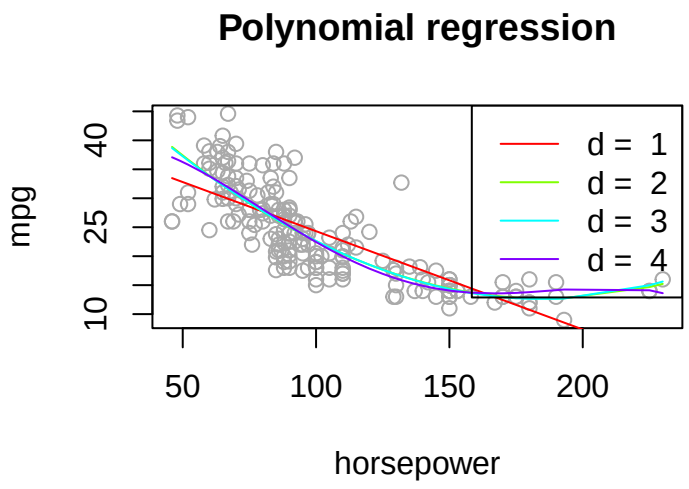
Øving 6

2022-05-30

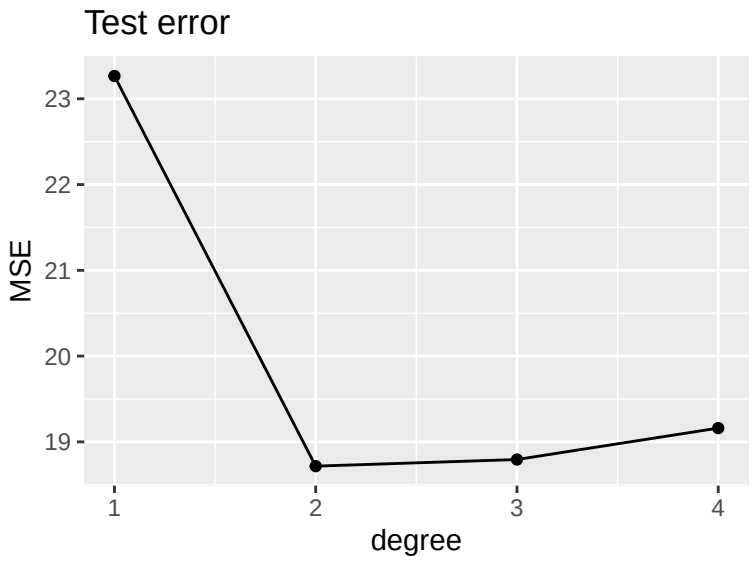
Problem 1

Perform polynomial regression of degree 1,2,3 and 4. Use lines() to add the fitted values in the plot. Also plot the test error depending on polynomial degree.

```
library(ISLR)

ds = Auto[c("horsepower", "mpg")]
n = nrow(ds)

degrees = 1:4

set.seed(1)

train_id = sample.int(n,n/2)

plot(ds[train_id,], col = "darkgrey", main = "Polynomial regression")

co = rainbow(length(degrees))


MSE = sapply(degrees,function(d){
  model = lm(mpg ~ poly(horsepower, degree = d), data = ds[train_id,])
  lines(cbind(ds[train_id,1],model$fitted.values)[order(ds[train_id,1]),], col=co[d])
  mean((predict(model, ds[-train_id,])-ds[-train_id,2])^2)
})

legend("topright", legend = paste("d = ", degrees), lty = 1, col = co)
```




```
MSE = data.frame(MSE = MSE, degree = 1:4)

ggplot(data = MSE, aes(x = degree, y = MSE)) + geom_line() + geom_point() + labs(title = "Test error")
```



Problem 2

Use factor(origin) for conversion to factor variable for the origin variable. Predict mpg by origin with a linear model. Plot the fitted values and approximate 95% confidence intervals.