Pre Live Session Unit 2 Assignment

1. Examine the glmselect output from the first two proc glmselect (labeled M1,M2 in the code) calls and compare them in the following way.
   1. What is different between the two OLS models in terms of the predictors? (note we have tricked glmselct in doing OLS by specifying Forward feature selection with no stopping criterion)

Procedure M1 selected only 7 features as being significant in the model and stopped at step 6 in the forward selection process. M2, selected 27 features as being significant and stopped at step 26. Having more features to choose from, M2 appears to be somewhat overfit, though it’s hard to interpret since it seems like the values of features V12-V31 aren’t given any context that I can find.

* 1. What are the two models R-square values and adjusted R-squared values?

R/R2=.6350/ .6098 for M1

R/R2=.6977/ .5804 for M2

* 1. Examine the Fit criteria and ASE plots. In terms of prediction do you think there is much harm in using all of the predictors versus using a feature selection approach to reduce the model down?

The extra predictors seem to contribute to large overfitting and generally a poorer model. Using the model selected at step 26 when the reduced model was able to achieve similar ASE after only 6 steps is a clear indicator that the extra variables are extraneous and should be eliminated.

1. Compare the second and third proc glmselct calls (M2, M3). These both have the same predictors but one is OLS and the other is using LASSO feature selection using cross validation.
   1. Note the R-squared and Adjusted Rsquared and compare them.

R/R2= .5830 .5564 for M3

R/R2= .6977/ .5804 for M2

* 1. What variables are included using the LASSO as a feature selection technique?

The lasso selection chooses the same abase variables as the simple model, but adds two more – V23/V25. If those are representative of recent scores, perhaps they are a proxy for a player’s momentum?

* 1. Suppose now that I told you that all of the predictors with generic names are just a bunch of random numbers, how does that piece of information potentially change your feeling on whether it matters or not to do feature selection.

I don’t need to be convinced….. see answer to q1 😊

1. Compare the fourth and fifth glmselect calls. These models include interaction terms so the model is even more complex and the potential for overfitting becomes even greater.
   1. In model 5, examine the the CVpress fit criterion panel and compare it to the ASE plot for the test set. Does the CV fit panel mimick the ASE test performance pretty well?

Yes – right around step 7 in the selection proves the CVPRESS is at a minimum, the training ASE is at a saddle/min, and the test ASE is at its absolute min.

* 1. In model 5 that uses the CV approach for feature selection, if we have used Adj-Rsquared rather than CV press, how good would you feel about the predictions you made with that particular model?

Not very good…. In all likelihood, that model includes all the parameters since its peak is at step 30.

Bonus/Critical thinking: When comparing ASE plots of OLS and LASSO from our given code, you may have noticed that OLS seems to yield smaller test error values than LASSO. That may seem contradictory. Why do you think this is happening and why the actual values of the ASE for the OLS and LASSO models we ran are not directly comparable?