Matthew Van Ginneken
EG Data Analyst Assessment Results

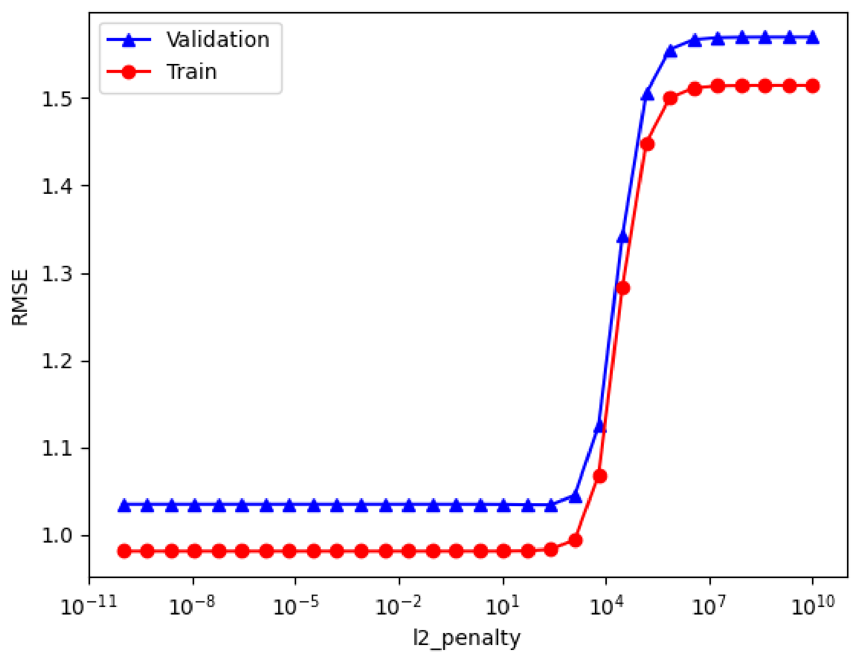
Linear Regression Results:

Training RMSE: 0.9679543303054372
Test RMSE: 1.0250636189639606

Actual Rank	Predicted Rank	Difference
2.0	3.951180	1.951180
3.0	2.225882	-0.774118
4.0	3.228693	-0.771307
4.0	2.563969	1.436031
3.0	3.800691	-0.800691

The plain linear regression model was able to predict most users' ranks within 1 level. While this is useful for predictions, it would be better to reign in the predicted range, and decrease the number of features needed to predict the user’s rank.

Ridge Regression Results:

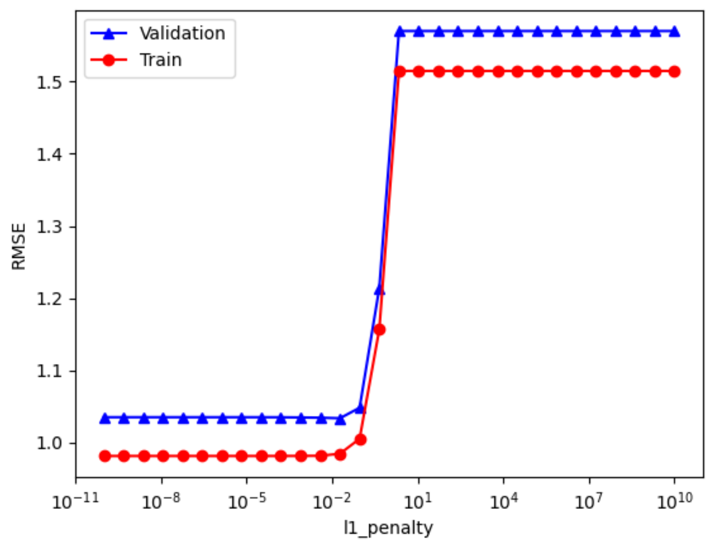


Overall the ridge regression did not provide much detail. It was sufficient to say that using a LASSO regression method would be better at determining what features were contributing most to the rank prediction and should be measured for, rather than weighing all features by a lesser degree.

Matthew Van Ginneken
EG Data Analyst Assessment Results

Lasso Regression Results:

L1 Value	Eliminated Parameters	RMSE	Coefficients
0.452035365636024 03	All but APM, AssignToHotkeys, Action Latency	1.1444942590916345	[-0. 0. 0. 0.29605818 0. 0.02290362 0. 0. 0. 0. -0. -0.35436904 0. 0. 0. 0. 0. 0.]
0.0038566204211634724	APM	0.9439973525969899	[0.0477671 0.07933698 -0.02020145 0. 0.23295588 0.19074395 0.06585008 0.18471124 0.02239838 0.27234628 -0.16651509 -0.42022937 0.03827801 -0.05696083 0.1037875 -0.06146975 0.02960733 0.01755323]
1e-10	None	0.9502969475308995	[0.05541592 0.08900464 -0.02338125 -0.11418892 0.29704978 0.17245369 0.07347909 0.19381757 0.02337418 0.3114932 -0.181024 -0.41636076 0.07363136 -0.0343897 0.11095547 -0.05930946 0.02923639 0.02098292]
0.018873918221350997	ActionsInPAC, ComplexAbilitiesUsed	0.9441268346290297	[0.03172533 0.05952396 -0.00455301 0.0866097 0.17970855 0.19622868 0.07851539 0.17326027 0. 0.17936438 -0.16642151 -0.41772668 0. -0.03384079 0.11893512 -0.0146912 0.00827935 0.0122956]
0.0007880462815669921	None	0.9582422857412952	[0.05488457 0.07687256 -0.02239268 -0.03444495 0.24106913 0.18079823 0.08415542 0.17936062 0.02146682 0.27762728 -0.1756732 -0.42649102 0.05308967 -0.04925757 0.10172604 -0.0605395 0.03144912 0.02070495]



Conclusions:

Across all the LASSO models, the Action Latency, or the time it takes the player to take an action after a perceived input is the strongest predictor of a player's rank. This makes sense, as reaction time to important in game events would make the difference between a successful response and a failure. Other consistently strong predictors of rank across multiple models include SelectByHotKeys, AssignToHotkeys, MinimapAttacks, NumberOfPACs, and GapBetween PACs. All of these features boil down to time being the most important predictor of a rank. Being able to assign hotkeys and use them will make a player respond faster in game. Minimap Attacks will be quicker inputs than regular attacks in game, and the number of PACs and the gap between them both characterize either the reaction time or the response time of the player.

Given the current data I would recommend using the coefficients determined when we eliminated the actions per minute feature. Of the LASSO models it has the lowest RMSE, and it eliminates the need to collect a piece of data which will decrease the storage load. If more features need to be eliminated I would then recommend eliminating age, total hours, minimap right clicks, complex units made, and complex abilities used from consideration as they are very weak predictors of rank. Additionally

Next Steps:

This was a very simple model training scenario. The coefficients determined in the model have yet to be optimized. Cross validation was not done on the data and could provide more valuable insights. Overall the RMSE of all models produced is still relatively large. Some users can be predicted to have a rank that differs by 2 rank categories which can result in some difficult scenarios. Characterization of the models and investigation of their accuracy across ranks should also be done, as these models may favor lower or higher rank predictions.