Quantitative Analysis HW9

R11922045 陳昱行

1.

(a)

The `best subset` model will have the smallest training RSS since it choses the least RSS model among all possibilities.

(b)

It depends on the testing data set and does not have a definite answer.

(c)

True, since we only add predictor to our model in each step and do not eliminate them.

(d)

True, since we have to eliminate a predictor from `k+1` predictor model to get `k` predictor model.

2.

Let $Q(\beta_R) = (y - X\beta_R)'(y - X\beta_R) - \lambda\beta_R'\beta_R$

To minimize Q , we need to solve $\hat{\beta}_R$ for the F.O.C given by:

$$
\begin{aligned}
&-2X'(y - X\hat{\beta}_R) + 2\lambda\hat{\beta}_R = 0 \\
\implies &(X'X + \lambda I)\beta_R = X'y \\
\implies \hat{\beta}_R = (X'X + \lambda I)^{-1}X'y &= \left(\sum_{i=1}^{n} x_i x_i' + \lambda I\right)^{-1}\left(\sum_{i=1}^{n} x_i y_i\right)
\end{aligned}
$$

Note that this model does not contain the intercept term.

3.

(a)

No, as $\lambda \to \infty$, $\beta_{\forall j \neq 0} \to 0$ except the intercept term. Therefore the LASSO regression will predict $\bar{y}$.

(b)

No, none of them should be necessary identical, there might be a coinicidence for which some of them are identical but not guaranteed.