CAS MA575: Assignment #9

Hu Hengchang(U34497427)
hhc@bu.edu

Boston University — December 10, 2018

1

(a)

From the approach based on all possible subsets, we can see that model with predictors X_1 and X_2 is the optimal model based on R^2_{adj}, AIC, BIC.

(b)

From the approach based on forward selection, we can see that model with predictor X_3 is the optimal model based on AIC and BIC.

(c)

Since model chosen in (a) is global optimal model based on R^2_{adj}, AIC and BIC. But by using forward selection to choose model, the first step we are going to do is to put the most relevant predictor in the model. Therefore, forward selection often produce a local optimal model.

(d)

I would choose model with predictors X_1 and X_2, because this model has smaller AIC and BIC and bigger R^2_{adj}.

3

(a)

From the approach based on all possible subsets, we can see that model with predictors X_2, X_4, X_5, X_6, X_7 is optimal based on R^2_{adj} and AIC & AICc, and model with predictors X_2, X_4, X_6 is optimal based on BIC.

(b)

From the approach based on backward selection, we can see that model with predictors X_2, X_4, X_5, X_6, X_7 is optimal based on AIC, and model with predictors X_2, X_4, X_6 is optimal based on BIC.

(c)

From the approach based on forward selection, we can see that model with predictors X_2, X_4, X_5, X_6, X_7 is optimal based on AIC, and model with predictors X_2, X_4, X_6, X_7 is optimal based on BIC.

(d)

Optimal model in (a) and (c) is different when based on BIC, since forward and backward selection sometimes reach local optimal model. However, the model chosen by all possible subsets always reach the global optimal model. In this case, BIC selection is just more "lucky" than AIC selection

(e)

I would choose model with predictors X_2, X_4, X_6. Because the AIC value and R^2_{adj} of this model is very close to the other model with predictors X_2, X_4, X_5, X_6, X_7, and from the summary of R code, we can see that the predictors of this model are all highly significant. However, the predictors of the other model of size 5 is not all significant. So I would choose model with predictors X_2, X_4, X_6.

(f)

$log(Y) = -11.08314 + 0.15658X_2 + 0.20625X_4 + 0.09178X_6$, where X_2 is GIR, X_4 is BirdieConversion, X_6 is Scrambling.

For example, if X_4 rises 10, the prize money will expand to $\exp(2.06) = 7.846$ times.

I do not think a greater player can rises 10 at X_4. Even one player can rises 10 at X_4, the prize is not possible to expand to that large.

5

(1)

$\mu_1 = E[Y|U_2 = ... = U_d = 0] = \beta_0$

for $j = 2, ..., d$. $\mu_j = E[Y|U_j = 1, U_{others} = 0] = \beta_0 + \beta_j$

(2)

$RSS(\beta) = \sum_{j=1}^{d} \sum_{i=1}^{n_j} (y_{ji} - \beta_0 - \beta_2 U_2 - ... - \beta_d U_d)^2$

for j = 2,...d, $\frac{\partial RSS(\beta)}{\partial \beta_j} = 0 \Rightarrow \sum_{i=1}^{n_j} (y_{ji} - \hat{\beta}_0 - \hat{\beta}_j) = 0 \Rightarrow \bar{y}_j = \hat{\mu}_j$

$\frac{\partial RSS(\beta)}{\partial \beta_0} = 0 \Rightarrow \sum_{i=1}^{n_1} (y_{1i} - \hat{\beta}_0)^2 + \sum_{i=1}^{n_2} (y_{2i} - \hat{\beta}_0 - \hat{\beta}_2) + ... + \sum_{i=1}^{n_d} (y_{di} - \hat{\beta}_0 - \hat{\beta}_d) = 0 \Rightarrow \sum_{i=1}^{n_1} (y_{1i} - \hat{\beta}_0) = 0 \Rightarrow \bar{y}_1 = \hat{\beta}_0 = \hat{\mu}_1$

(3)

$$\begin{aligned}
RSS &= \sum_{j=1}^{d} \sum_{i=1}^{n_j} (y_{ji} - \hat{\beta}_0 - \hat{\beta}_2 U_2 - ... - \hat{\beta}_j U_j - ...)^2 \\
&= \sum_{i=1}^{n_1} (y_{1i} - \hat{\beta}_0)^2 + \sum_{i=1}^{n_2} (y_{2i} - \hat{\beta}_0 - \hat{\beta}_2)^2 + ... + \sum_{i=1}^{n_j} (y_{ji} - \hat{\beta}_0 - \hat{\beta}_j)^2 + ... \\
&= \sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2 + ... + \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j)^2 + ... \\
&= \sum_{j=1}^{d} (n_j - 1) SD_j^2
\end{aligned}$$

$df = n - d$, $SD_j^2 = \frac{\sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j)^2}{n_j - 1}$

(4)

$var(\hat{\beta}_0) = var(\bar{y}_1) = var(\frac{1}{n_1} \sum_{i=1}^{n_1} y_i) = \frac{1}{n_1} var(y)$, since y_i are i.i.d

$var(\hat{\beta}_0 + \hat{\beta}_j) = var(\bar{y}_j) = var(\frac{1}{n_j} \sum_{i=1}^{n_j} y_j) = \frac{1}{n_j} var(y)$, since y_i are i.i.d

Since n_j are equal, $var(\hat{\beta}_0) = var(\hat{\beta}_0 + \hat{\beta}_j) \Rightarrow \sqrt{var(\hat{\beta}_0)} = \sqrt{var(\hat{\beta}_0 + \hat{\beta}_j)} \Rightarrow se(\hat{\beta}_0) = se(\hat{\beta}_0 + \hat{\beta}_j)$

$var(\hat{\beta}_j) = var(\bar{y}_j - \hat{\beta}_0) = var(\bar{y}_j - \bar{y}_1) = var(\bar{y}_j) + var(\bar{y}_1) = \frac{1}{n_j} var(y) + \frac{1}{n_1} var(y) = \frac{2}{n_j} var(y)$

So $se(\hat{\beta}_j)$ are all equal for $j = 2, ...d$.