DATA 881 - Lesson 10 Homework

Katie Haring

11/3/2019

Use the unemployment.csv dataset and subset it to include only data from the United States. Use at least two of the approaches we covered this week (cubic splines, natural splines, smoothing splines, local regression, or GAMs) and model the unemployment rate over the years in these data. Show a figure with two curves that show the results for the two methods, making sure to label the curves so I can tell which is which. Do some interpretation of what you are seeing in the data.

Data Manipulation Code

```
library(data.table)
data<-fread("C:/Users/Katie/Documents/Homework/Stat Learning I - Fall 19/
            Lesson 10/unemployment.csv",header=TRUE)

datasub <- data[which(data$country == "United States")]
datasub = datasub[,-1]
datasub<-t(datasub)

Years<-data.frame(row.names(datasub))
Unemp<-datasub[,1]

newdata<-cbind(Years,Unemp)
newdata<-data.frame(newdata,row.names=NULL)
colnames(newdata)<-c("Year","Unemp")
```

Smoothing Splines Code

```
## Letting R determine lambda, using leave-one-out cross-validation
model<-smooth.spline(newdata$Year,newdata$Unemp)
preds = predict(model)

## Setting lambda
model1 = smooth.spline(newdata$Year, newdata$Unemp,lambda = .00075)
preds1 = predict(model1)
```

Local Regression Code

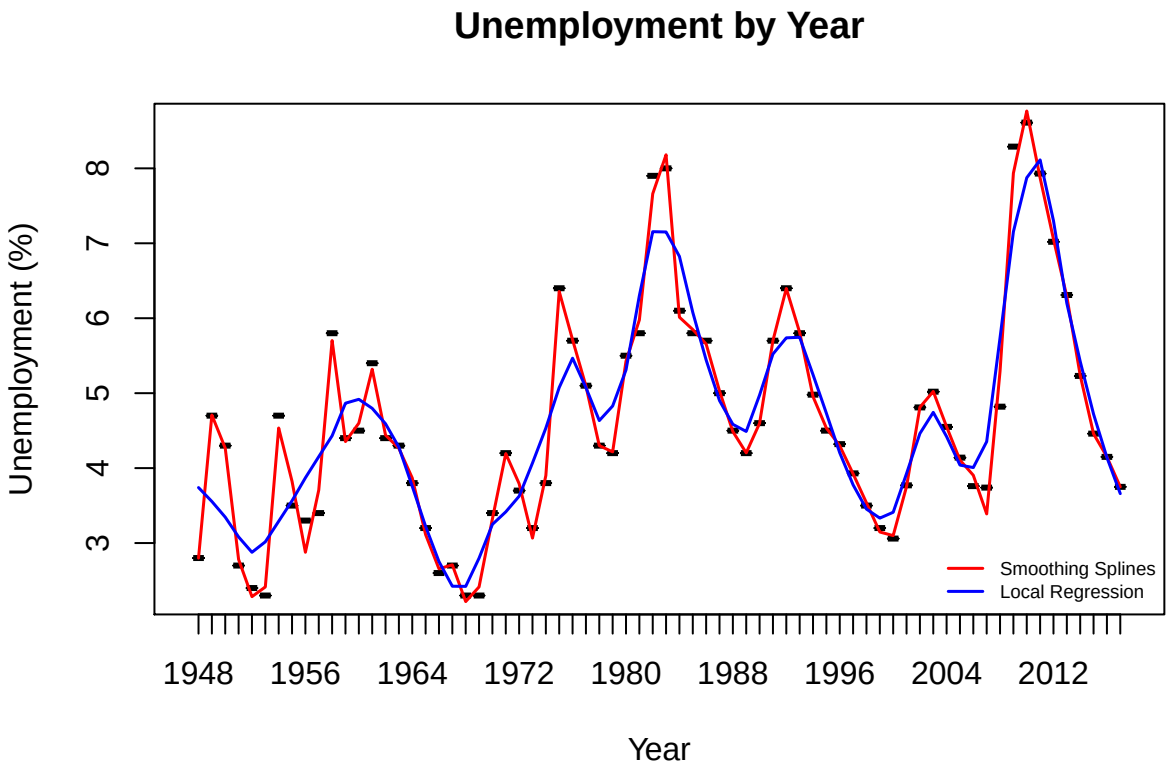
```
## Span=0.15
model2<-loess(newdata$Unemp~as.numeric(newdata$Year),span=.15)
preds2 = predict(model2)
ord = order(newdata$Year)

## Span =0.5
model3<-loess(newdata$Unemp~as.numeric(newdata$Year),span=.5)
preds3 = predict(model3)
ord = order(newdata$Year)
```

Plotting Code

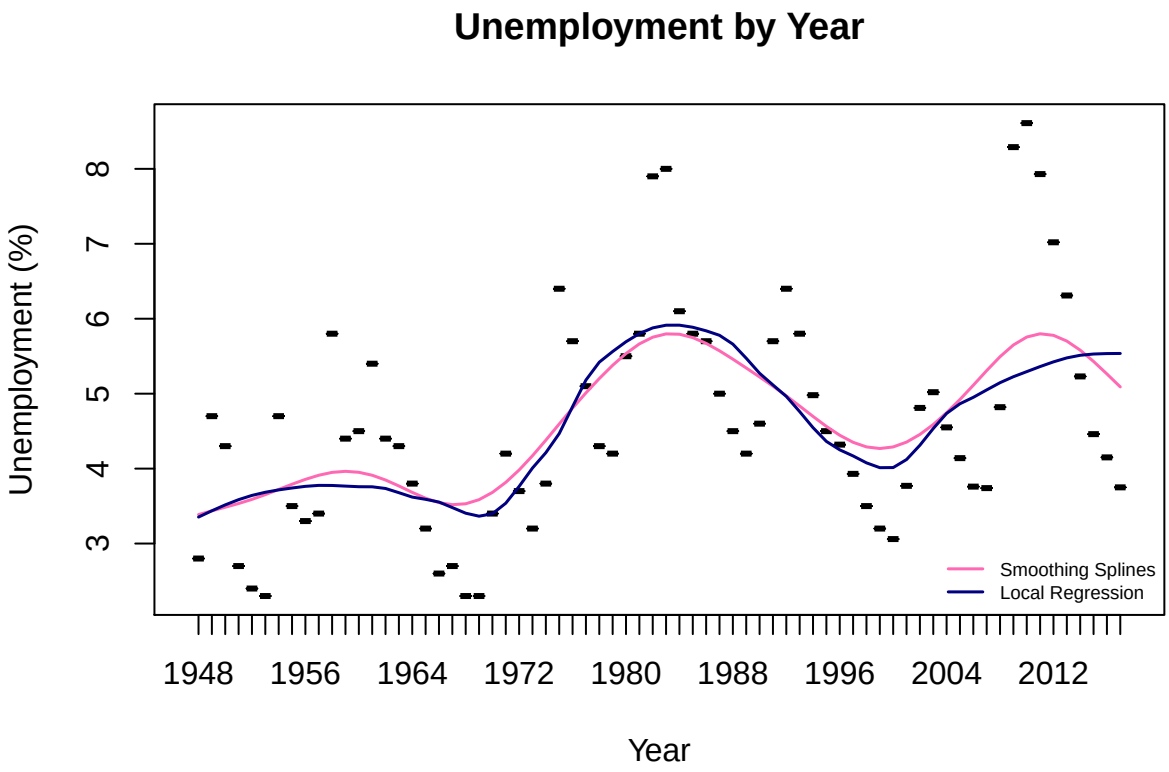
```
## Plot Smoothing Splines with R-determined lambda and Local Regression with Span=0.15
with(newdata, plot(Year, Unemp,
                   main = "Unemployment by Year",
                   xlab="Year",ylab="Unemployment (%)"))
legend("bottomright",legend=c("Smoothing Splines","Local Regression"),
                    col=c("red","blue"),lwd=1.5,lty=1:1,cex=.6,box.lty=0)
lines(preds$x, preds$y, col='red', lwd=1.5)
lines(newdata$Year[ord], preds2[ord], col='blue', lwd=1.5)

## Plot Smoothing Splines with user-set lambda and Local Regression with span=0.5
with(newdata, plot(Year, Unemp,
                   main = "Unemployment by Year",
                   xlab="Year",ylab="Unemployment (%)"))
legend("bottomright",legend=c("Smoothing Splines","Local Regression"),
                    col=c("hot pink","navy blue"),lwd=1.5,lty=1:1,cex=.6,box.lty=0)
lines(preds1$x, preds1$y, col='hot pink', lwd=1.5)
lines(newdata$Year[ord], preds3[ord], col='navy blue', lwd=1.5)
```

Plots

Smoothing Spline with R-determined lamdba and Local Regression with span=0.15



Smoothing Spline with user-set lambda (0.00075) and Local Regression with span=0.5



Interpretation

Unemployment was at its lowest in the late 1960s and peaked in 2010. The unemployment rate appears to be somewhat cyclical, based on the data. The rate increases until it hits a peak, then decreases until it hits a nadar; it generally does not hold steady at any one point.

Both smoothing splines and local regression can fit smoother or rougher lines reflecting this, depending on how they are utilized.

If the user does not choose a lambda for their smoothing spline, R will set one using LOOCV. For this data, that results in a highly biased line that passes through almost every data point. This is similar to using Local Regression with the span set to 0.15.

However, if the user does set a lambda for their smoothing spline, they can choose a value that creates a much smoother fit. In this case, a lambda of 0.00075 creates a very smooth line showing general peaks and valleys. This is very simliar to using local regression with a span of 0.5.

For local regression, a closer but less smooth fit can be obtained by decreasing the span. For smoothing splines, this can be done by reducing lambda .