Untitled

Harish Kunaparaju

2023-03-05

```
library(ISLR)
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-6
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

QB1. Build a Lasso regression model to predict Sales based on all other attributes (“Price”, “Advertising”, “Population”, “Age”, “Income” and “Education”). What is the best value of lambda for such a lasso model?

using Carseats dataset to predict the sales.

```
Data<-Carseats
set.seed(333)
```

#To select specific columns from a data set’Carseats’and creating a new database called’Carseats_filtered with those coulmns to selected coulumns showns to be ’sales,price,advertising,population,age,income, and education.

```
Carseats_Filtered <- Carseats %>% select("Sales", "Price",
"Advertising","Population","Age","Income","Education")
```

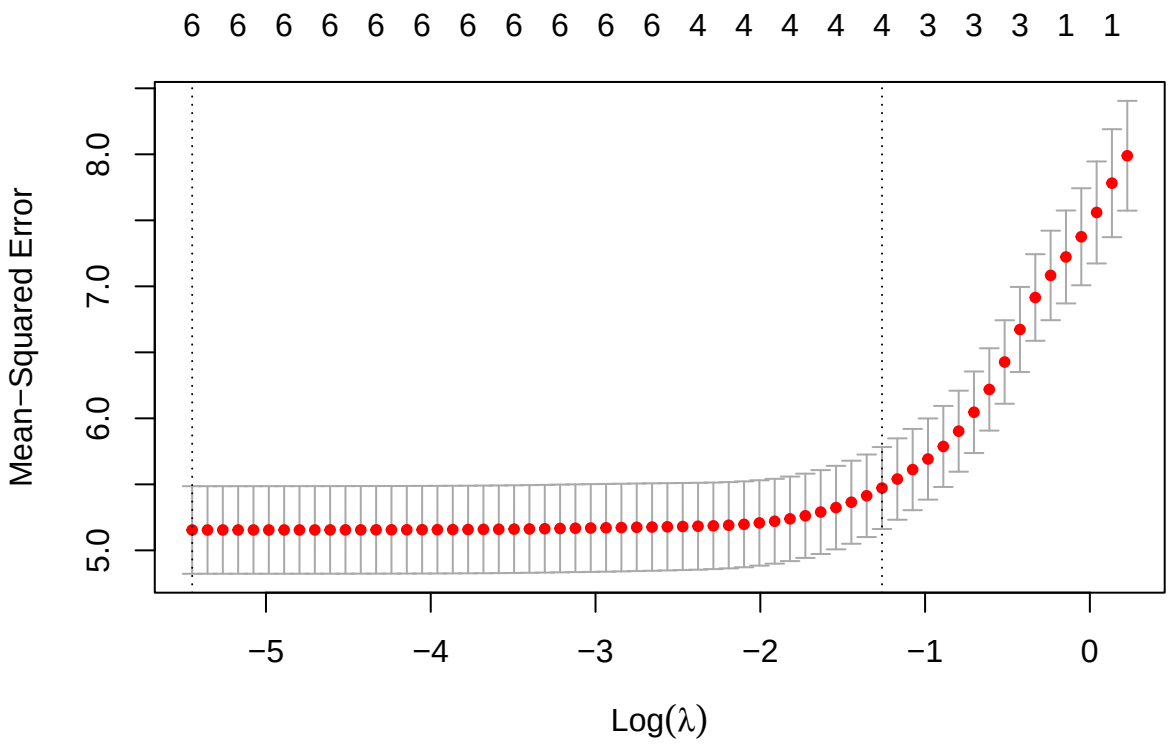
Using normalization

```
carseats_norm<- preProcess(Carseats_Filtered[-1], method = c("scale", "center"))
carseats<- predict(carseats_norm, Carseats_Filtered)
```

```
y <- carseats$Sales
x<-data.matrix(carseats[,c("Price","Advertising","Population","Age","Income","Education")])
```

#The cv.glmnet() function performs cross-validation to select the optimal value of the regularization parameter (lambda) for the LASSO regression model. The arguments y and x correspond to the response variable and predictor variables, respectively.

```
model<- cv.glmnet(x,y, alpha = 1)
plot(model)
```



#The model$lambda.min expression extracts the value of lambda that gives the minimum mean cross-validated error in the LASSO regression model,

```
lambda_value<- model$lambda.min
lambda_value
```

```
## [1] 0.004305309
```

The minimum value of the lambda is ~0.0015

QB2. What is the coefficient for the price (normalized) attribute in the best model (i.e. model with the optimal lambda)?

#The coef() function returns the coefficients of the LASSO regression model at a specified value of lambda. In this case, s="lambda.min",specifies that the function should return the coefficients at the value of lambda that gives the minimum mean cross-validated error,

```
price_co<-coef(model, s= "lambda.min")
price_co
```

```
## 7 x 1 sparse Matrix of class "dgCMatrix"
##                       s1
```

```
## (Intercept)  7.49632500
## Price       -1.35383399
## Advertising  0.82805813
## Population  -0.13061347
## Age         -0.78854992
## Income       0.28931898
## Education   -0.09102484
```

The coefficient for the price gone to negetives[-1]

QB3. How many attributes remain in the model if lambda is set to 0.01? How that number changes if lambda is increased to 0.1? Do you expect more variables to stay in the model (i.e., to have non-zero coefficients) as we increase lambda?

#The first model, model1, uses alpha = 0.01, which specifies a very small amount of L2 regularization in addition to the L1 (LASSO) penalty #The second model, model2, uses alpha = 0.1, which specifies a larger amount of L2 regularization in addition to the L1 penalty.

```
model1<- cv.glmnet(x, y, alpha = 0.01)

model2<- cv.glmnet(x, y, alpha = 0.1)
```

QB4. Build an elastic-net model with alpha set to 0.6. What is the best value of lambda for such a model?

```
elastic<-cv.glmnet(x, y, alpha = 0.6)

lambda_value1<- model$lambda.min
lambda_value1
```

```
## [1] 0.004305309
```

The best value of lamda is ~ 0.0043