

Week2_김민회HW

```
library(mlbench)
library(leaps)
```

Bostonhousing 데이터 정보 (<https://rdrr.io/cran/mlbench/man/BostonHousing.html>)

```
data("BostonHousing")
dt = BostonHousing
head(dt)
```

```
##      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
```

```
sum(is.na(dt))
```

```
## [1] 0
```

```
dim(dt)
```

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

결측값 없으므로 바로 forward selection 수행

```
fwd = regsubsets(medv~., data=dt, nvmax=4, method='forward')
summary(fwd)
```

```
## Subset selection object
## Call: regsubsets.formula(medv ~ ., data = dt, nvmax = 4, 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 4
## Selection Algorithm: forward
##           crim zn  indus chas1 nox rm  age dis rad tax ptratio b  lstat
## 1  ( 1 ) " " " " " " " " " " " " " " " " " " " " " " " " "*"
## 2  ( 1 ) " " " " " " " " " " "*" " " " " " " " " " " " " "*"
## 3  ( 1 ) " " " " " " " " " " "*" " " " " " " " " "*" " " "*"
## 4  ( 1 ) " " " " " " " " " " "*" " " "*" " " " " "*" " " "*"

```

새롭게 선택된 변수는 'dis'이다.

```
coef(fwd, 4)
```

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

```

```
summary(fwd)$rsq
```

```
## [1] 0.5441463 0.6385616 0.6786242 0.6903077

```

'dis' 변수의 계수는 -0.5519이며 모델의 결정계수는 0.6903이다.