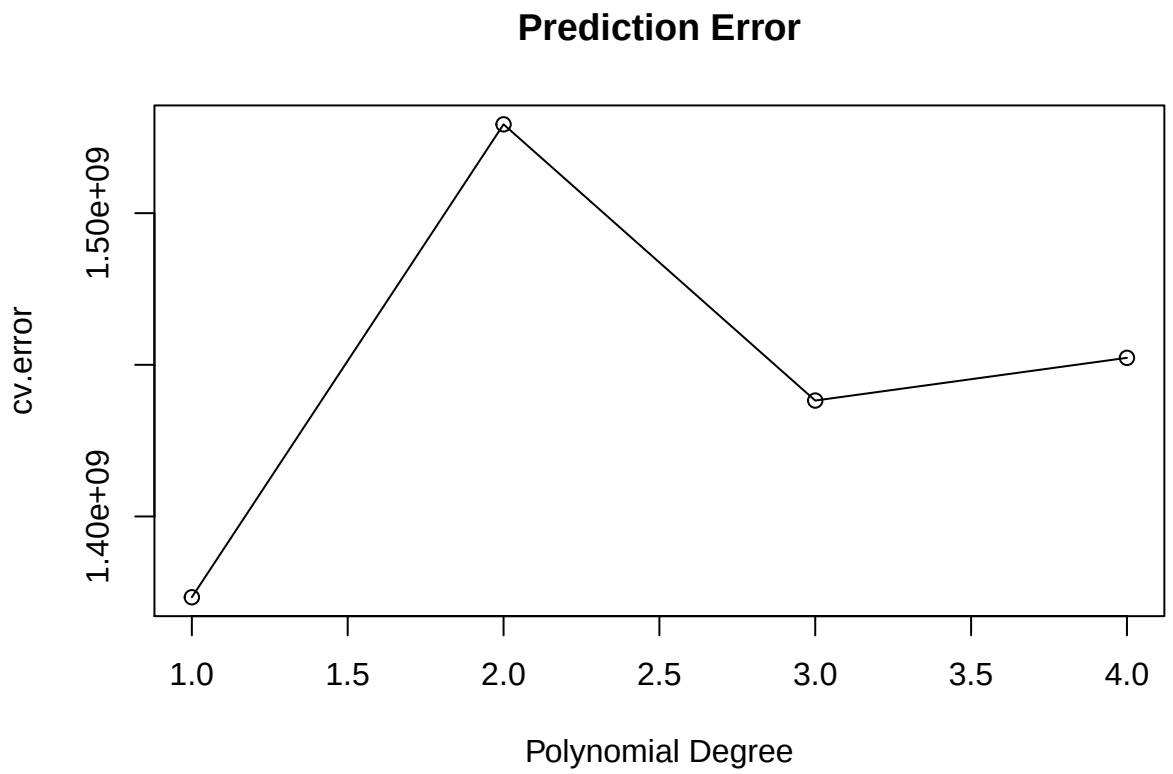
HW 3

Ben Buzzee

9/20/2019

1. Cross Validation

```
library(boot)
library(MASS)
income <- read.csv("income.txt", sep = "", header = T)

glm.fit1 = glm(y~x, data=income)
cv.error1=cv.glm(income, glm.fit1, K=5)


glm.fit2 = glm(y~poly(x,2), data=income)
# quadratic model
cv.error2=cv.glm(income, glm.fit2, K=5)

glm.fit3 = glm(y~poly(x,3), data=income)
# cubic model
cv.error3=cv.glm(income, glm.fit3, K=5)

glm.fit4 = glm(y~poly(x,4), data=income)
# quartic model
cv.error4=cv.glm(income, glm.fit4, K=5)


cv.error=c(cv.error1$delta[1],cv.error2$delta[1],cv.error3$delta[1],cv.error4$delta[1])

d=c(1,2,3,4)
plot(d,cv.error, main = "Prediction Error", xlab = "Polynomial Degree")
lines(d,cv.error)
```



Depending on the situation a degree of either 1 or 4 would be acceptable. A linear model would be much more interpretable, while a polynomial of degree 4 has a smaller prediction error.