Compulsory exercise 2: Group 5

TMA4268 Statistical Learning V2022

Aleksander Johnsen Solberg and Gjermund Oscar Lyckander

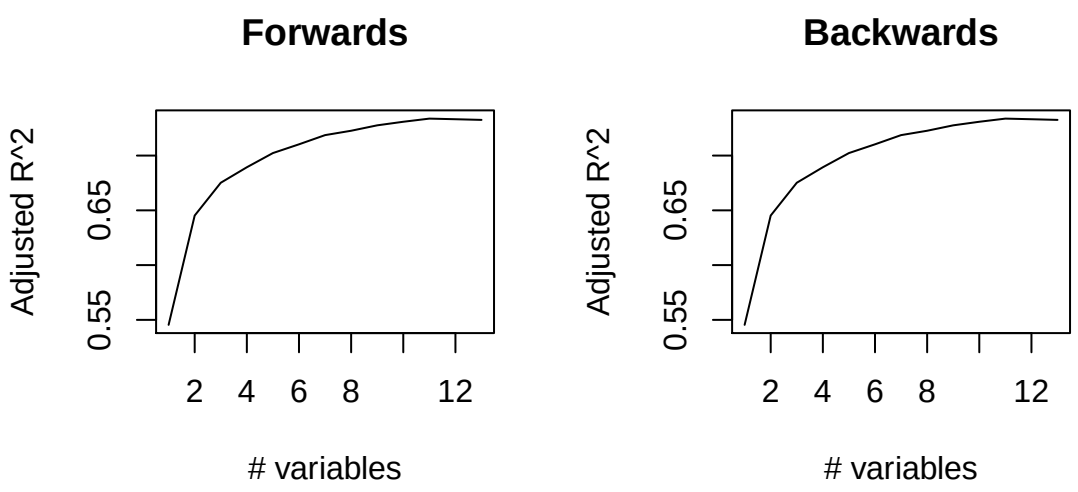
02 April, 2022

Problem 1

```
set.seed(1)
boston <- scale(Boston, center = T, scale =  T)

train.ind = sample(1:nrow(boston), 0.8 * nrow(boston))
boston.train = data.frame(boston[train.ind, ])
boston.test = data.frame(boston[-train.ind, ])
```

a)

```
set.seed(1)
forward_stepwise = regsubsets(medv ~ ., data = boston.train, nvmax = 13, method = 'forward')
backward_stepwise = regsubsets(medv ~ ., data = boston.train, nvmax = 13, method = 'backward')

forward_stepwise_summary = summary(forward_stepwise)
backward_stepwise_summary = summary(backward_stepwise)
#forward_stepwise_summary
#backward_stepwise_summary

par(mfrow=c(1,2))
plot(forward_stepwise_summary$adjr2, xlab = '# variables', ylab = 'Adjusted R^2', type='l', main='Forwa
plot(forward_stepwise_summary$adjr2, xlab = '# variables', ylab = 'Adjusted R^2', type='l', main='Backwa
```



b)

```
forward_stepwise_summary$outmat
```

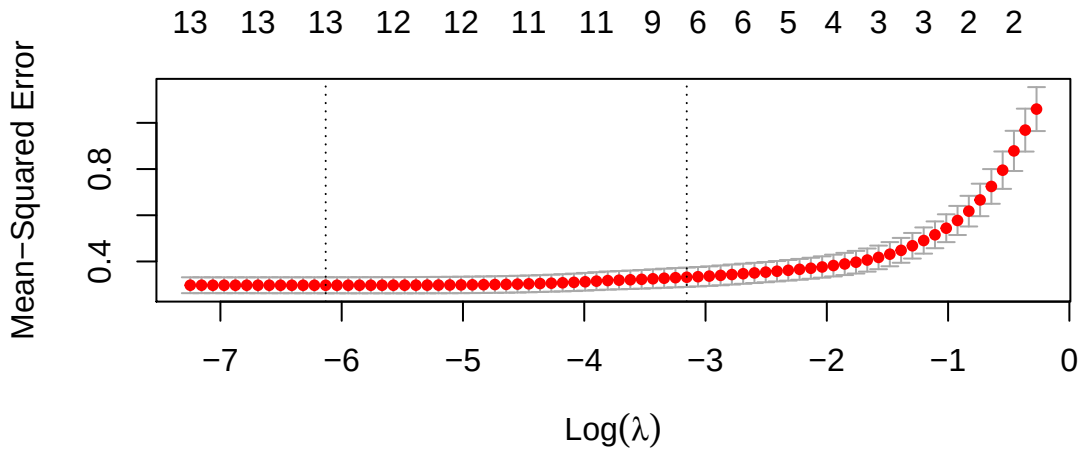
```
##           crim zn  indus chas nox rm  age dis rad tax ptratio black lstat
## 1  ( 1 )  " "  " " " "   " "  " " " " " " " " " " " "     " "   "*"
## 2  ( 1 )  " "  " " " "   " "  " " "*" " " " " " " " "     " "   "*"
## 3  ( 1 )  " "  " " " "   " "  " " "*" " " " " " " " "*"     " "   "*"
## 4  ( 1 )  " "  " " " "   " "  " " "*" " " "*" " " " " "*"     " "   "*"
## 5  ( 1 )  " "  " " " "   " "  " " "*" " " "*" " " " " "*"     "*"   "*"
## 6  ( 1 )  " "  " " " "   " "  "*" "*" " " "*" " " " " "*"     "*"   "*"
## 7  ( 1 )  " "  " " " "   "*"  "*" "*" " " "*" " " " " "*"     "*"   "*"
## 8  ( 1 )  " "  " " " "   "*"  "*" "*" " " "*" "*" " " "*"     "*"   "*"
## 9  ( 1 )  " "  " " " "   "*"  "*" "*" " " "*" "*" "*" "*"     "*"   "*"
## 10  ( 1 ) " "  "*" " "   "*"  "*" "*" " " "*" "*" "*" "*"     "*"   "*"
## 11  ( 1 ) "*"  "*" " "   "*"  "*" "*" " " "*" "*" "*" "*"     "*"   "*"
## 12  ( 1 ) "*"  "*" "*"   "*"  "*" "*" " " "*" "*" "*" "*"     "*"   "*"
## 13  ( 1 ) "*"  "*" "*"   "*"  "*" "*" "*" "*" "*" "*" "*"     "*"   "*"
```

We choose the predictors 'rm, 'dis', 'ptratio' and 'lstat'.

c)

i)

```
set.seed(1)
y = boston.train$medv
x = data.matrix(boston.train[, -14])

cv_lasso = cv.glmnet(x, y, alpha=1, nfolds=5)
plot(cv_lasso)
```



ii)

```
lasso_best_lambda = cv_lasso$lambda.min
lasso_best_lambda
```

```
## [1] 0.002172032
```

iii)

```
coef(glmnet(x, y, alpha=1, lambda=lasso_best_lambda))
```

```
## 14 x 1 sparse Matrix of class "dgCMatrix"
##                       s0
## (Intercept)  0.023622904
## crim        -0.081992849
## zn           0.094717791
## indus        0.002619428
## chas         0.087341100
## nox         -0.175365927
## rm           0.312648954
## age         -0.011212120
## dis         -0.317143728
## rad          0.270168177
## tax         -0.207314714
## ptratio     -0.204052488
## black        0.102877803
## lstat       -0.428298373
```

d)

i) TRUE

ii) FALSE

iii) FALSE

iv) TRUE

Problem 2

a)

b)

Problem 3

a)

i) TRUE

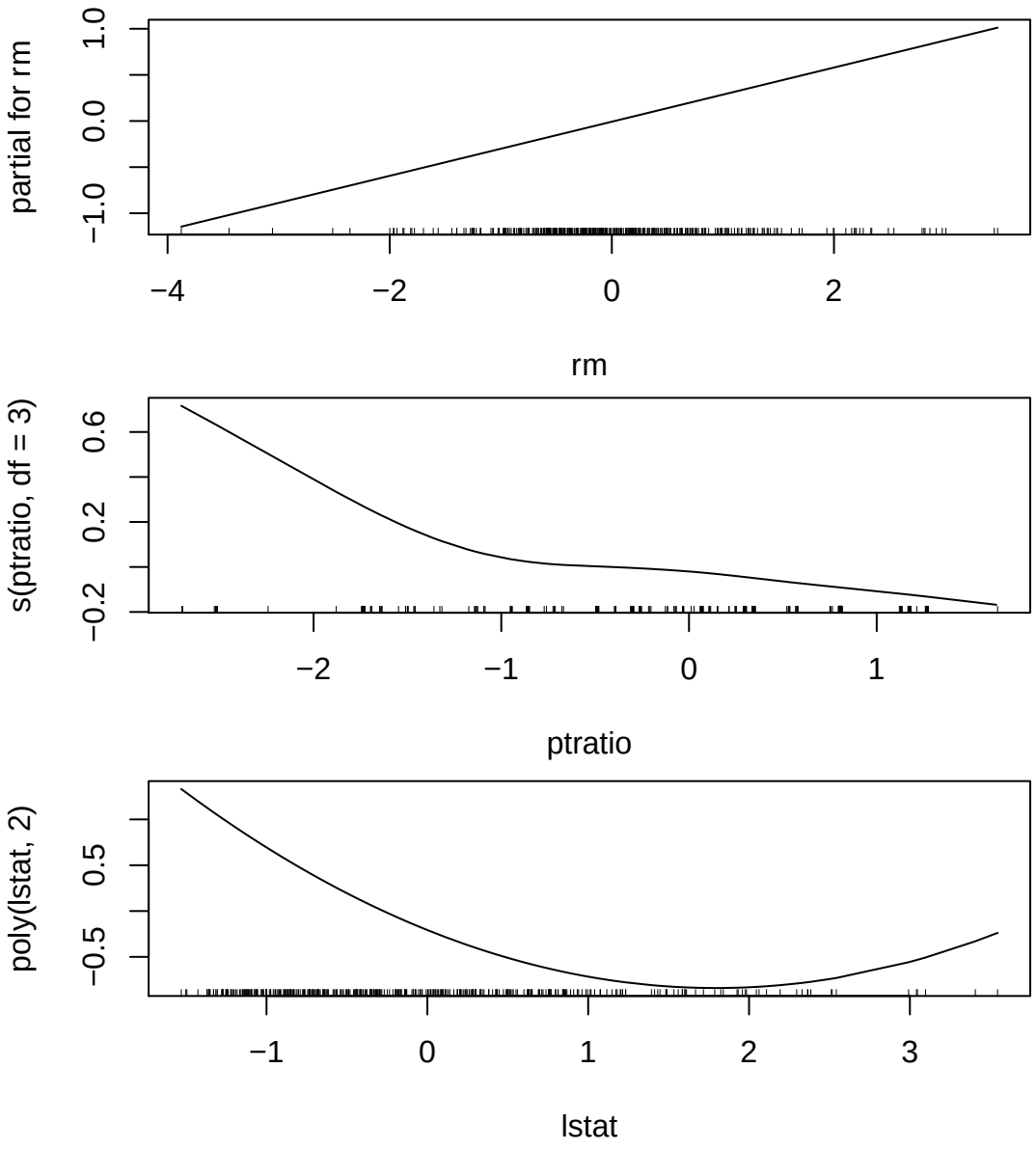
ii) FALSE

iii) FALSE

iv) TRUE

b)

```
additive_model = gam(medv ~ rm + s(ptratio, df=3) + poly(lstat, 2), data=boston.train)
plot(additive_model)
```



Problem 4

a)

i) FALSE

ii) TRUE

iii) TRUE

iv) TRUE

b)

c)

testest

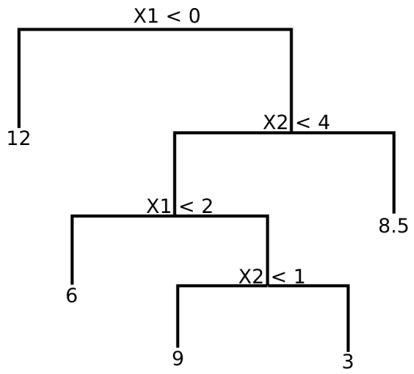


Figure 1: Tree