Fantasy Football Predictions Using Ridge Regression

Blake Washburn and Ryan Carney
Clemson University
bwashbu, rjcarne

Abstract

Fantasy Football is a statistical online game which can have high stakes. Some users have put their hope in statistical learning models to get an edge on their competition. Due to the nature of the scoring methods used by most fantasy football, machine learning could potentially provide an insight into the probability that a NFL player will preform well or poorly in an upcoming game. This paper will consider a ridge regression model for predicting the future performance of players, as well as explore the different clustering and comparison models between players of the same position.

1 Motivation

Fantasy Football has grown in popularity over the last two decades. Leagues can vary in intensity, from casual play with random opponents to high-stakes leagues costing hundreds to participate. Naturally, players want to maximize their odds of winning. The authors of this paper are no different. In fact, all of major fantasy football platforms, such as ESPN and YaHoo, provide their own player projections using their own models. Meanwhile, freelancers online also sell their predictions to users looking for a different perspective on projections. Thousands of users consider these predictions week after week during the NFL season when setting their player lineup. It is evident that there is both local and national interest in the development of an accurate prediction model for fantasy football scoring.

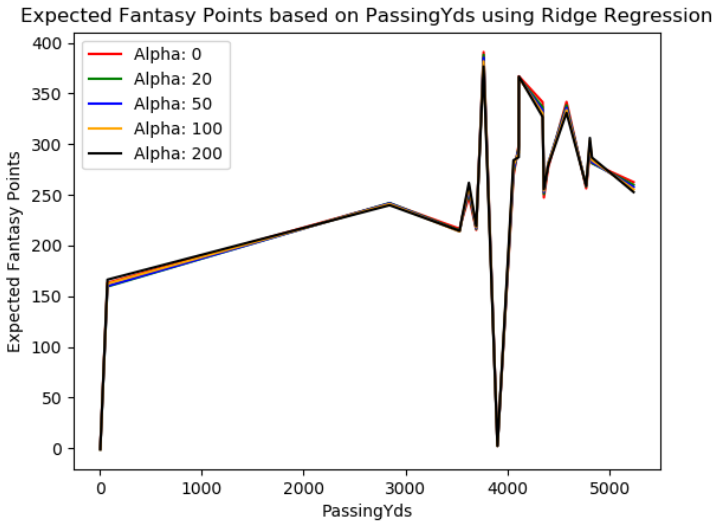
2 Method

The data was collected by fantasyfootballpros.com from NFL stats ranging back to 1970 [1]. It provides the most comprehensive set of football statistics that the authors could obtain for free on the internet. Although there is a plethora of information in this data set, it still needed to be cleaned for the purposes of this research. Before 1992, some player attributes, such as attempted passes or completed catches were not tracked. Furthermore, the culture of the game of football was drastically different prior to 1992, as all the teams in the league have shifted overtime from a run-dominated game to the more fast-paced, passing-heavy, type of offense that Americans are use to seeing today. For that reason, any data prior to 1992 was not considered in the creation of our models.

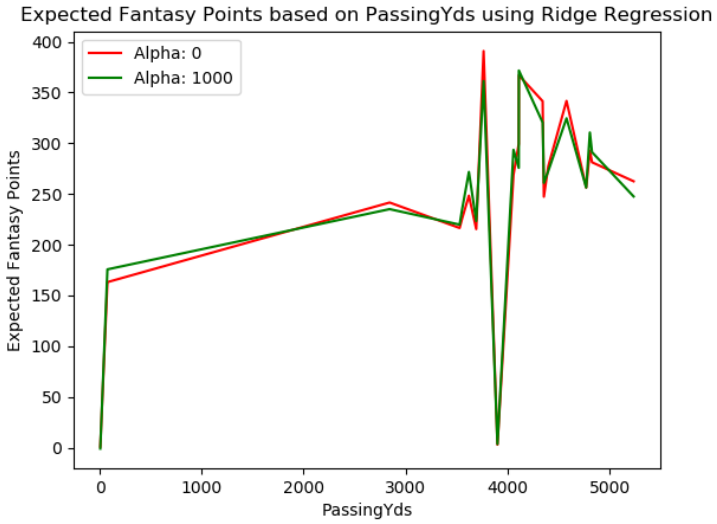
When considering which technique would best model the data once it had been cleaned, some offhand graphs were produced and analysed. From these, the authors could conclude that the data set boasted a large number of features and tended to over-fit most preliminary models, as will be described below. As such, the team believes that a ridge-regression type model will most accurately predict the future fantasy football performance of NFL players.

3 Preliminary Experiments

As a proof of concept for the regression strategy, a ridge regression was carried out on to predict the Fantasy Point value of a certain player based on a given feature. It was found that the regression coefficients were over-fit to the data set, likely due to the small sample size.



In addition, there are some interesting things to consider with the data beyond the regression, namely that increasing the reliance on the minimum sum of square coefficients (alpha) did not have a drastic effect on the projected data.



In order to aid our research, we have set up various modeling capabilities so that we can easily refine our data and quickly evaluate new statistical possibilities. Some of the baseline tests that we have run have been difficult to comprehend when transformed into a 2D graph. Thus, we are currently attempting to create other methods for predictions. One that we have experimented with is using the mean values as inputs for prediction of future data.

4 Next Steps

Moving Forward, ridge regression will be implemented and used across all applicable positions from the data set. In the case of inaccuracies, we may pursue other forms of regression such as Lasso, Linear or Polynomial regression. Graphs will be generated to visually represent the results where applicable. Because the intent is to perform forward projection, we will focus our efforts on tuning a model that is flexible enough that it is useful in projecting fantasy football results in real world scenarios. In order to do that, our goal moving forward is to transition from a data exploration phase into more focused application of regression models to answer specific questions we may have that could be prudent to our research goals.

5 Contributions

The authors worked closely on the development of the project, but focused on different aspects of the project to maximize efficiency. The authors utilized Discord to facilitate collaboration and host regular meetings to mutually decide on important research questions. Mr. Carney lead the way on researching related work and implementing the preliminary experiments, while Mr. Washburn focused on composing the proposal and milestone report. However, the final product from both focus areas is a result of collaboration between the authors directly and verbally.

References

[1] Data Pros, Fantasy Football. Data Sets, FFDP, 2020, www.fantasyfootballdatapros.com/csvfiles.