# CFerrari_Assignment12

*Charley Ferrari*

*Thursday, April 23, 2015*

Using the stats and boot libraries in R perform a cross-validation experiment to observe the bias variance tradeoff. You'll use the auto data set from previous assignments. This dataset has 392 observations across 5 variables. We want to fit a polynomial model of various degrees using the glm function in R and then measure the cross validation error using cv.glm function Fit various polynomial models to compute mpg as a function of the other four variables acceleration, weight, horsepower, and displacement using glm function.

```
library(stats)
library(boot)
library(dplyr)
```

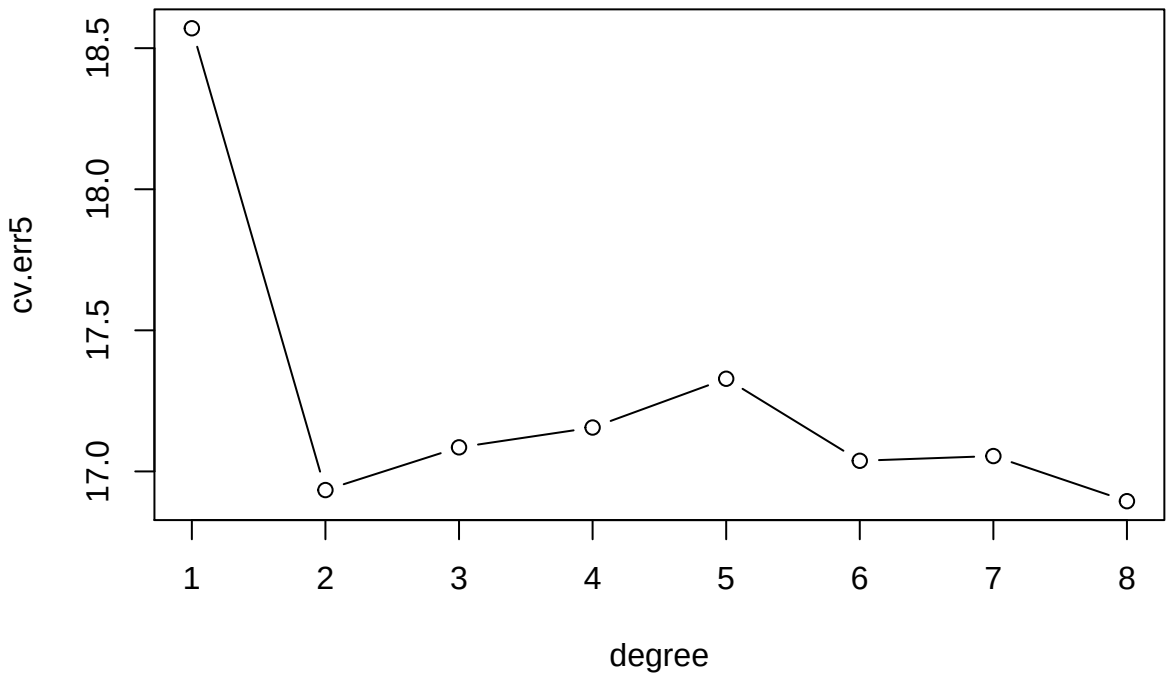
```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)

setwd("E:/Downloads/Courses/CUNY/SPS/Git/IS 605 Fundamentals of Computational Mathematics/Assignment 12

autodata <- scan("auto-mpg.data")
autodata <- t(matrix(autodata, nrow = 5))

autodata <- data.frame(displacement = autodata[,1],
                       horsepower = autodata[,2],
                       weight = autodata[,3],
                       acceleration = autodata[,4],
                       mpg = autodata[,5])

glm.fit=glm(mpg~poly(displacement+horsepower+
                       weight+acceleration,2), data=autodata)

n <- 8

degree <- 1:n

cv.err5 <- c()

for(i in 1:n){
  glm.fit <- glm(mpg~poly(displacement+horsepower+
                       weight+acceleration,i), data=autodata)
```

```
  cv.err5 <- c(cv.err5, cv.glm(autodata, glm.fit, K=5)$delta[1])
}
```

Once you have fit the various polynomials from degree 1 to 8, you can plot the crossvalidation error function

```
plot(degree,cv.err5,type='b')
```




```
#how to interpret glm summary R poly
```

I would have expected more of a u-shape. I tried this with larger numbers and got more erratic behavior. Trying the test again and again, choosing different values for cross validation, led to more wider changes. It seemed to be a u-shape, but for higher degree polynomials, the error scores themselves seemed to have a wide variance.

```
n <- 17

degree <- 1:n

cv.err5 <- c()

for(i in 1:n){
  glm.fit <- glm(mpg~poly(displacement+horsepower+
                          weight+acceleration,i), data=autodata)
  cv.err5 <- c(cv.err5, cv.glm(autodata, glm.fit, K=5)$delta[1])
```

```
}

plot(degree,cv.err5,type='b')
```