Shot quality analysis

The code, 'shot_quality.R', for the analysis can be also found in Github code repository.

The main idea of this analysis is to investigate the factors that can influence the goals making. We consider the Lasso Logistic regression model. For the response, $y = 1$ indicates a goal while $y = 0$ indicates a shot which did not lead to a goal. The predictors are chosen to be distance, type of shots and session. 'Distance', a numeric variable, describes the distance between the place making the shot and the opponent's goal. 'Type of shots', a categorical variable with 9 levels, gives different types of the shots, like 'Wrist', 'Backhand' and etc. Meanwhile, another categorical variable, 'Session', indicates whether the game is in 'Regular' season or 'Playoffs'.

A polynomial of 'distance' up to order 5 is considered in the model. Furthermore, the interaction between the polynomial of 'distance' and 'type of shots' variables are also included in the model. The model matrix is briefly shown as

$$\text{Indicator of goals} \sim Session + (Dis + Dis^2 + Dis^3 + Dis^4 + Dis^5) * Type$$

We have used 10-fold cross validation for variable selection (cv.glmnet in R). λ ranges from $\lambda^1 \approx exp(-8)$ to $\lambda^{100} \approx exp(-30)$. Figure 2 gives the binomial deviance. The AIC selections are the same with the minimum OOS deviance rule, and the BIC selections are to the left of the 1se OOS deviance rule. The AIC and BIC paths are shown in figure 3. Meanwhile, regularization path is given in figure 4. There are 22 non-zero coefficients given by the model following 1se OOS deviance rule, which are shown as follows.

(Intercept)	sessionPlayoffs	1	2	3	4
-2.58971481	-0.08147769	-565.67981418	362.25819810	-12.46277561	-25.03842383
typeBackhand	typeSlap	typeSnap	typeWrist	typeWrap	typeWrap-around
-0.33533948	0.23062800	0.17066213	-0.04837840	-1.58234062	-1.59548554
typeTip-in	typeDeflected	3:typeBackhand	1:typeSlap	2:typeSlap	5:typeSnap
0.23755194	0.35814627	-1.15792930	-28.10104978	-27.68401312	14.88273867
2:typeWrist	3:typeWrist	4:typeWrist	5:typeWrist		
42.93320826	-83.73106544	-13.54603292	36.48986062		

Figure 1: All 22 non-zero coefficients in the model selected by 1se OOS deviance rule. Here 1-4 indicates the 'distance' variable with power 1-4.

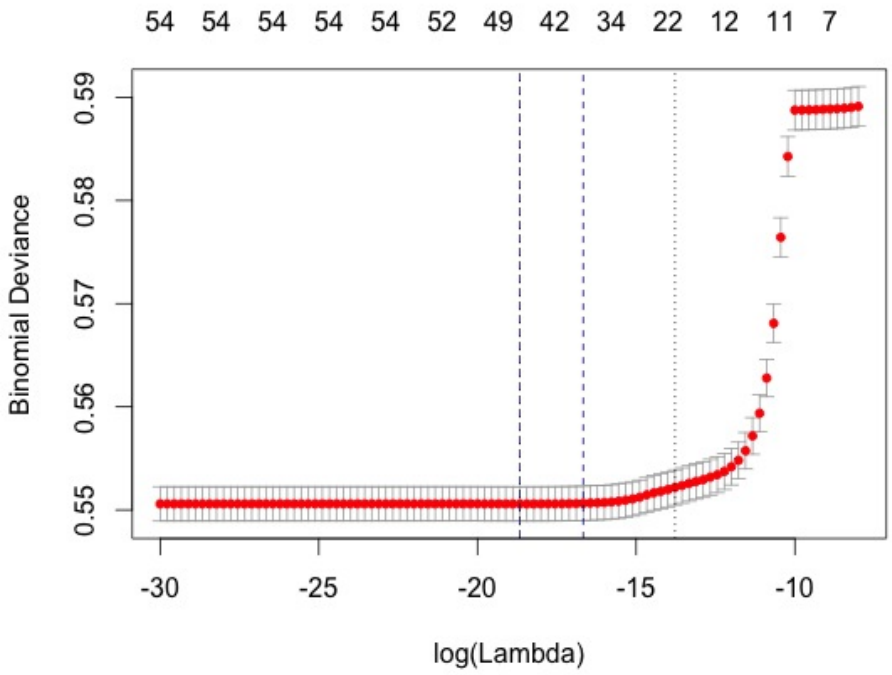
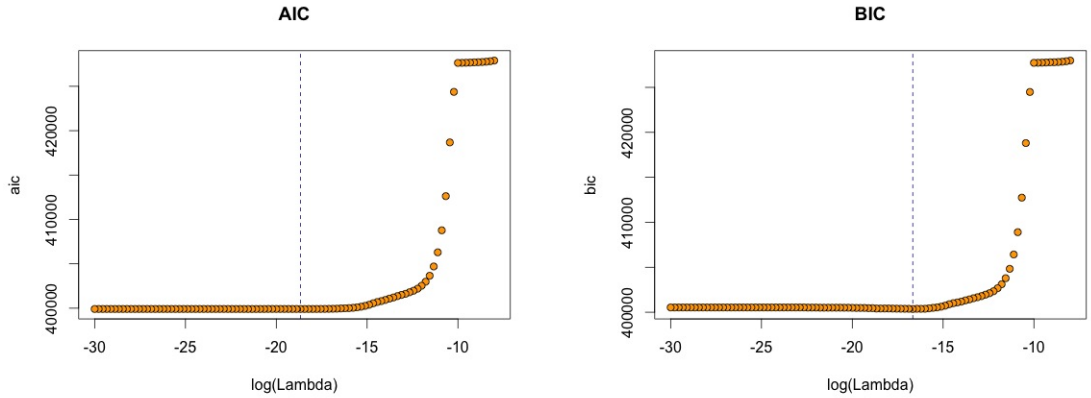


Figure 2: 10-fold CV: mean OOS deviance ± 1 se. Minimum and 1se selection rules are marked with black dotted lines. And minimum AIC and BIC are denoted as dashed blue lines.



(a) AIC (b) BIC

Figure 3: The figure describes the AIC and BIC paths for the model selected by 10-fold cross validation. The 100 $log(\lambda)$ ranges from -8 to -30. The minimum AIC and BIC are denoted by dashed blue lines in the graphs.

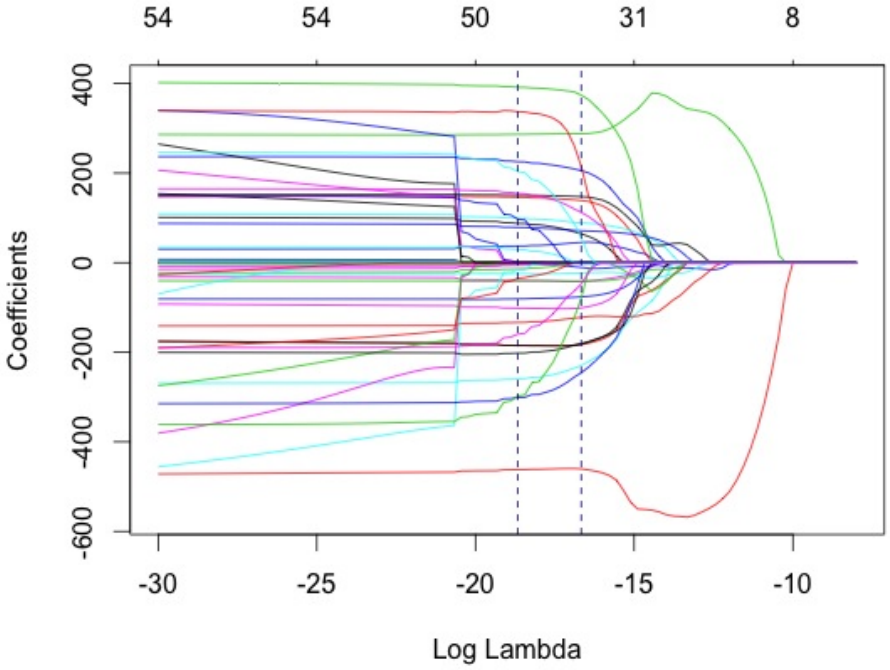


Figure 4: Regularization paths. Minimum AIC and BIC are marked with dashed blue lines.