Stat 4320 Project Starcraft 2 Analysis

Andres Gonzalez
Michael Tsui
Ariel De leon

Introduction

- In the multiplayer game Starcraft 2, there is a competitive mode in which players can compete against each other and advance in rank. These rankings are Bronze, Silver, Gold, Platinum,Diamond, Master, Grand Master
- The goal of this project is to create the best predictive model to predict the overall skill of the player base with 18 various predictors to fit into our model.

Description and Ideas

- The data set of Starcraft2 has 3338 observations of 19 variables.
- The response is LeagueIndex, which leaves 18 predictors.
- Seeing that the number of predictors is a handful, we will use three models to predict LeagueIndex.
- Ridge, Lasso, and forward/back selection depending which one performs better in LOOCV.

Methodology

- Split the data into Training Set and Validation set 50/50
- Fit models for Training Set and find LOOCV to compare
- Find MSE of each model and compare results.

Model Selection

- Ridge
- Lasso
- Back Selection or Forward

Forward and Backward Selection

Selection Summary

Variable	AIC	Sum Sq	RSS	R-Sq	Adj. R-Sq
ActionLatency	5038.805	1623.493	1993.335	0.44887	0.44854
AssignToHotkeys	4889.088	1796.702	1820.126	0.49676	0.49616
APM	4819.001	1873.643	1743.185	0.51803	0.51717
MinimapAttacks	4768.671	1927.451	1689.377	0.53291	0.53179
TotalHours	4739.153	1959.054	1657.774	0.54165	0.54027
GapBetweenPACs	4721.649	1978.314	1638.514	0.54697	0.54534
UniqueHotkeys	4715.442	1986.352	1630.477	0.54920	0.54730
WorkersMade	4709.009	1994.569	1622.259	0.55147	0.54931
SelectByHotkeys	4704.675	2000.714	1616.114	0.55317	0.55074
HoursPerWeek	4701.894	2005.336	1611.492	0.55445	0.55176
Age	4698.943	2010.110	1606.718	0.55577	0.55282
NumberOfPACs	4697.599	2013.326	1603.502	0.55666	0.55344
UniqueUnitsMade	4694.904	2017.830	1598.998	0.55790	0.55443
ActionsInPAC	4692.735	2021.819	1595.009	0.55900	0.55527
ComplexAbilitiesUsed	4692.616	2023.843	1592.985	0.55956	0.55557

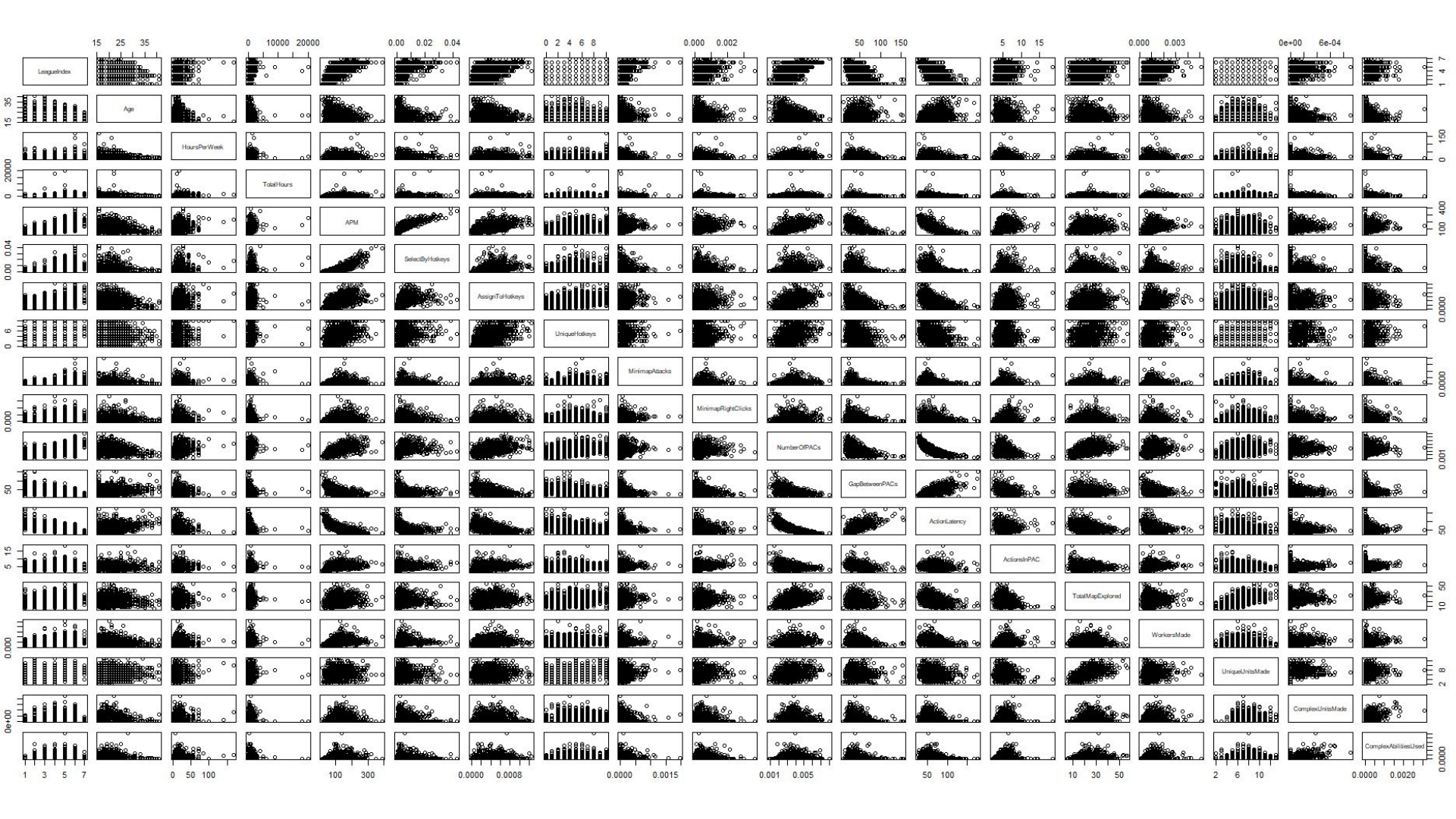
Backward Elimination Summary

Variable	AIC	RSS	Sum Sq	R-Sq	Adj. R-Sq
Full Model	4697.676	1592.088	2024.740	0.55981	0.55501
MinimapRightClicks	4695.690	1592.101	2024.727	0.55981	0.55527
ComplexUnitsMade	4693.981	1592.379	2024.449	0.55973	0.55547
TotalMapExplored	4692.616	1592.985	2023.843	0.55956	0.55557
APM	4692.346	1594.637	2022.191	0.55911	0.55537

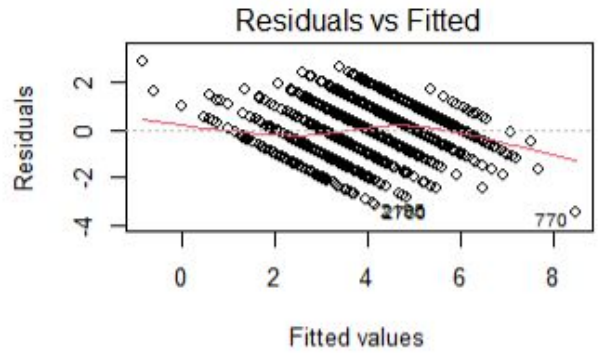
Model selection continued...

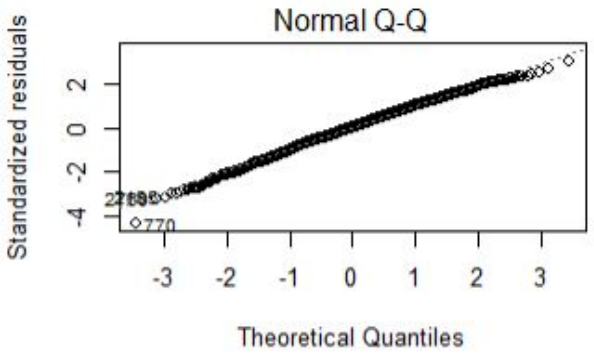
So we had chosen the back selection between full model and forward selection because it had the lowest AIC value.

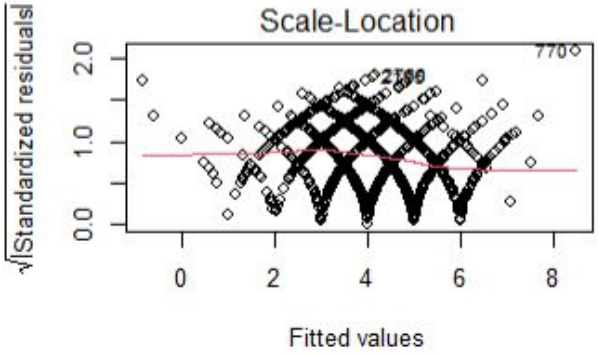
Then we plotted the scatter plot in order to see if there was any obvious relationship between the response LeagueIndex and the other predictors. So that we can transform the model we get from back selection.

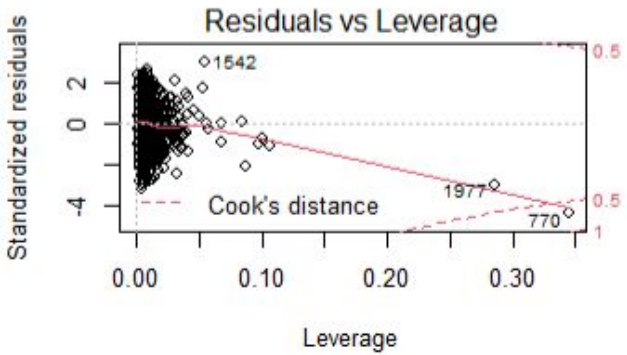

LeagueIndex
Age
HoursPerWeek
TotalHours
APM
SelectByHotkeys
AssignToHotkeys
UniqueHotkeys
MinimapAttacks
MinimapRightClicks
NumberOfPACs
GapBetweenPACs
ActionLatency
ActionsInPAC
TotalMapExplored
WorkersMade
UniqueUnitsMade
ComplexUnitsMade
ComplexAbilitiesUsed

The original residual graphs

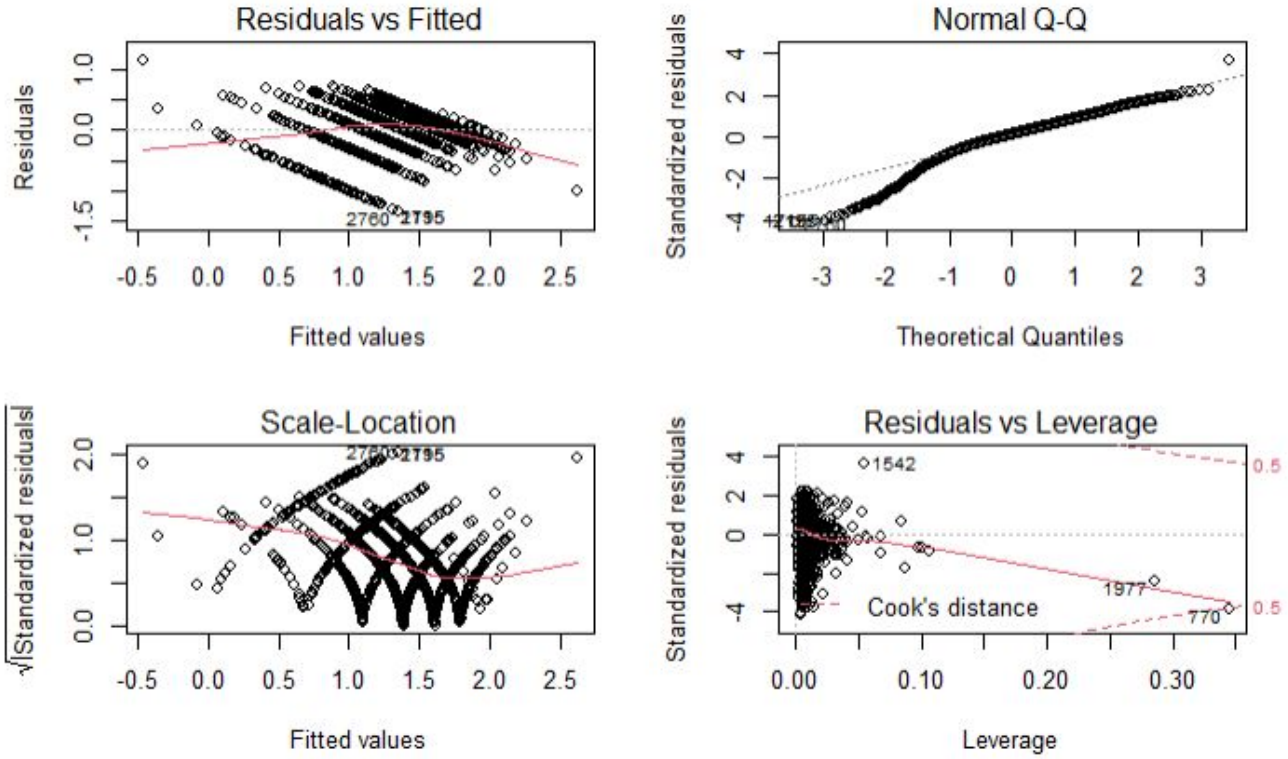




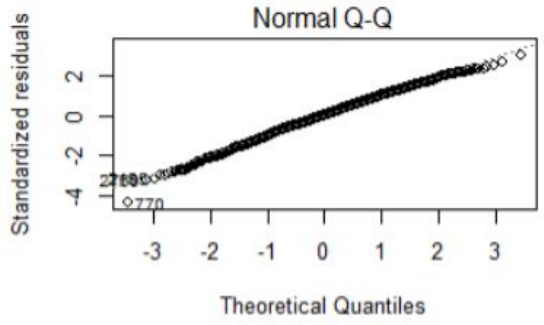


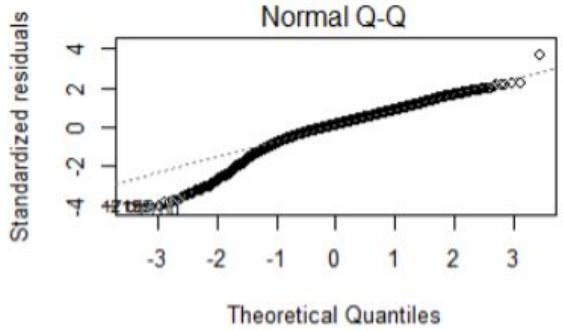


Looking at the scatter plot, we thought some of the patterns resembled a log graph. So we did a log transform on the response and plotted the residual.



From what we observed, the normal distribution actually became worse.





It was at this point we decided to go ahead and do the Ridge, Lasso models and just compare LOOCVs to see which model had the lowest.

Training Set Results

In LOOCV

- Ridge: 0.9872421
- Lasso: 0.9865503
- Back Selection: 0.9848855

- Using LOOCV to compare, we can see that the back selection model won out compared to Ridge and Lasso.

- The next step is to use Mean Squared Error (MSE) in order to test the performance of these three models using the validation set.

Validation Set of MSE

- Ridge: 14.13684
- Lasso: 14.51549
- Back Selection: 14.8740

Conclusion

Based on the results of the MSE, we conclude that the Ridge model performed the best when predicting on the validation set.