ZZSC5806- Regression analysis for Data Scientists (Week 4)

Atefeh Zamani

Deparment of Statistics

University of New South Wales

Packages

Hide

```
library(ISLR)
library(SemiPar)
library(doBy)
library(ResourceSelection)
```

4.1 Model Assessment and Selection

Simulated a dataset with $n = 20$ of the form $y_i = x_i + \epsilon_i$, where $\epsilon_i \sim N(0, 0.25^2)$.

Hide

```
set.seed(20242)
n=20
x <- runif(n) # covariates
err <- rnorm(n, sd=0.25)
y <- x + err
```

Fit polynomial of order k=5,10 and 15

Hide

```
linear<- lm(y~x)
poly5 <- lm(y~poly(x,5))
poly10 <- lm(y~poly(x,10))
poly15 <- lm(y~poly(x,15))
```

Plot the data and the fitted models

Hide

```
newx <- seq(min(x),max(x),length=100)

predict.poly5 <- predict(poly5, newdata=list(x=newx))
predict.poly10 <- predict(poly10, newdata=list(x=newx))
predict.poly15 <- predict(poly15, newdata=list(x=newx))

plot(x,y, pch=16, ylim=range(predict.poly10), xlab="X", ylab="Y")
points(newx,predict.poly5, type="l", col="blue", lwd=2)
```
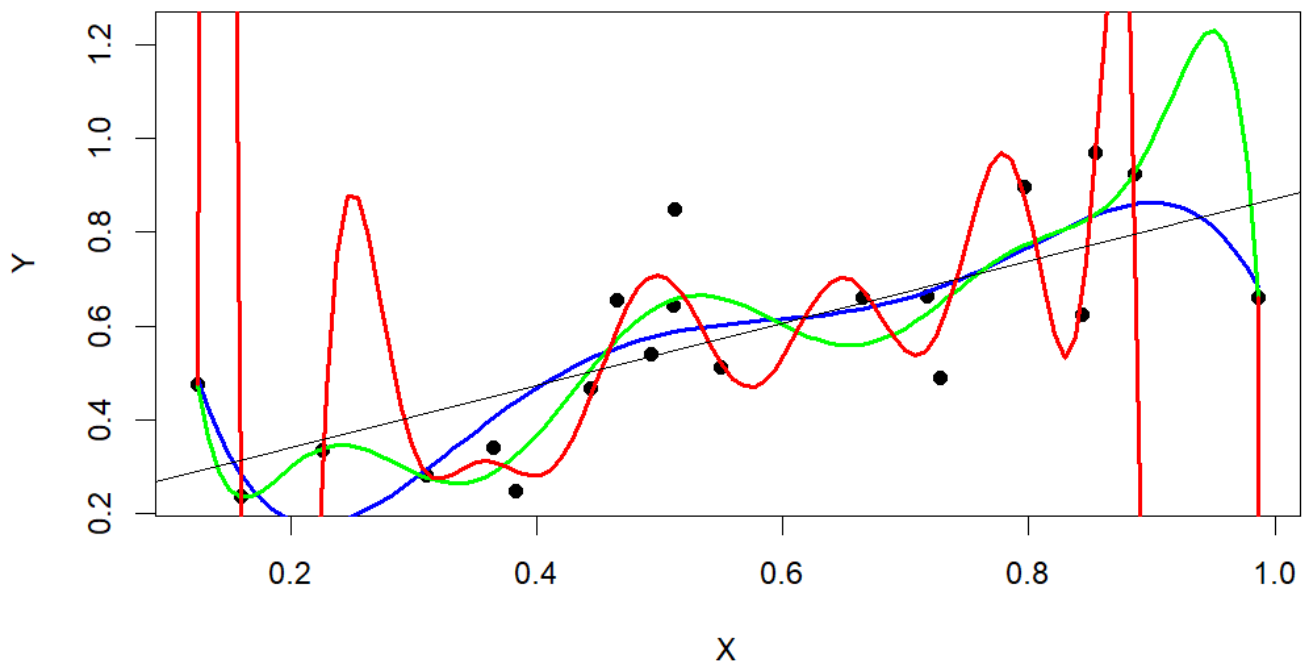
Hide

```
points(newx,predict.poly10, type="l", col="green", lwd=2)
points(newx,predict.poly15, type="l", col="red", lwd=2)
```

Hide

```
abline(linear$coefficients[1],linear$coefficients[2])
```


Hide

```
# dev.copy2pdf(file="Lect7_PolyFits.pdf", height=5, width=7)
```

4.2 Cross Validation

Hide

```
library(faraway)
```

```
Attaching package: 'faraway'

The following object is masked _by_ '.GlobalEnv':

    melanoma

The following object is masked from 'package:doBy':

    prostate
```

Hide

```
data(package="faraway")
data("cheddar")
attach(cheddar)
```

How to plot Figure 3 (MSE values obtained for the cheddar dataset in order to estimate the test errors.)

Hide