Progress Report I

Writer: Eric Foote

October 11, 2018

Supervisor: Dr. Stewart

Since the proposal presentation, I have made major simplifications to the scope of the project. While I was originally going to consider each players individual accomplishments to the whole team, I decided instead to just look at the just overall team statistics. Work has been done on the linear regression model. I was originally going to build individual models for all thirty one teams. However, there were issues with this approach, once I removed the lockout shortened season each team was only going to have nine data points for 26 variables which will not give an adequate regression line. So instead all of the teams were combined together which gives 270 data points for 26 variables. Then I did an AIC step-wise regression model selector which gave me: an $AIC = -1655.03$ and corresponding model

$$\text{points} = \text{point percentage} + \text{shots per game} + \text{even strength goals against average ratio} + \text{intersect}$$

I have also read into doing a ridge regression for variable selection and have started looking at its implementation in R but I am probably just going to stick to the AIC model instead just simply do to the time. For the linear model I am starting to look at the various plots and see if there is any areas that the model can be improved. For the logistic regression model I have determined that I am going to predict wins instead of points because wins only have two states (i.e. you win or you do not win) while points can not due to a overtime loss giving a team a point. I have not started to look at gradient boosted decision trees at this present time.

The remaining work is to start doing is collecing all of my research on the remaining two topics; finalizing my sources for the linear regression model; wrapping up the linear regression model by using it to predict 2017-2018; begin work on the final two models by reading various implementations or packages in R.