HW week2

[#1] Forward Stepwise Selection

```
library(mlbench)
library(leaps)
data(BostonHousing)
head(BostonHousing)
```

```
##      crim zn indus chas   nox    rm  age    dis rad tax ptratio      b lstat
## 1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90  4.98
## 2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8 396.90  9.14
## 3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242    17.8 392.83  4.03
## 4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222    18.7 394.63  2.94
## 5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7 396.90  5.33
## 6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222    18.7 394.12  5.21
##   medv
## 1 24.0
## 2 21.6
## 3 34.7
## 4 33.4
## 5 36.2
## 6 28.7
```

```
dim(BostonHousing)
```

```
## [1] 506  14
```

```
#set 'medv' variable as dependent variable
regfitfwd=regsubsets(medv ~.,data=BostonHousing,method ="forward")
summary(regfitfwd)
```

```
## Subset selection object
## Call: regsubsets.formula(medv ~ ., data = BostonHousing, method = "forward")
## 13 Variables  (and intercept)
##         Forced in Forced out
## crim        FALSE      FALSE
## zn          FALSE      FALSE
## indus       FALSE      FALSE
## chas1       FALSE      FALSE
## nox         FALSE      FALSE
## rm          FALSE      FALSE
## age         FALSE      FALSE
## dis         FALSE      FALSE
## rad         FALSE      FALSE
## tax         FALSE      FALSE
## ptratio     FALSE      FALSE
## b           FALSE      FALSE
## lstat       FALSE      FALSE
## 1 subsets of each size up to 8
```

```
## Selection Algorithm: forward
##          crim zn  indus chas1 nox rm  age dis rad tax ptratio b   lstat
## 1  ( 1 ) " "  " " " "   " "   " " " " " " " " " " " "     " " "*"
## 2  ( 1 ) " "  " " " "   " "   " " "*" " " " " " " " "     " " "*"
## 3  ( 1 ) " "  " " " "   " "   " " "*" " " " " " " " "*"     " " "*"
## 4  ( 1 ) " "  " " " "   " "   " " "*" " " "*" " " " "*"     " " "*"
## 5  ( 1 ) " "  " " " "   " "   "*" "*" " " "*" " " " "*"     " " "*"
## 6  ( 1 ) " "  " " " "   "*"   "*" "*" " " "*" " " " "*"     " " "*"
## 7  ( 1 ) " "  " " " "   "*"   "*" "*" " " "*" " " " "*"     "*" "*"
## 8  ( 1 ) " "  "*" " "   "*"   "*" "*" " " "*" " " " "*"     "*" "*"
```

What is new variable under the model with 4 variables?

dis

What is the estimated regression coefficient of that variable and the R^2 value of that model?

```
coef(regfitfwd,4)
```

```
## (Intercept)          rm         dis     ptratio       lstat
##  24.4713576   4.2237922  -0.5519263  -0.9736458  -0.6654360
```

```
summary(regfitfwd)$adjr2[4]
```

```
## [1] 0.6878351
```

-0.5519263 and 0.6878351

Ex. 7.4 Consider the in-sample prediction error (7.18) and the training error $\overline{\text{err}}$ in the case of squared-error loss:

$$\begin{aligned} \text{Err}_{\text{in}} &= \frac{1}{N}\sum_{i=1}^{N} \text{E}_{Y^0}(Y_i^0 - \hat{f}(x_i))^2 \\ \overline{\text{err}} &= \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{f}(x_i))^2. \end{aligned}$$

Add and subtract $f(x_i)$ and $\text{E}\hat{f}(x_i)$ in each expression and expand. Hence establish that the average optimism in the training error is

$$\frac{2}{N}\sum_{i=1}^{N}\text{Cov}(\hat{y}_i, y_i),$$

as given in (7.21).

$$\begin{aligned} w = E_y(op) &= E_y(Err_{in} - \overline{err}) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left[E_y E_Y(Y_i^{0\,2} - 2Y_i^0\hat{f}(x_i) + \hat{f}(x_i)^2) \right. \\ &\qquad \left. - E_y(y_i^2 - 2y_i\hat{f}(x_i) + \hat{f}(x_i)^2) \right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\left[2E_y(y_i\hat{f}(x_i)) - 2E_y(\hat{f}(x_i))E_Y(Y_i^0) \right] \\ &= \frac{2}{N}\sum_{i=1}^{N} Cov(y_i, \hat{f}(x_i)) \end{aligned}$$