Chapter 6 - Linear Model Selection and Regularization

1. (a) You would expect the best subset selection method to have the smallest training RSS for k predictors, as it is possible that forward and backward stepwise selection may have missed out on certain combinations of predictors due to their iterative nature. However, if forward and backward stepwise selection have the same combination of predictors as best subset selection, then the training RSS will be the same in all three approaches.

   (b) Any of the approaches can produce the subset of predictors which gives the best test set performance. However, best subset selection considers more models and hence has a greater chance of producing the model which gives the lowest test set error.

   (c) i. True

      ii. True

      iii. False

      iv. False

      v. False

2. (a) Lasso regression compared to least squares is less flexible as the values of the coefficients are constrained. It will give improved prediction accuracy when its increase in bias is less than its decrease in variance.

   (b) Ridge regression compared to least squares is less flexible as the values of the coefficients are constrained. It will give improved prediction accuracy when its increase in bias is less than its decrease in variance.

   (c) Non-linear methods compared to least squares are more flexible as the values of the coefficients are no longer constrained. It will give improved prediction accuracy when its increase in variance is less than its decrease in bias.

3. (a) Steadily decreases. This is because adding more predictors allows greater flexibility in the model. When s is increased from 0, more predictors are added to the model leading to overfitting. This leads to lower training RSS.

(b) When there are few predictors, the model underfits the data. When there are many predictors the model will tend to overfit the data. In both these cases test RSS will be really high. However, when only the relevant predictors are included, the model will produce low test RSS. Thus test RSS will initially decrease and then eventually start increasing in a U-shape.

(c) Steadily increases. When s is increased from 0, more predictors are added to the model leading to increased flexibility. As flexibility increases variance increases.

(d) Steadily decreases. When s is increased from 0, more predictors are added to the model leading to increased flexibility. As flexibility increases bias decreases.

(e) Remains constant. The irreducible error is random noise and is independent of the model, hence it is not affected by the number of predictors included.

4. (a) Steadily increases. This is because removing predictors reduces flexibility in the model. When λ is increased from 0, more predictors are removed from the model leading to underfitting. This leads to higher training RSS.

(b) When there are few predictors, the model underfits the data. When there are many predictors the model will tend to overfit the data. In both these cases test RSS will be really high. However, when only the relevant predictors are included, the model will produce low test RSS. Thus test RSS will initially decrease and then eventually start increasing in a U-shape.

(c) Steadily decreases. When λ is increased from 0, more predictors are removed from the model leading to decreased flexibility. As flexibility decreases variance decreases.

(d) Steadily increases. When λ is increased from 0, more predictors are removed from the model leading to decreased flexibility. As flexibility decreases bias increases.

(e) Remains constant. The irreducible error is random noise and is independent of the model, hence it is not affected by the number of predictors included.

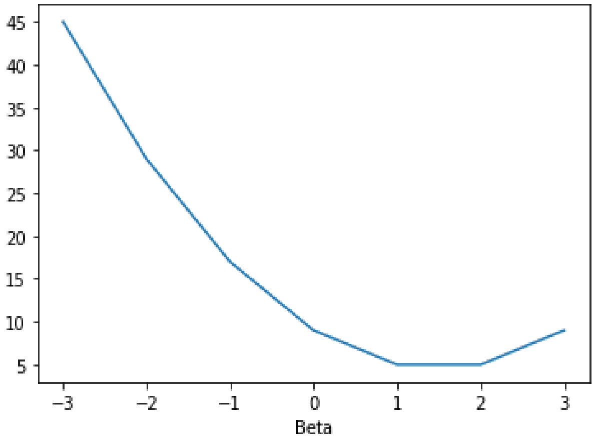
5. (a) $E = (y_1 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_1 x_{21} - \beta_2 x_{22})^2 + \lambda(\beta_1^2 + \beta_2^2)$

(b) $x_{11} = x_{12} = -x_{21} = -x_{22} = a$
$y_1 = -y_2 = b$
$\frac{\partial E}{\partial \beta_1} = 2\lambda\beta_1 - 2(y_1 - \beta_1 x_{11} - \beta_2 x_{12})x_{11} - 2(y_2 - \beta_1 x_{21} - \beta_2 x_{22})x_{21} = 0$
$\lambda\beta_1 = (b - \beta_1 a - \beta_2 a)a - (-b - \beta_1(-a) - \beta_2(-a))(-a)$
$\lambda\beta_1 = 2ab - 2\beta_1 a^2 - 2\beta_2 a^2$
$\frac{\partial E}{\partial \beta_2} = 2\lambda\beta_2 - 2(y_1 - \beta_1 x_{11} - \beta_2 x_{12})x_{12} - 2(y_2 - \beta_1 x_{21} - \beta_2 x_{22})x_{22} = 0$
$\lambda\beta_2 = (b - \beta_1 a - \beta_2 a)a - (-b - \beta_1(-a) - \beta_2(-a))(-a)$
$\lambda\beta_2 = 2ab - 2\beta_1 a^2 - 2\beta_2 a^2$
Thus $\beta_1 = \beta_2$.

(c) $E = (y_1 - \beta_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_1 x_{21} - \beta_2 x_{22})^2 + \lambda(|\beta_1| + |\beta_2|)$

(d) $x_{11} = x_{12} = -x_{21} = -x_{22} = a$
$y_1 = -y_2 = b$
$\frac{\partial E}{\partial \beta_1} = \pm\lambda - 2(y_1 - \beta_1 x_{11} - \beta_2 x_{12})x_{11} - 2(y_2 - \beta_1 x_{21} - \beta_2 x_{22})x_{21} = 0$
$\pm\lambda = 2(b - \beta_1 a - \beta_2 a)a - 2(-b - \beta_1(-a) - \beta_2(-a))(-a)$
$\pm\lambda = 4ab - 4\beta_1 a^2 - 4\beta_2 a^2$
$\frac{\partial E}{\partial \beta_2} = \pm\lambda - 2(y_1 - \beta_1 x_{11} - \beta_2 x_{12})x_{11} - 2(y_2 - \beta_1 x_{21} - \beta_2 x_{22})x_{21} = 0$
$\pm\lambda = 2(b - \beta_1 a - \beta_2 a)a - 2(-b - \beta_1(-a) - \beta_2(-a))(-a)$
$\pm\lambda = 4ab - 4\beta_1 a^2 - 4\beta_2 a^2$

6. (a) $E_R = \sum_{j=1}^{p}(y_j - \beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$
Given $p = 1$, $E_R = (y_1 - \beta_1)^2 + \lambda\beta_1^2$
Say $\lambda = 1$ and $y_1 = 3$, $E_R = 2\beta_1^2 - 6\beta_1 + 9$
Now the ridge regression coefficient is given by $\beta_1 = \frac{y_1}{1+\lambda} = 1$.



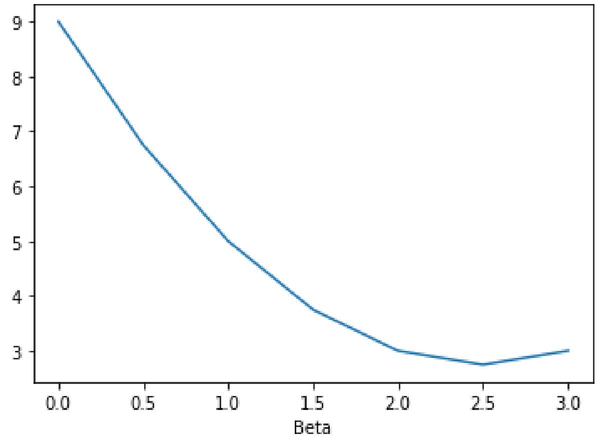
(b) $E_R = \sum_{j=1}^{p} (y_j - \beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$

Given $p = 1,\ E_R = (y_1 - \beta_1)^2 + \lambda|\beta_1|$

Say $\lambda = 1$ and $y_1 = 3,\ E_R = \beta_1^2 - 6\beta_1 + 9 + |\beta_1|$

Now the lass regression coefficient is given by $\beta_1 = \begin{cases} y_1 - \lambda/2 & if y_1 > \lambda/2; \\ y_1 + \lambda/2 & if y_1 < -\lambda/2; \\ 0 & if |y_1| \leq \lambda/2. \end{cases}$

Now since $y_1 = 3$ is greater than $\lambda/2 = 0.5,\ \beta_1 = y_1 - \lambda/2 = 2.5$.


Beta