hw8.R

Muhammad

2021-10-20

library(MASS)  
library(glmnet)

## Loading required package: Matrix

## Loaded glmnet 4.1-2

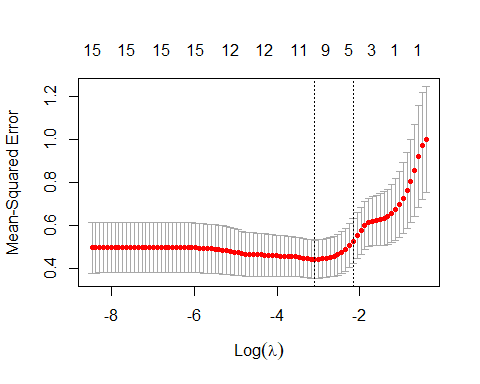
setwd("C:/Users/Muhammad/ISYE/hw8")  
data <- read.table("uscrime.txt", stringsAsFactors = FALSE, header = TRUE)  
  
#Q11.1 1:  
  
reg1 <- lm(Crime ~., data = data)  
  
#Training for stepwise regression, validation through AIC   
stepwise <- stepAIC(reg1, direction = "both", trace = FALSE)  
  
#Best model using stepwise regression using both backward and forward selection  
summary(stepwise)

##   
## Call:  
## lm(formula = Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob,   
## data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -444.70 -111.07 3.03 122.15 483.30   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6426.10 1194.61 -5.379 4.04e-06 \*\*\*  
## M 93.32 33.50 2.786 0.00828 \*\*   
## Ed 180.12 52.75 3.414 0.00153 \*\*   
## Po1 102.65 15.52 6.613 8.26e-08 \*\*\*  
## M.F 22.34 13.60 1.642 0.10874   
## U1 -6086.63 3339.27 -1.823 0.07622 .   
## U2 187.35 72.48 2.585 0.01371 \*   
## Ineq 61.33 13.96 4.394 8.63e-05 \*\*\*  
## Prob -3796.03 1490.65 -2.547 0.01505 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 195.5 on 38 degrees of freedom  
## Multiple R-squared: 0.7888, Adjusted R-squared: 0.7444   
## F-statistic: 17.74 on 8 and 38 DF, p-value: 1.159e-10

#Data too small otherwise I would split the data into training, validation and testing set  
  
#sacling data  
data\_scaled <- scale(data, center = TRUE, scale = TRUE)  
x<- data\_scaled[,1:15]  
y<- data\_scaled[,16]  
  
# Q11.1.2  
fit\_lasso <- glmnet(as.matrix(x), as.matrix(y), family="mgaussian", alpha = 1)  
summary(fit\_lasso)

## Length Class Mode   
## a0 99 -none- numeric  
## beta 1485 dgCMatrix S4   
## df 99 -none- numeric  
## dim 2 -none- numeric  
## lambda 99 -none- numeric  
## dev.ratio 99 -none- numeric  
## nulldev 1 -none- numeric  
## npasses 1 -none- numeric  
## jerr 1 -none- numeric  
## offset 1 -none- logical  
## call 5 -none- call   
## nobs 1 -none- numeric

cv.glm <- cv.glmnet(as.matrix(x), as.matrix(y), alpha=1)  
plot(cv.glm)



#best value of lambda:  
best\_lambda <- cv.glm$lambda.min  
best\_lambda

## [1] 0.04580907