Partial Least Squares Regression in Football Projections

Author:
Kalyan Chadalavada

Advisor:
Dr. Jo Hardin

May 5, 2018

Abstract

Partial least squares (PLS) is a developing modeling technique that is gaining traction in the world of predictive models. It has the capability to compress a large number of explanatory variables into fewer components that are in turn used in a linear model to predict response variables. What sets PLS apart from principle components analysis (PCA) is that PLS also relates the response variables to the explanatory variables for more accurate predictive capabilities, as well as predicting multiple response variables. In this paper, I will examine the mechanism that PLS utilizes and apply it to NFL quarterback data from 2009 to 2017 to predict the total yardage a quarterback may have in future seasons.

Contents

Chapter 1

Introduction

Every day models are used to predict an outcome, and many of them are barely even noticed. From choosing what to eat to following the behavior of the United States' economy and everything in between, people utilize different techniques to decide what is the best method of action for the future. When ordering a meal, a customer will think about what he already knows he likes to eat, what his stomach is craving in the moment, what has been recommended to him, the restaurant's specials, any news on health issues on certain foods, the type of diet he is on, and the list goes on and on. Although all of these questions are answered in just a short amount of time, choosing a meal is still considered a trivial action that is almost second nature.

Now, predicting the ups and downs of the seemingly-capricious movement of the New York Stock Exchange is a task that even professionals have trouble mastering. Every day people buy shares of hundreds of different companies in the hopes of falling on the fruitful side of luck and making a fortune. When they buy a stock, they believe that based off of current events in the world, trends in that stock from the past, advice from financial advisors and other professionals they, too, will be able to catch a boom in that stock and make some profit from the investment. Figure 1.1 portrays the volatile nature of the stock market. Although the general trend of the graph is positive after 2009, there are many dips that would result in a loss of principal. The dive in the stock market in 2008 would have resulted in a huge loss of money. With such a temperamental behavior, it would seem that plotting the course of stocks for the future is an insurmountable task.

The movie *The Big Short(2015)* argues that it actually is possible to predict even the stock market. In this movie, a small group of economists were able to foresee the economic collapse of 2008, use this foresight to their advantage, make specific financial actions to brace for the collapse, and make millions while many others suffered. Now although these economists may not have used multiple linear regression to become wealthy, it is one technique to create a model in order to predict future data. Multiple linear regression (MLR) creates a linear model as a function of multiple explanatory variables to predict

Figure 1.1: NYSE Composite Index from 2006-2015[1]

a single response variable. However, MLR may not be suitable for all data sets, particularly ones that do not satisfy the assumptions required for multiple linear regression. Another analysis method that can be used is principal components regression. Principal components creates linear combinations between the most important explanatory variables to best match the existing data. It gives weights to each explanatory variable, assigning individual weights to each predictor to best represent the data set. An extension to principal components analysis is partial least squares regression. In this report, I will introduce multiple linear regression and principle components analysis to provide an in depth explanation of partial least squares regression. Finally, I will apply partial least squares analysis to National Football League(NFL) data to show how player productivity and value can be predicted with a model. In turn, this model can be used by teams to recruit more accurately and more confidently.

Chapter 2

Background

2.1 Multiple Linear Regression

Simple linear regression is a model that only uses a single explanatory variable or predictor, x, to predict a single response variable, y. A simple linear regression model can be represented as:

$$\hat{y} = b_0 + b_1 x \qquad (2.1)$$

where b_0 represents the y-intercept of the model and b_1 is the estimate of the slope needed to predict y from x. In Figure 2.1, we see that x is the explanatory variable that is being used to predict the response variable y. Values for x were generated in R as one hundred normal random natural numbers with a mean of zero and a standard deviation of one. Y was calculated similarly however each natural number of x was multiplied by x. In this case, the b_0 is 0.035 and b_1 is 1.046. The regression line is the representation of the simple linear regression model, and the actual data is represented by the unfilled dots around the regression line. For these data, the model only has one explanatory variable, but what if there are two explanatory variables? More than 2? For example, in Figure 2.2, how would an analyst be able to correlate all ten explanatory variables to create a model for his desired response variable.

Multiple linear regression(MLR) is a method of incorporating multiple explanatory variables to predict a response variable with a linear function. In order for MLR to be possible, certain assumptions must be met. First, the relationship between the explanatory and response variables must be roughly linear. The best way to confirm linearity is by creating scatter plots of the data. For example, in Figure 2.1, the original scatter plot of x versus y shows a linear relationship between the two variables. However, in Figure 2.3, y_2 is defined as x multiplied by 100 normal random natural numbers, that have a mean of zero and a standard deviation of 1. Here, Figure 2.3 shows a non-linear relationship between x and y_2. These two variables would, therefore, not be suitable for MLR.

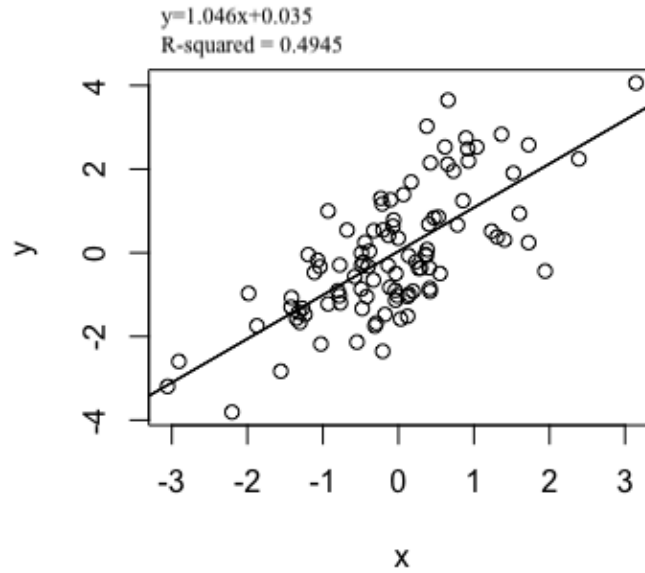Figure 2.1: Simple linear regression model using randomized data in R

y=1.046x+0.035
R-squared = 0.4945

Figure 2.2: Sample Data Table on Top Running Backs[2]

PLAYER	TEAM	ATT	YDS	YDS/A	LONG	20+	TD	YDS/G	FUM	1DN
Kareem Hunt, RB	KC	272	1,327	4.9	69	12	8	82.9	1	60
Todd Gurley II, RB	LAR	279	1,305	4.7	57	8	13	87.0	5	66
Le'Veon Bell, RB	PIT	321	1,291	4.0	27	3	9	86.1	2	74
LeSean McCoy, RB	BUF	287	1,138	4.0	48	12	6	71.1	3	55
Mark Ingram, RB	NO	230	1,124	4.9	72	11	12	70.3	2	50
Jordan Howard, RB	CHI	276	1,122	4.1	53	5	9	70.1	1	61
Melvin Gordon, RB	LAC	284	1,105	3.9	87	5	8	69.1	1	58
Leonard Fournette, RB	JAX	268	1,040	3.9	90	4	9	80.0	2	46
C.J. Anderson, RB	DEN	245	1,007	4.1	40	7	3	62.9	1	47
Ezekiel Elliott, RB	DAL	242	983	4.1	30	5	7	98.3	1	55

The second assumption is that all variables of the data set must be normally distributed. In other words, the residuals at any subset of explanatory variables form a normal distribution around the linear regression line as seen in Figure 2.4. There are many ways to test this assumption, including checking to see that the Q-Q plot of the data is linear.

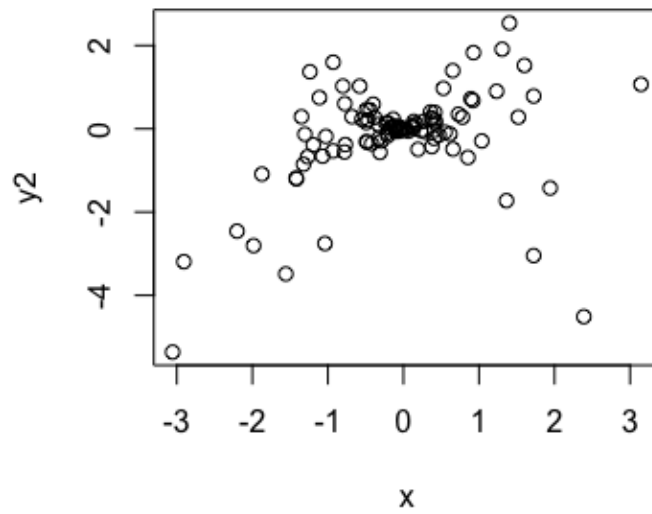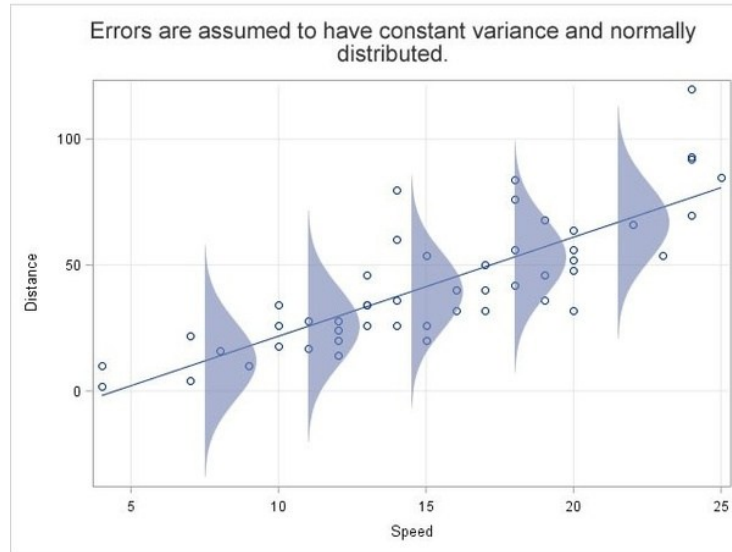Figure 2.3: Scatter Plot of x and the New Response Variable, y2

Figure 2.4: Depiction of Normal Errors[6]

The third assumption is that there should constant variance amongst the data. If we are to plot the fitted values of a model versus the residuals, the data should be randomly spread out around the zero line of the plot. There should be no increase or decrease in error across the fitted values.

The fourth assumption is independence amongst the observations. In other words, the appearance of one observation should have no effect on any other observation.

If these four assumptions are met, the data can accurately be matched to a multiple linear regression in the form of:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + ... + b_m X_m + \epsilon \tag{2.2}$$

where Y is the response variable, $\{b_0, ..., b_m\}$ is the coefficient vector that minimizes the sum of square errors, X_m is each explanatory variable, m is the number of predictors, and ϵ is the error of the model. Even without these four assumptions, the parameter estimates can be fit using least squares, but the accuracy of the model would falter.

## 2.2 Principle Components Analysis

Multiple linear regression is quick and accurate method to model linear data, but what if the value for m in Equation 2.2 is incredibly high? Looking back at Figure 2.2, we see a data set that has eleven variables to work with. Although these variables are the ones recorded by ESPN for running backs, they do not fully encompass all factors that effect running back output. For

example, one variable that is not accounted for in Figure 2.2 is the capability of one's offensive line. No matter how talented a running back is, the offensive line is necessary in creating openings to run through and keeping defenders at bay. Now what are all the factors that contribute to the effectiveness of the offensive line? We could still use MLR for a large number of explanatory variables, but some variables like the offensive line, speed of the running back, and others are more substantial in affecting running back output than variables like a team's wide receiver output. The less important pieces of information can be left out of models as they have little effect on the response variable. However, how can we decide which variables have more influence on the data?

Principal components analysis (PCA) is a method of analysis that can solve both of these questions that MLR cannot. When dealing with m number of of variables, where m is very large, PCA serves to compress the data to be focused on k components, where k is less than or equal to m. Although the full dataset includes all m variables, the k variables will be able to explain approximately the same amount of variability in the response variable that the m explanatory variables do. Figure 2.5 shows a broad overview of how PCA zooms the scope of the model to only incorporate the most important variables.

In other words, PCA creates linear combinations, or principle components (PC), of all variables to represent the data set. The top image of Figure 2.5 depicts a model that uses a set of the original explanatory variables to predict y1. M1 portrays a model that does not have any relation between y1 and x5. M2 utilizes the same five explanatory variables to predict y1. This shows how linear combinations of x1-x4 can create just two components, C1 and C2. These components are then used in a linear model to predict y1. Similarly, PCA can be used in football data like from Figure 2.2 to compress the enormous number of variables into fewer variables which are then utilized in a linear model to predict a response variable[4].

Figure 2.5: Principle Components Analysis Flow Chart [3]
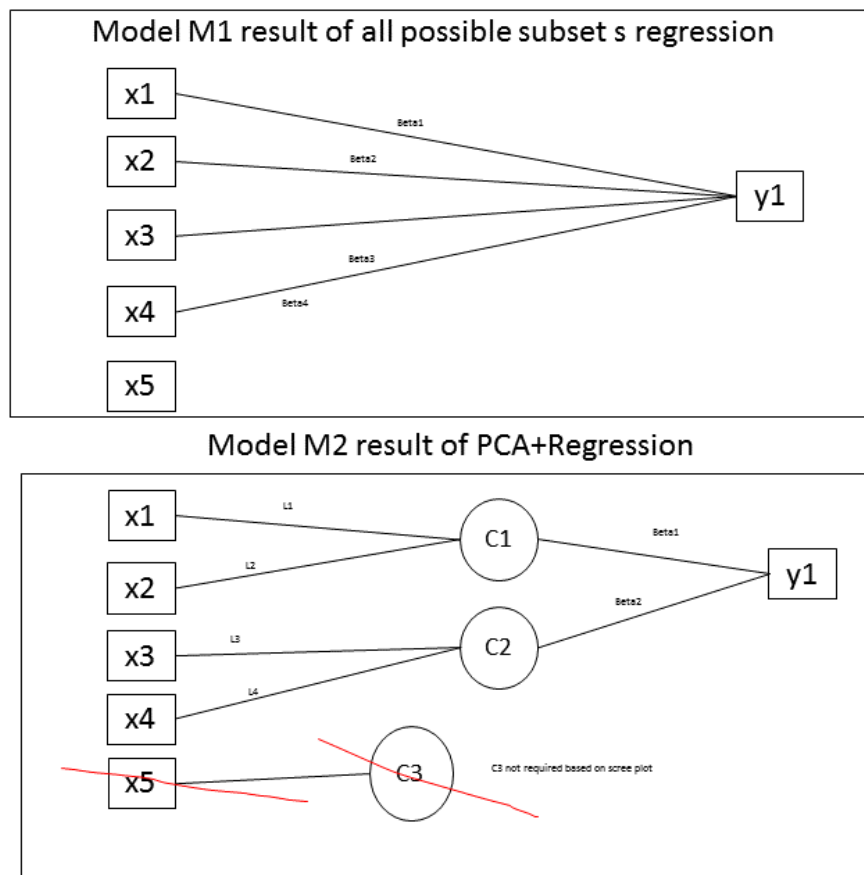
Chapter 3

Mathematical Derivation of Partial Least Squares

Much like PCA, Partial Least Squares (PLS) aims to compress the original data matrix in order to create a regression model with fewer explanatory components than number of variables in the original data space. Similarly, PLS creates linear combinations to represent the explanatory variables with latent variables, and the greater the number of latent variables or components, the more representative of the data these latent variables will be. The sets of linear combinations for both PLS and PCA are from the explanatory variables; however, the key difference between the two modeling techniques lies in the optimization criteria that created the coefficients. PCA aims to create a linear combination that will project the data onto the one dimensional space that captures the maximum variability in X, but PLS's goal is to do the same for $X^T Y$. We will dive into more depth on how PLS achieves this goal later in this chapter.

Let's suppose we have a sample of size N which contains the response variables $Y_1, Y_2, ..., Y_p$ and the explanatory variables $X_1, X_2, ..., X_m$, where p and m are the number of response and explanatory variables, respectively. When p is only one, we have the option of creating a linear regression model of the following form:

$$Y = XB \tag{3.1}$$

where the regression coefficients, B, can be estimated with:

$$\hat{B} = X^+ Y \tag{3.2}$$

Here, X^+ is defined as $(X^T X)^{-1} X^T$, otherwise known as the matrix pseudoinverse. However, collinear variables in the X matrix create trouble in this set of solutions because $X^T X$ must be invertible. The issue of non-invertible $X^T X$ matrices can be solved by using PLS. We must decompose the X and Y matrices into the following forms:

$$X_{N \times m} = T_{N \times r} P^T_{r \times m} \quad Y_{N \times r} = T_{N \times r} C^T_{r \times p} \tag{3.3}$$

where T is an orthogonal matrix that is common to both X and Y. In PLS, the common matrix T acts to relate the explanatory variables with the response variables for more accurate latent variables. The T matrix is arbitrary at the moment, but it is the same matrix in the decomposition of X and Y matrices to maximize the variability of $X^T Y$ captured by the linear combination between X and Y. P and C are the loadings matrices for X and Y, r is the rank of the input matrix. T, P, and C must be computed iteratively. We must calculate a component, $w \in \mathbb{R}^{m \times 1}$, such that the covariance between X and Y is maximized. Let's define the matrix, S, as such:

$$S_1 = X_1^T Y_1 \tag{3.4}$$

where X_1 and Y_1 represent the initial conditions of the response and explanatory variables. That is, $X_1 = X$ and $Y_1 = Y$. The subscripts for these matrices indicate that PLS is a iterative process where each covariance matrix after the first is an approximated version of the original data matrices. Because the covariance equation is $cov(X) = XX^T$ if X is centered, we know that the $cov(X^T Y) = X^T Y X Y^T = SS^T$. Based on the initial conditions X_1 and Y_1, the first component vector w_1 is the solution to the following optimization problem:

$$w_1 = \underset{||w||=1}{\arg\max}\, w^T S_1 S_1^T w \tag{3.5}$$

which then maximizes the $cov(w_1 X^T Y)$. w_1 is also the largest eigenvector of $cov(X^T Y)$,

$$w_1 = Eig_{max}\{S_1 S_1^T\}. \tag{3.6}$$

Here, $Eig_{max}\{\}$ is the operator that extracts the maximum eigenvector from a given matrix. We, then, take our newly calculated first component, w_1, and use it to calculate the first column of the scores matrix T in the following way:

$$t_1 = X_1 w_1. \tag{3.7}$$

Eventually, we want to rewrite equation (3.3) as a summation of the multiplication between the columns of T and P for X and between the columns of T and C for Y.

$$X = \sum_{d=1}^{r} t_d p_d^T, \quad Y = \sum_{d=1}^{r} t_d c_d^T \tag{3.8}$$

However, our PLS equation only calculated the first score, t_1, and we do not know the other column values for the T matrix. This first column of the scores is used as an approximation of the X and Y matrices as such:

$$X \approx t_1 p_1^T, \quad Y \approx t_1 c_1^T \tag{3.9}$$

where p_1 and c_1 are the first columns of their respective loadings matrices. Later, I will show how to calculate X_2. P and C have not yet been calculated so we must figure out a way to determine these values. Upon closer inspection, the equations in (3.9) are similar to equation (3.1). Using least squares as in

equation (3.2), we apply the same method to determine the first columns of P and C. We do so by projecting the values of X_1 onto t_1 as seen in equation (3.10)

$$X_1 = t_1 p_1^T \quad \rightarrow \quad \hat{p_1} = (t_1^T t_1)^{-1} t_1^T X_1 \tag{3.10}$$

The Y matrix is treated similarly to calculate c_1 as such:

$$Y_1 = t_1 c_1^T \quad \rightarrow \quad \hat{c_1} = (t_1^T t_1)^{-1} t_1^T Y_1, \tag{3.11}$$

which simplifies to equation (3.12) determining p_1 and c_1.

$$p_1 = \frac{X_1^T t_1}{t_1^T t_1}, \quad c_1 = \frac{Y_1^T t_1}{t_1^T t_1} \tag{3.12}$$

We then subtract $t_1 p_1^T$ from X to create a new deflated matrix. This process removes the influence of t_1 on the original X and Y matrices, and we are left with the following:

$$X_2 \leftarrow X_1 - t_1 p_1^T, \quad Y_2 \leftarrow Y_1 - t_1 c_1^T \tag{3.13}$$

In other words, we subtract the component in the direction of the maximum $cov(w_1 X^T Y)$ onto the remaining $m - 1$ space. Figure 3.1 is a graphical representation of this process. PCA 1st Dimension is removed from the dataset, and the data is then projected onto the remaining space to form X_2. The component in the direction of maximum variability then becomes PCA 2nd Dimension for X_2. The decomposition above is applied to the response variables when $r < p$; however, as will be seen in Chapter 4, I will only be predicting a single response variable. Therefore, there is no decomposition for the response variables, only the X matrix.

Figure 3.1: Components in Direction of the Maximum $cov(w_1 X^T Y)$[5]
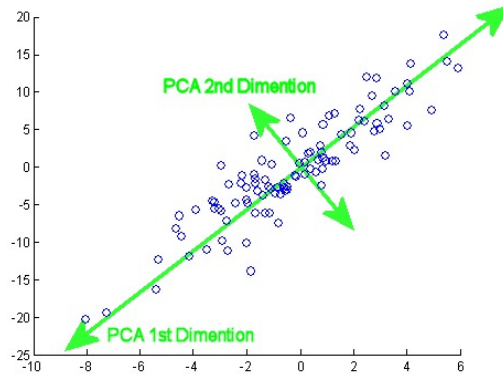
Equations (3.1)-(3.13) are repeated r times in order to calculate the remaining vectors for W, T, P, and C. We then compile the vectors in their

respective matrices as such:

$$W = w_1, ..., w_r \quad T = t_1, ..., t_r \quad P = p_1, ..., p_r \quad C = c_1, ..., c_r \tag{3.14}$$

Finally, we are left with a newly formed X matrix marked as:

$$\tilde{X} = TP^T = TP^T \times I = TP^T WW^T = T(P^T W)W^T \tag{3.15}$$

where I is the identity matrix. The coefficients for this PLS model are calculated in a similar way as equation (3.2).

$$\hat{B} = \tilde{X}^+ Y \tag{3.16}$$

$\tilde{X}^+$ is calculated as such:

$$\tilde{X}^+ = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tag{3.17}$$

Figure 3.2 provides a summarized version of the steps PLS takes to model data.

Figure 3.2: NIPALS Algorithm[8]

Algorithm 1 The NIPALS Algorithm

1: **Initialise:** $\boldsymbol{X}_1 \leftarrow \boldsymbol{X}, \boldsymbol{Y}_1 \leftarrow \boldsymbol{Y},$
2: **for** $d = 1, \ldots, r$ **do**
3:      $\boldsymbol{S}_d = \boldsymbol{X}_d^\top \boldsymbol{Y}_d$
4:      $\boldsymbol{w}_d = \text{Eig}_{\max}\{\boldsymbol{S}_d \boldsymbol{S}_d^\top\}$
5:      $\boldsymbol{t}_d = \boldsymbol{X}_d \boldsymbol{w}_d$
6:      $\boldsymbol{c}_d = \boldsymbol{Y}_d^\top \boldsymbol{t}_d / \boldsymbol{t}_d^\top \boldsymbol{t}_d$
7:      $\boldsymbol{p}_d = \boldsymbol{X}_d^\top \boldsymbol{t}_d / \boldsymbol{t}_d^\top \boldsymbol{t}_d$
8:      $\boldsymbol{X}_{d+1} = \boldsymbol{X}_d - \boldsymbol{t}_d \boldsymbol{p}_d^\top, \boldsymbol{Y}_{d+1} = \boldsymbol{Y}_d - \boldsymbol{t}_d \boldsymbol{c}_d^\top$
9: **end for**
10: Create matrices $\boldsymbol{W}, \boldsymbol{T}, \boldsymbol{P}$ and $\boldsymbol{C}$ from **(3.14)**
11: Find $\boldsymbol{B}_{\text{PLS}}$ from **(3.16)**

This algorithm will be applied to NFL play-by-play data to extract a model that predicts player productivity for future seasons[8].

Chapter 4

PLS Application to NFL Data

In order to fully grasp the depth of the PLS technique, I have applied it to nine seasons worth of NFL play-by-play data scraped from ryurko's github repository [7]. The play-by-play data displays the time of each play, what type of play was run, the players involved in the play, the result of the play in terms of yardage or any other event including touchdowns and turnovers, and other details about the play. When approaching these datasets, my focus was to observe a player or team's output over an entire season rather than just a play at a time. I then began running analysis on the data sets that presented data in such a fashion seen in Figure 4.1, which presents statistics each team's quarterback from 2009 to 2017. In Figure 4.1, there seem to only be a few variables, so wouldn't I just use MLR to model these data? In actuality, this image is only a small section of the full dataset. The full dataset has 288 observations and 88 variables, which makes MLR a less appropriate technique. Thankfully, one of PLS's prime objectives is to compress the number of explanatory variables. And so, using R, I applied a PLS Regression model to this dataset that resulted in the output seen in Figure 4.2. The model consisted of just one response variable (Total Yardage) and the rest of the dataset as the explanatory variables. I chose twenty components to put into the model in order to as an arbitrary maximum number of components. From the output in Figure 4.2, I can see that the percent variance explained in the response variable begins to plateau at four components; however, there seems to be another plateau at thirteen components as well. In order to better determine what the appropriate number of components are to be used in the model. I plotted the root mean squared error (RMSE) of this model in Figure 4.3. The goal is to minimize the RMSE value without running into the dilemma of over-fitting. To detect over-fitting, the model was constructed with a training dataset consisting of data from 2009 to 2016. The 2017 season was set aside as the test data set for later use. Just as expected, there is a significant drop until the model reaches four components

and another drop at thirteen components. So I am still left wondering what the appropriate number of components for my model should be. I decided to plot the predicted values for each of these models against the measured values. Figure 4.4 represents the side-by-side comparison of these two plots. Although the model with thirteen components has a much tighter fit for the predicted values versus the measured values, I still am unable to properly determine whether or not the model is over-fitted. Because I had trained the model with a subset of the original data set, I now want to compare the percent variance explained by the training set and the test set. According to the model output in Figure 4.2, with four components, the percent variance explained is 0.95. Using R, I predicted the 2017 season yardage for quarterbacks with the test set, and compared that to the actual 2017 statistics and received a correlation of 0.96. Therefore, coupled with the RMSE being so low at four components, the similar correlation values between training and test data sets leads me to believe that my model is capable of accurate predictions without suffering from over-fitting.

Figure 4.3 displays the cross-validation plot of our model. Both CV and adjCV represent the test error of the predictive model, so the lower these values the more accurate our model is. We witness a large dip in both of these estimates at four components so this form of validation has told us that four components produces an accurate model for predicting future data.Finally, in Figure 4.4, I plotted the true values for the 2017 season Total yards and compared them to the predicted values from my model. The predicted values are not exactly aligned with true values, but there is a positive linear relationship that shows our model is quite accurate with predicting future data. When I used MLR to create a linear model for the data set, it formed a model with an R-squared value of 0.9999. At first it seems as this is highly accurate; however, upon observing Figure 4.5 we see that this model suffers from over fitting as the prediction values are exactly one-to-one for the 2017 true values.

Figure 4.1: Partial Representation of Team Passing Season Data

	Season	Team	Player_Name	Attempts	Completions	Drives	Comp
1	2009	SEA	M.Hasselbeck	611	372	173	
2	2009	IND	P.Manning	602	402	160	
3	2009	NE	T.Brady	595	391	166	
4	2009	HOU	M.Schaub	590	399	165	
5	2009	ARI	K.Warner	589	393	181	
6	2009	DET	M.Stafford	581	317	176	
7	2009	ATL	M.Ryan	564	331	158	
8	2009	CHI	J.Cutler	560	342	172	
9	2009	DEN	K.Orton	558	341	174	
10	2009	GB	A.Rodgers	555	360	170	
11	2009	MIN	B.Favre	553	377	165	
12	2009	PHI	D.McNabb	553	335	176	

Figure 4.2: PLS Output of Total Yards against all other variables for quarter-backs

```
Data:    X dimension: 256 86
         Y dimension: 256 1
Fit method: kernelpls
Number of components considered: 20
TRAINING: % variance explained
             1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
X              87.70    97.77    98.97    99.26    99.78    99.88    99.94
Total_Yards    53.33    82.95    93.86    95.10    95.35    95.82    95.96
             8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
X              99.95    99.97    99.98     99.99     99.99     100.0
Total_Yards    96.30    96.47    97.13     98.25     99.11      99.7
             14 comps  15 comps  16 comps  17 comps  18 comps  19 comps
X             100.00    100.00    100.00    100.00    100.00    100.00
Total_Yards    99.77     99.79     99.81     99.82     99.83     99.84
             20 comps
X             100.0
Total_Yards    99.9
```
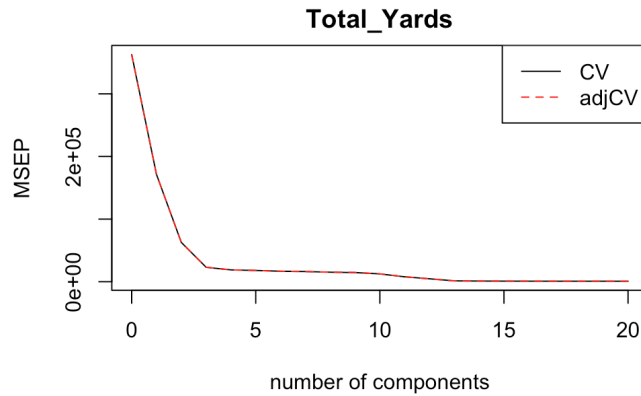
Figure 4.3: RMSE of the PLS model for Quarterbacks

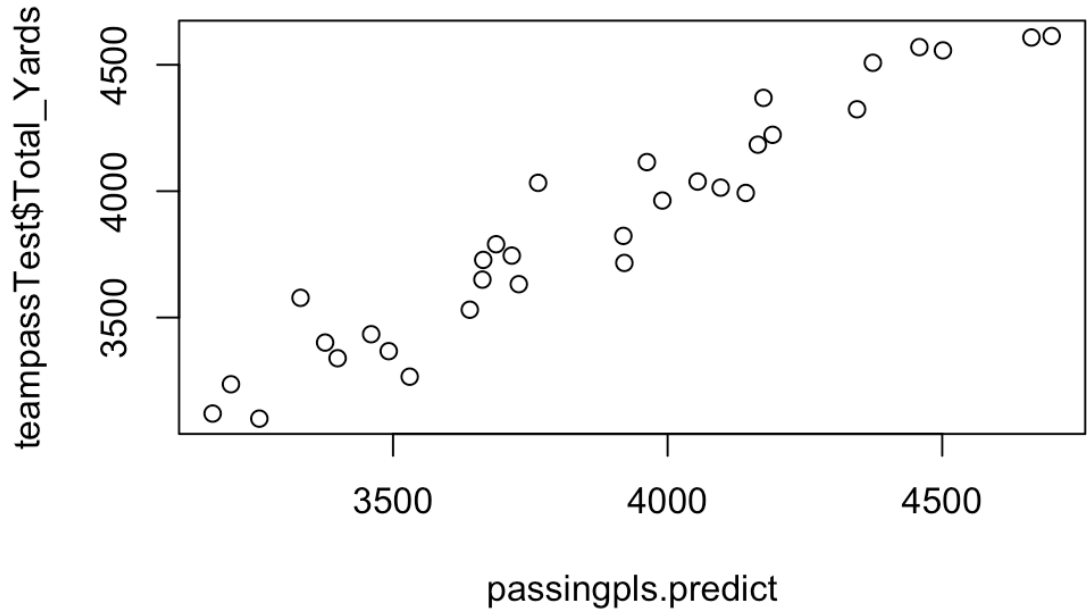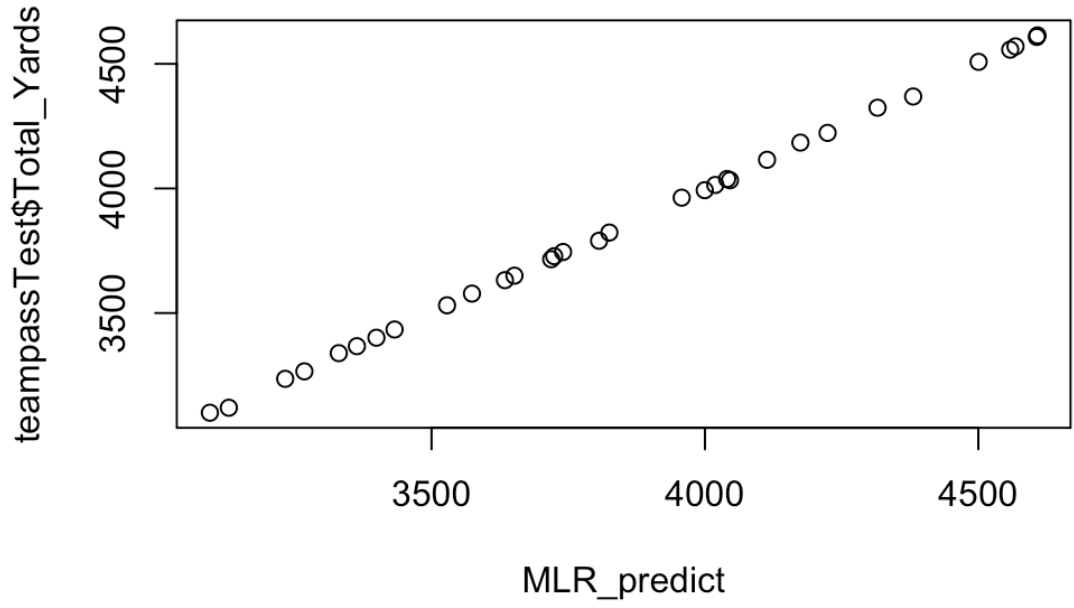Figure 4.4: Quarterback Model with 4 components predictive capability

Figure 4.5: Quarterback Model with 4 components predictive capability with MLR

Chapter 5

Conclusion

Through the data analysis, I was able to create a model to best represent the quarterback passing data over 8 seasons. The appropriate number of components for the model was found to be four components in order to avoid over-fitting while still providing a strong enough model for predictive ability. As a general rule of thumb for PLS regression models, the appropriated number of components for the model can be found by choosing the value with the first significant drop in RMSE value.

In the future, I would like to apply this type of analysis to players including the running back, wide receivers, tight end, etc. I would also like to research the effects that these players have upon each other. In other words, how does a running back's productivity affect that of the quarterback and vice versa. To do so, I would like to create a numerical rating system for each player based off their statistics, and include these values with other player's stat charts to observe the correlation between players on the same team. Currently, there is little data on the effectiveness of the offensive line in these datasets, but I would like to be able to incorporate a method of measuring their efficiency and strength as a variable for the other players as well. One additional metric that must be analyzed is each team's defensive opponents. Clearly, a quarterback will perform at a higher level against weaker defenses, but how do we measure a team's defensive value. Just as there is the quarterback rating(QBR), I would like to create a scale for defenses to be able to account for all factors that influence a player's productivity. In the long run, these models will be used to create a full team depth chart based off of style of play and effectiveness with each play style. The factors that have the most influence on Total Yards or Touchdowns can be isolated using coefficient plots, leading to the areas that need the most attention while practicing or building a team.

In a perfect world (with more time), I would also create models with MLR and PCA. I could then compare the accuracy and analyze the difference between all three modeling techniques. They each have their own advantages and disadvantages with respect to predictive capability, so it is important to understand which data sets require each technique.

Bibliography

[1] Chris Ciovacco. Bullish or bearish stocks? `www.seeitmarket.com/bullish-or-bearish-stocks-what-the-charts-are-saying-14335/`.

[2] ESPN. Rushing yards leaders. `www.espn.com/nfl/statistics/player/-/stat/rushing/sort/rushingYards/seasontype/2`.

[3] Stack Exchange. Do the problems of stepwise variable selection exist in fa, pca, sem? `https://stats.stackexchange.com/questions/231412/do-the-problems-of-stepwise-variable-selection-exist-in-fa-pca-sem`.

[4] Richard Johnson and Dean Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, Inc., 1992.

[5] Guoxin Li. How do i interpret the results of a pca analysis? `https://www.quora.com/How-do-I-interpret-the-results-of-a-PCA-analysis`.

[6] Quora. What are the difference errors and residuals in regression analysis, for example, linear regression?

[7] ryurko. nflscrapr-data. `https://github.com/ryurko/nflscrapR-data`.

[8] Alexander E. Stott, Sithan Kanna, Danilo P. Mandic, and William T. Pike. An online nipals algorithm for partial least squares.