PARTH RATHOD

CSP 571 - DPA

A20458817

FALL 2021

Recitation Answers

Chapter 6

Question 1

a) The Best Subset selection will have the smallest training RSS as it will consider all the possible models unlike the others which has a greedy approach.

b) The Best subset selection model will have the high chances of choosing a model with less test RSS as it contains 2^p models whereas as the other two models will consider only (1+ $p(p$+1))/2 models.

c) i) TRUE :- because the (k+1) variable model has one more predictor in addition to all of the predictors chosen for the k-variable model in the forward stepwise selection.
ii) TRUE :- because in the backward stepwise selection, k-variable model is obtained by removing one predictor from (k+1)-variable model which will reduce the RSS of the model.
iii) FALSE :- because both the models follow different criteria.
iv) FALSE :- because both the models follow different criteria. Also, there is not link between the models obtained from forward and backward model.
v) FALSE :- because the best subset approach selects the model with (k+1) predictors from among all feasible models with (k+1) predictors. As a result, it does not ensure that the same predictors will be used for the k predictor model.

Question 2

a) Option (iii) is correct.
Because Regularization reduces the Test MSE by adding a penalty by decreasing variance and increasing bias. This penalty shrinks the coefficient and slop gets less steep.
b) Option (iii) is correct.
Ridge reduces predictors that do not have a significant link with the target variable, making them less flexible. It also decreases variance as the expense of increasing bias. To enhance prediction accuracy, the increase in bias should be smaller than the decrease in variance.
c) Option (ii) is correct.
Because non-linear techniques are more flexible than least squares, they may provide more accurate predictions.

Question 3

```
a) Option (iv) is correct.
As we increase s, the model becomes more and more flexible as the
restriciton on beta is reducing, thus the coefficients increase from 0 to
their least square estimate values. Thus, resulting in decreased RSS.
b) Option (ii) is correct.
As model is becoming more and more flexible the test RSS wil reduce first
and then start increasing when overfitting will start.
c) Option (iii) is correct.
Variance steadily increase with increase in model flexibility.
d) Option (iv) is correct.
Bias decreases with increase in model flexibility.
e) Option (v) is correct.
Irreducible error is model independent and does not depend on s.
```

Question 4

```
a) Option (iii) is correct.
As we increase lambda, the model becomes less and less flexible as the
retriciton on beta is increasing, thus the coefficients come close to 0
from their least square estimate values. Thus, resulting in increased
RSS.
b) Option (ii) is correct.
As the model is becoming less and less flexible the test RSS will  reduce
first and then start increasing when overfitting will start.
c) Option (iv) is correct.
Variance steadily decreases with the decrease in model flexibility.
d) Option (iii) is correct.
Bias increases with decrease in model flexibility.
e) Option (v) is correct.
Irreducible error is model independent and does not depend on lambda.
```

Question 5

a)

Question 5

a) Ridge Regression is given by:-

Minimize :- $\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_j \right)^2 + \lambda \sum_{i=1}^{p} \hat{\beta}_j^{2}$

here

$n = p = 2$ & $\hat{\beta}_0 = 0$

$\therefore \min \rightarrow \left[(y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_{21} x_{22})^2 \right]$
$+ \lambda (\beta_1^2 + \beta_2^2)$

b)

b) ∴ On expanding above Equation.

$\rightarrow (y_1^2 + \hat{\beta}_1^2 x_{11}^2 + \hat{\beta}_2^2 x_{12}^2 - 2\hat{\beta}_1 x_{11} y_1 - 2\hat{\beta}_2 x_{12} y_1 + 2\hat{\beta}_1 \hat{\beta}_2 x_{11} x_{12})$
$+ (y_2^2 + \hat{\beta}_1^2 x_{21}^2 + \hat{\beta}_2^2 x_{22}^2 - 2\hat{\beta}_1 x_{21} y_2 - 2\beta_2 x_{22} y_2 + 2\beta_1 \beta_2 x_{21} x_{22})$
$+ \lambda \beta_1^2 + \lambda \beta_2^2$

Take dairative & equate it to zero.

$\therefore \frac{\partial}{\partial \hat{\beta}_1} = 0$

Also,

$x_{11} = x_{12} = x$, & $x_{21} = x_{22} = x_2$ & Divide by 2.

$$\rightarrow (\hat{\beta}_1 x_1^2 - x_1 y_1 + \hat{\beta}_2 x_1^2) + (\hat{\beta}_1 x_2^2 - x_2 y_2 + \hat{\beta}_2 x_2^2) + \lambda \beta_1 = 0.$$

$$\rightarrow \hat{\beta}_1 (x_1^2 + x_2^2) + \beta_2 (x_1^2 + x_2^2) + \lambda \hat{\beta}_1 = x_1 y_1 + x_2 y_2$$

Add $2\hat{\beta}_1 x_1 x_2$ and $2\beta_1 x_1 x_2$ on both sides.

$$\rightarrow \hat{\beta}_1 (x_1 + x_2)^2 + \beta_2 (x_1 + x_2)^2 + \lambda \beta_1 = x_1 y_1 + x_2 y_2 \quad \ldots (1)$$

As $x_1 + x_2 = 0$

$$\therefore \lambda \hat{\beta}_1 = x_1 y_1 + x_2 y_2 + 2\hat{\beta}_1 x_1 x_2 + 2\hat{\beta}_2 x_1 x_2 \quad \ldots (1)$$

Similarly taking partial derivative wrt $\hat{\beta}_{21}$ we get.

$$\lambda \hat{\beta}_2 = x_1 y_1 + x_2 y_2 + 2\hat{\beta}_1 x_1 x_2 + 2\hat{\beta}_2 x_1 x_2 \quad \ldots (2)$$

From (1) & (2)

$$\underline{\beta_1 = \beta_2}$$

c)

c] $\min\left[(y_1 - \hat{\beta}_1 x_{11} - \beta_2 x_{12})^2 + (y_2 - \beta_1 x_{21} - \beta_2 x_{22})^2\right]$

$+ \lambda (|\hat{\beta}_1| + |\hat{\beta}_2|)$

d)

d) Replacing penalty term from Ridge Regression

$\therefore$ the derivative term.

$$= \frac{\partial}{\partial \hat{\beta}} (\lambda |\beta|) = \frac{\lambda |\beta_1|}{\beta}$$

Same like ridge regression we get,

$$\frac{\lambda |\beta_1|}{\beta_1} = \frac{\lambda |\beta_2|}{\beta_2}$$

Provide β_1 & β_2 are both positive or both negative.