HW9

Lucy Lin

2022-11-06

Problem 1a

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0.0000   Median :4.000   Median :2.000  
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812  
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```
#help(mtcars)

# Output of help():
# Usage
# mtcars
# Format
# A data frame with 32 observations on 11 (numeric) variables.
#
# [, 1] mpg Miles/(US) gallon
# [, 2] cyl Number of cylinders
# [, 3] disp    Displacement (cu.in.)
# [, 4] hp  Gross horsepower
# [, 5] drat    Rear axle ratio
# [, 6] wt  Weight (1000 lbs)
```

```
# [, 7] qsec    1/4 mile time
# [, 8] vs  Engine (0 = V-shaped, 1 = straight)
# [, 9] am  Transmission (0 = automatic, 1 = manual)
# [,10] gear    Number of forward gears
# [,11] carb    Number of carburetors
```

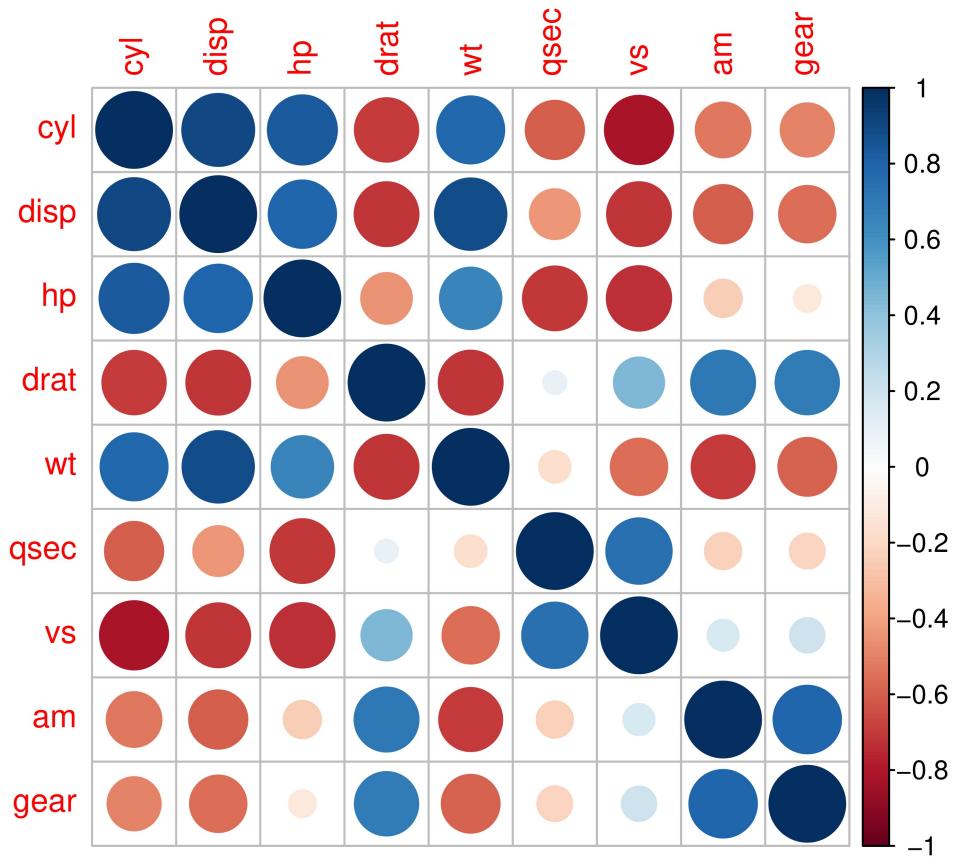
Problem 1b

Findings: hp and disp are positively correlated with cyl. # drat and vs are negatively correlated with cyl, dis, hp, and wt. # am and gear are positively correlated with each other. #

```
fit <- lm(mpg ~ ., data = mtcars)
round(cor(mtcars[,2:10]), 2)
```

```
##        cyl  disp    hp  drat    wt  qsec    vs    am  gear
## cyl   1.00  0.90  0.83 -0.70  0.78 -0.59 -0.81 -0.52 -0.49
## disp  0.90  1.00  0.79 -0.71  0.89 -0.43 -0.71 -0.59 -0.56
## hp    0.83  0.79  1.00 -0.45  0.66 -0.71 -0.72 -0.24 -0.13
## drat -0.70 -0.71 -0.45  1.00 -0.71  0.09  0.44  0.71  0.70
## wt    0.78  0.89  0.66 -0.71  1.00 -0.17 -0.55 -0.69 -0.58
## qsec -0.59 -0.43 -0.71  0.09 -0.17  1.00  0.74 -0.23 -0.21
## vs   -0.81 -0.71 -0.72  0.44 -0.55  0.74  1.00  0.17  0.21
## am   -0.52 -0.59 -0.24  0.71 -0.69 -0.23  0.17  1.00  0.79
## gear -0.49 -0.56 -0.13  0.70 -0.58 -0.21  0.21  0.79  1.00
```

```
corrplot::corrplot(cor(mtcars[,2:10]))
```



Problem 1c

```
X = mtcars[,2:10]
round(faraway::vif(X), 3)
```

```
##    cyl   disp     hp   drat     wt   qsec     vs     am   gear
## 14.574 11.784  7.105  3.231  7.839  6.985  4.923  4.631  4.393
```

Problem 1d Large range in eigenvalues and condition values.

```
X = as.matrix(X)
(lambda = (eigen(t(X) %*% X))$val)
```

```
## [1] 2.973062e+06 5.109143e+04 2.584759e+03 1.849463e+01 1.825114e+01
## [6] 4.554178e+00 2.559894e+00 1.869618e+00 1.240618e+00
```

```
sqrt(max(lambda)/ lambda)
```

```
## [1]    1.000000    7.628304   33.915013  400.939826  403.605419  807.973339
## [7] 1077.682953 1261.030444 1548.042783
```

Problem 1e

Almost all of the predictors are not significant.

```
fit1 <- lm(mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb, data = mtcars)
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + disp + hp + drat + wt + qsec + vs +
##     am + gear + carb, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

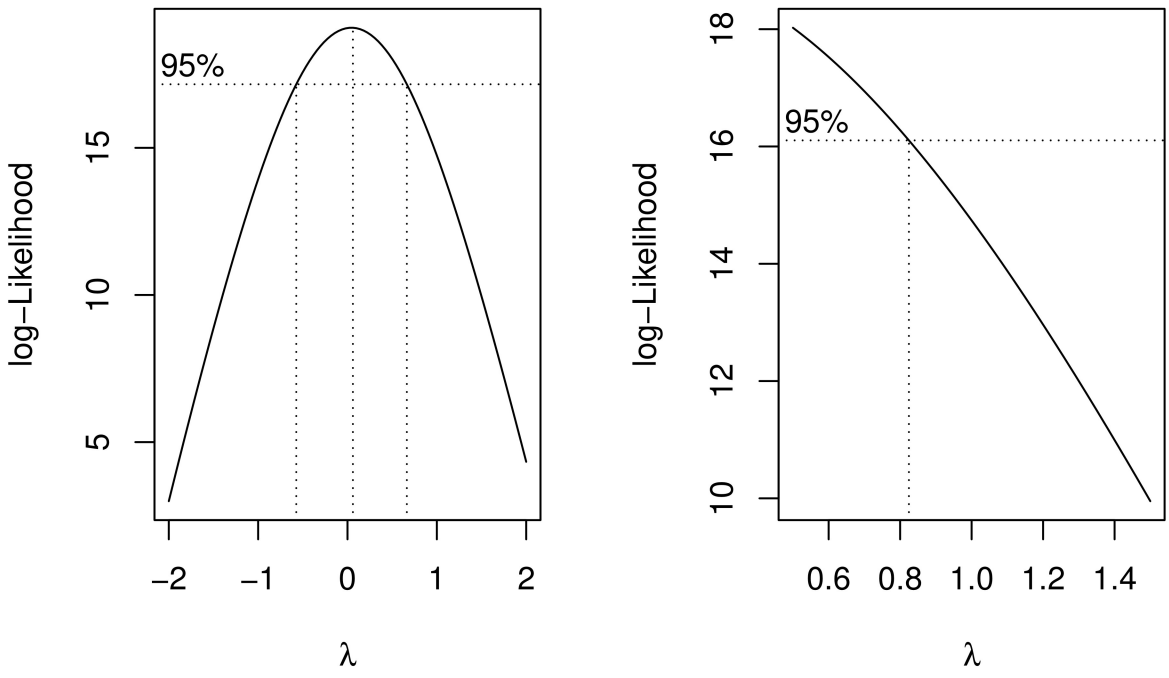
Problem 2 optimal lambda: 0.06060606

```
data = data(gala, package = "faraway")

fit2 = lm(Species ~ .,data = gala)
require(MASS)
```

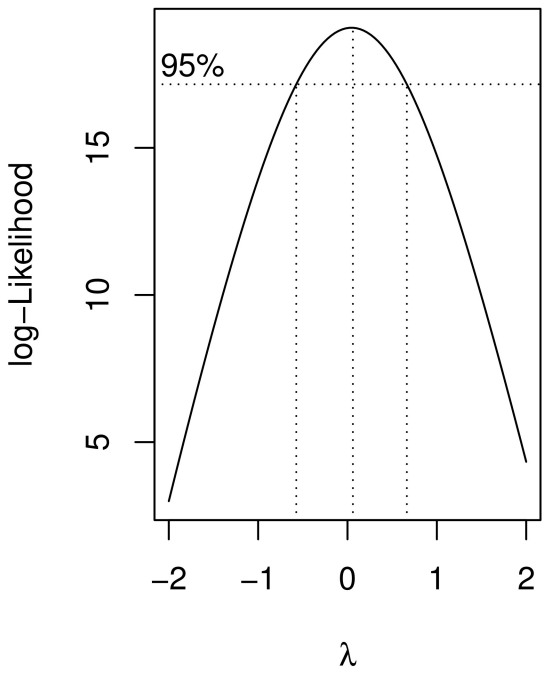
```
## Loading required package: MASS
```

```
par(mfrow = c(1,2))
boxcox(fit, plotit = TRUE)
boxcox(fit, plotit = TRUE, lambda = seq(0.5, 1.5, by=0.1))
```




```
b = boxcox(fit)
b$x[which.max(b$y)]
```

```
## [1] 0.06060606
```



#Problem 3a #(a) The estimate in (2) has a higher variance than the estimate in (1).q True, as complexity of a model increases, so does variance.

#Problem 3b #(b) The estimate in (2) has a higher bias than the estimate in (3). True, this is due to a lower complexity in (2) than (3).

#Problem 3c #(c) The estimate in (3) has the smallest training error among the three. True, training errors are minimized as the model complexity increases.

#Problem 3d #(d) The estimate in (1) has the smallest test error among the three. False, the model is very biased since it has low complexity. Additionally the line of fit does not pass through many of the points.

#Problem 4a True, this shows that the beta_2 is significant at the 0.0001 level.

#Problem 4b True, nested models will not be able to explain as much as their larger model.

#Problem 4c False, we cannot tell because M1 and M2 are not a nested/larger models of each other.

#Problem 4d True, larger AIC values imply a worse model.

#Problem 4e False, you cannot compare Y to Yˆ2.