

HW 1

Carlyle Morgan

2/15/2022

3.

a.

```
set.seed(5)

y<-c()

mses<-data.frame(rep(NA,1000),rep(NA,1000),rep(NA,1000),rep(NA,1000),rep(NA,1000),rep(NA,1000),rep(NA,1000))

for (j in seq(1,1000)){

  x_sim<-sort(sample(seq(1,100,.001),200))
  for (i in 1:length(x_sim)){
    y[i]<-rnorm(1,2+4*x_sim[i],70)
  }

  trainind<-sort(sample(seq(1,200,1),100))
  y_train<-y[trainind]
  y_test<-y[-trainind]
  x_train<-x_sim[trainind]
  x_test<-x_sim[-trainind]

  x_train<-data.frame(x_train,x_train^2, x_train^3, x_train^4, x_train^5)
  x_test<-data.frame(x_test,x_test^2, x_test^3, x_test^4, x_test^5)

  linmodel<-lm(y_train~x_train[,1])
  quadmodel<-lm(y_train~x_train[,1]+x_train[,2])
  cubmodel<-lm(y_train~x_train[,1]+x_train[,2]+x_train[,3])
  x4model<-lm(y_train~x_train[,1]+x_train[,2]+x_train[,3]+x_train[,4])
  x5model<-lm(y_train~x_train[,1]+x_train[,2]+x_train[,3]+x_train[,4]+x_train[,5])

  mses[j,1]<-mean(linmodel$residuals^2)
  mses[j,2]<-mean(quadmodel$residuals^2)
  mses[j,3]<-mean(cubmodel$residuals^2)
  mses[j,4]<-mean(x4model$residuals^2)
  mses[j,5]<-mean(x5model$residuals^2)
  mses[j,6]<-mean((y_test-predict(linmodel,x_test))^2)
  mses[j,7]<-mean((y_test-predict(quadmodel,x_test))^2)
  mses[j,8]<-mean((y_test-predict(cubmodel,x_test))^2)
  mses[j,9]<-mean((y_test-predict(x4model,x_test))^2)
  mses[j,10]<-mean((y_test-predict(x5model,x_test))^2)
```

```

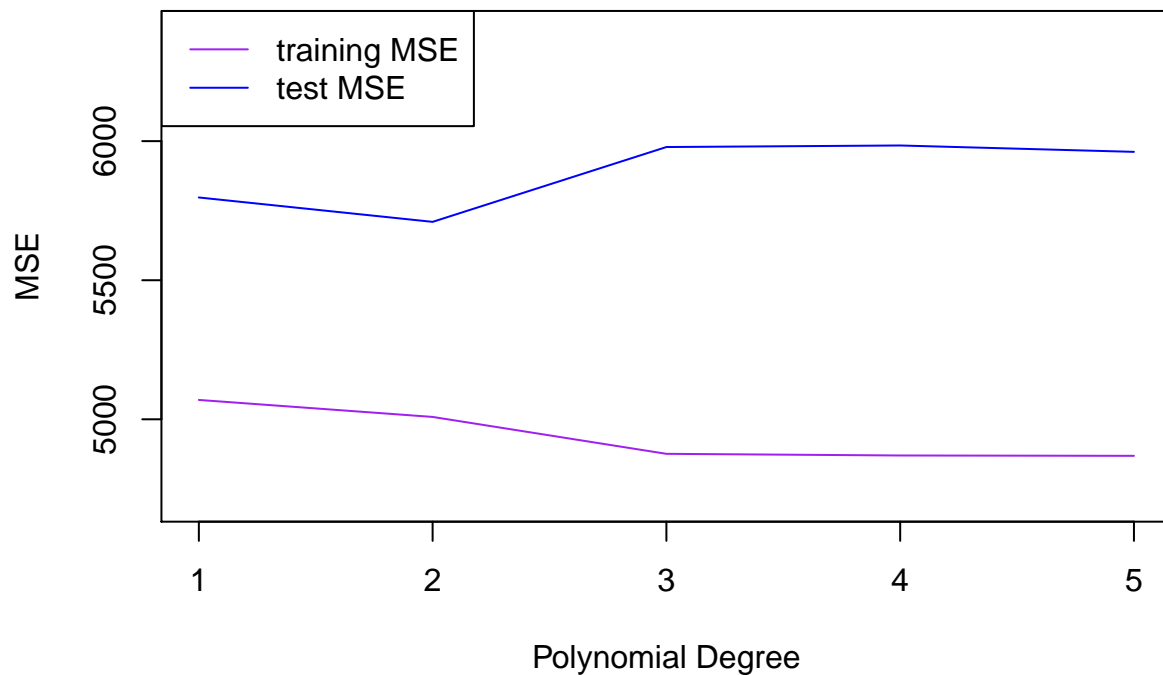
}

degdata1<-data.frame(seq(1,5),t(mses[1,c(1,2,3,4,5)]),t(mses[1,c(6,7,8,9,10)]))
colnames(degdata1)<-c("Degree", "Train MSE", "Test MSE")
means1000<-apply(mses,2,mean)
degdata1000<-data.frame(seq(1,5), means1000[1:5], means1000[6:10])
colnames(degdata1000)<-c("Degree", "Avg Train MSE", "Avg Test MSE")

plot(degdata1$Degree, degdata1$`Train MSE`, type = "l", col = "purple", ylab = "MSE", xlab = "Polynomial Degree",
lines(degdata1$Degree, degdata1$`Test MSE`, type = "l", col = "blue")
legend("topleft", c("training MSE", "test MSE"),
      lty = c(1,1),
      col = c("purple", "blue"))

```

R = 1 Simulations



There is no guarantee that simulations will have consistent patterns.

b.

```

plot(degdata1$Degree, degdata1000$`Avg Train MSE`, type = "l", col = "purple", ylab = "MSE", xlab = "Polynomial Degree",
lines(degdata1$Degree, degdata1000$`Avg Test MSE`, type = "l", col = "blue")
legend("topleft", c("training MSE", "test MSE"),
      lty = c(1,1),
      col = c("purple", "blue"))

```

R = 1000 Simulations

