional features. Specifically, it would be extremely interesting to get a more granular look at both the offense and the defense in a given week, passing in the percentage of available starters to both sides as a feature. One other feature which which would be difficult to implement but likely extremely interesting would be to pass in a sentiment analysis of how positive any article about the team had been in the week leading up to that game. One especially complicated yet valuable final feature to implement would be to attempt to quantify how a team changed over the offseason by using some combination of player/coach turnover, large-scale NLP analysis on team articles and talking-head generated NFL draft 'grades' of the players drafted. Such a feature would help to understand how a team changed across an offseason and thus aid in the use of the quarterback-specific model across multiple seasons.

Ultimately, however, most of the features discussed in the previous paragraph were well beyond the scope of this project and could even be large projects of their own. Our team was quite happy with the results we achieved and we were ultimately pleased that we managed to beat a fantasy expert. While there's certainly room for improvement, we think this project shows the potential machine learning has in the world of fantasy sports.

References

Bookman, Matt. “Predicting Fantasy Football - Truth in Data”. December 14, 2012. <http://cs229.stanford.edu/proj2012/Bookman-PredictingFantasyFootball.pdf>.

Boyd, Evan. “A New Method for Ranking Quarterback Fantasy Performance with Assessment Using Distances Between Rankings”. <http://mds.marshall.edu/cgi/viewcontent.cgi?article=1841&context=etd>.

Chen, Boris. “Drafting a Fantasy Football Team, With Help From Advanced Statistics”. *The New York Times*. August 6, 2014. <http://www.nytimes.com/2014/08/07/sports/football/drafting-a-fantasy-football-team-with-help-from-advanced-statistics.html>.

Fantasy Football Today. 2013-2014 Weekly NFL Fantasy Quarterback Data. <http://www.fftoday.com/>.

Kapania, Nitin. “Predicting Fantasy Football Performance with Machine Learning Techniques”. December 14, 2012. <http://cs229.stanford.edu/proj2012/Kapania-FantasyFootballAndMachineLearning.pdf>.

Pro Football Reference. 2013-2014 NFL Schedule Data. <http://www.pro-football-reference.com/years/2013/games.htm>.