Lab 11

Kristen Fitzgerald

2024-07-11

Question 1

This dataset has variables pertaining to fuel economy of various cars. Do not create a training and test set. Just use the whole cars2010 dataset for the following analysis. The cars2011 and cars2012 datasets will be used at later time periods.

Part (a)

Run a LASSO regression predicting the `FE` variable using all the remaining variables. Some of these predictor variables are coded as numeric, but should be treated as categorical. The only numeric variables in your dataset should be `EngDispl`. All remaining variables are categorical.

a. Plot the coefficients and how they change across different levels of lambda.

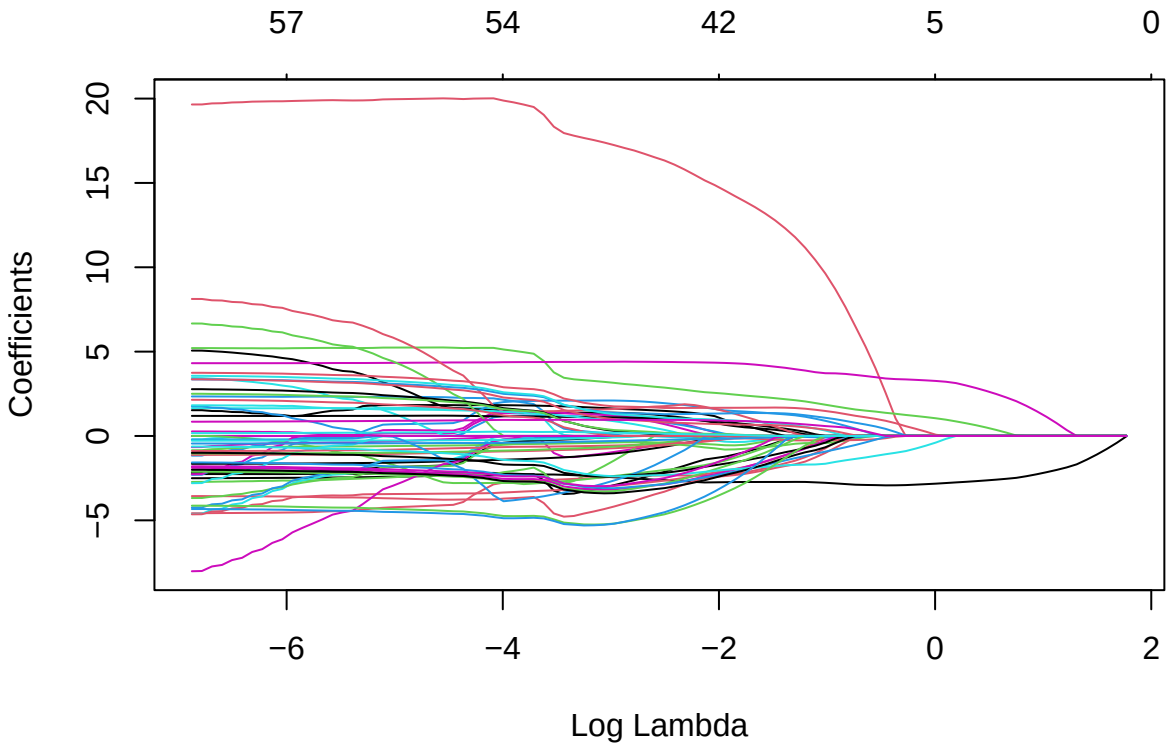
Solution:

```
train.x <- model.matrix(FE ~ ., data = cars2010)[, -1]

train.y <- cars2010$FE

cars.lasso <- glmnet(x = train.x, y = train.y, alpha = 1)

plot(cars.lasso, xvar = "lambda")
```


57
54
42
5
0
Coefficients
20
15
10
5
0
−5
−6
−4
−2
0
2
Log Lambda

Part (b)

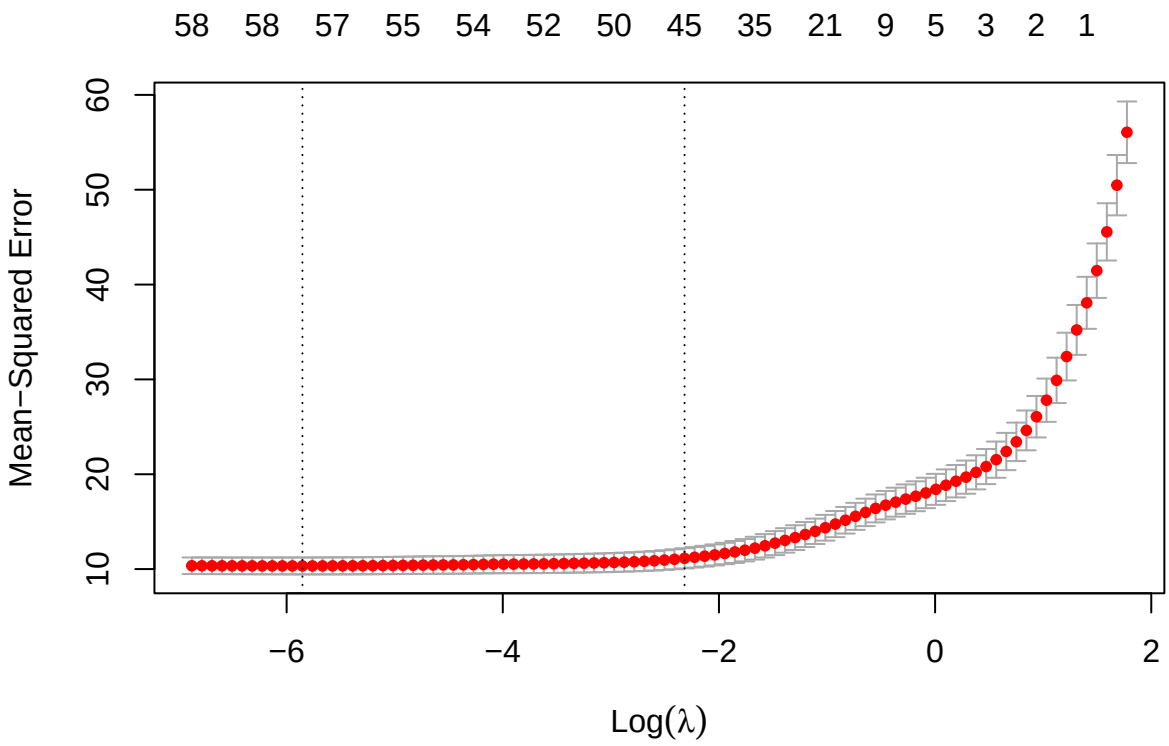
Perform a CV LASSO to optimize the lambda value

a. What is the value of lambda that minimizes the MSE?

Solution: $\lambda = \exp(0.003) = 1.003$

```
cars.lasso.cv <- cv.glmnet(x = train.x, y = train.y, alpha = 1)

plot(cars.lasso.cv)
```




```
print(cars.lasso.cv$lambda.min)
```

```
## [1] 0.002869537
```

b. What is the value of lambda one standard error above the minimum MSE value?

Solution: $\lambda = \exp(0.130) = 1.139$

```
print(cars.lasso.cv$lambda.1se)
```

```
## [1] 0.09843857
```

c. How many variables are left at the penalty value that is one standard error above the minimum MSE value (think of variables as a whole, not per category)? (HINT: Look at the coefficients from the model with coef function.)

Solution: 13 variables are left in the model.

```
coef(cars.lasso, s = cars.lasso.cv$lambda.1se)
```

```
## 61 x 1 sparse Matrix of class "dgCMatrix"
##                                                       s1
## (Intercept)                                  41.68860755
## EngDispl                                     -2.68316124
## NumCyl3                                      15.79473800
## NumCyl4                                       2.73707137
## NumCyl5                                       .
## NumCyl6                                       .
## NumCyl8                                      -0.23201183
## NumCyl10                                     -1.76380125
## NumCyl12                                     -2.87388329
## NumCyl16                                      .
## TransmissionA4                                .
## TransmissionA5                               -0.10233349
## TransmissionA6                                0.69055642
## TransmissionA7                                1.07688638
## TransmissionAM6                              -2.34348345
## TransmissionAM7                              -1.90035275
## TransmissionAV                                1.83751303
## TransmissionAVS6                              .
## TransmissionM5                                .
## TransmissionM6                                .
## TransmissionS4                               -2.48059644
## TransmissionS5                               -0.43130674
## TransmissionS6                                0.61722577
## TransmissionS7                                .
## TransmissionS8                                0.39897018
## AirAspirationMethodSupercharged              -0.42100160
## AirAspirationMethodTurbocharged              -0.44967341
## NumGears4                                    -0.54232198
## NumGears5                                    -0.01524099
## NumGears6                                     .
## NumGears7                                     .
## NumGears8                                     1.41245387
## TransLockup1                                 -0.39205975
## TransCreeperGear1                            -0.23141165
## DriveDescFourWheelDrive                      -0.20103061
## DriveDescParttimeFourWheelDrive               .
## DriveDescTwoWheelDriveFront                   4.37500789
## DriveDescTwoWheelDriveRear                    0.98498272
## IntakeValvePerCyl1                            1.88088706
## IntakeValvePerCyl2                            .
## IntakeValvePerCyl3                           -0.83647776
## ExhaustValvesPerCyl1                          0.05528322
## ExhaustValvesPerCyl2                          .
## CarlineClassDesc2Seaters                      .
```

```
## CarlineClassDescCompactCars                        1.66181194
## CarlineClassDescLargeCars                          0.15091593
## CarlineClassDescMidsizeCars                        1.30199377
## CarlineClassDescMinicompactCars                    1.03108474
## CarlineClassDescSmallPickupTrucks2WD              -2.66804269
## CarlineClassDescSmallPickupTrucks4WD              -2.03469417
## CarlineClassDescSmallStationWagons                 .
## CarlineClassDescSpecialPurposeVehicleminivan2WD -2.52712904
## CarlineClassDescSpecialPurposeVehicleSUV2WD     -2.67716385
## CarlineClassDescSpecialPurposeVehicleSUV4WD     -2.24002658
## CarlineClassDescStandardPickupTrucks2WD         -2.19417098
## CarlineClassDescStandardPickupTrucks4WD         -2.87855301
## CarlineClassDescSubcompactCars                     0.76721140
## CarlineClassDescVansCargoTypes                    -4.29639858
## CarlineClassDescVansPassengerType                 -4.10537585
## VarValveTiming1                                    0.11025301
## VarValveLift1                                      0.94858328
```

Part (c)

Obtain the variables from the LASSO regression at the penalty value that is one standard error above the minimum MSE value. The multiple linear regression with p-value selection (Lab 6) left the variables `EngDispl`, `NumCyl`, `Transmission`, `AirAspirationMethod`, `NumGears`, `TransLockup`, `DriveDesc`, `IntakeValvePerCyl`, `CarlineClassDesc`, and `VarValveLift`.

a. What variables were left in your LASSO at the 1SE above minimum MSE penalty value?

Solution: All 13 variables were left in the model.