

MATH5824 - Solution to Chapter 4 Exercises

Exercise 4.1 The following uses R code from examples and previous exercises.

```
# Define the variables from earlier
season = 1:13
cyclones = c(6,5,4,6,6,3,12,7,4,2,6,7,4)

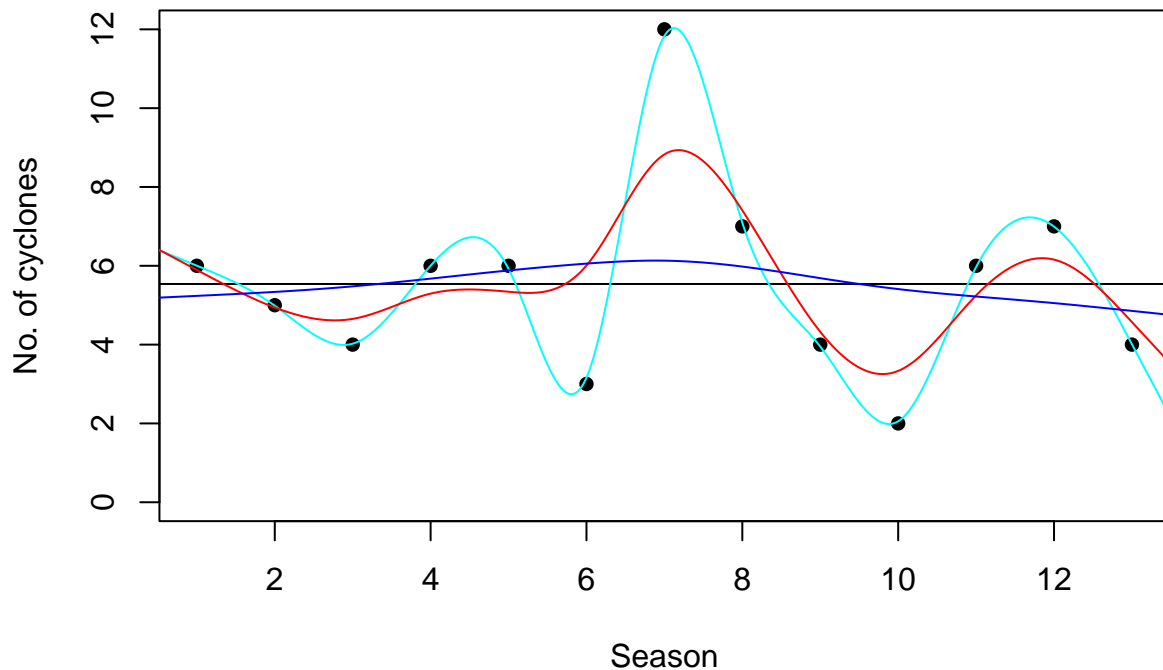
# Draw a scatter plot of the data and add a line at the mean value
plot(season, cyclones, pch=16,
     ylim=c(0,12),
     xlab="Season", ylab="No. of cyclones")
abline(h=mean(cyclones))

# Define a set of x-values at which to evaluate the splines
new.season = seq(0,14, 0.01)

# Fit and plot spline with varying lambda
my.fit = smooth.spline(season, cyclones, lambda=1e-6)
new.fitted = predict(my.fit, new.season)
lines(new.fitted, col="cyan")

my.fit = smooth.spline(season, cyclones, lambda=1e-4)
new.fitted = predict(my.fit, new.season)
lines(new.fitted, col="red")

my.fit = smooth.spline(season, cyclones, lambda=1e-2)
new.fitted = predict(my.fit, new.season)
lines(new.fitted, col="blue")
```



The first spline uses a very small value of the smoothing parameter, $\lambda = 1e^{-6}$. This is close to interpolating the data and hence may describe unimportant variation.. The next, has $\lambda = 1e^{-4}$, and seems to match the ups and down without following the data too closely. The final graph has $\lambda = 1e^{-2}$ which leads to a spline which is not too far from a constant value.

Of these, $\lambda = 1e^{-4}$ seems to be the best as it reflects changes through time without over fitting.

Exercise 4.2 Follow a similar approach to spline modelling of this data set as in the previous. It is not clear which is the response and which the explanatory but, to me, it seems reasonable to believe that if we have to wait longer for an eruption then the eruption will be longer and hence the following takes **waiting** as the explanatory and **eruption** as the response.

```
attach(faithful)
plot(waiting, eruptions, pch=16, col="grey")

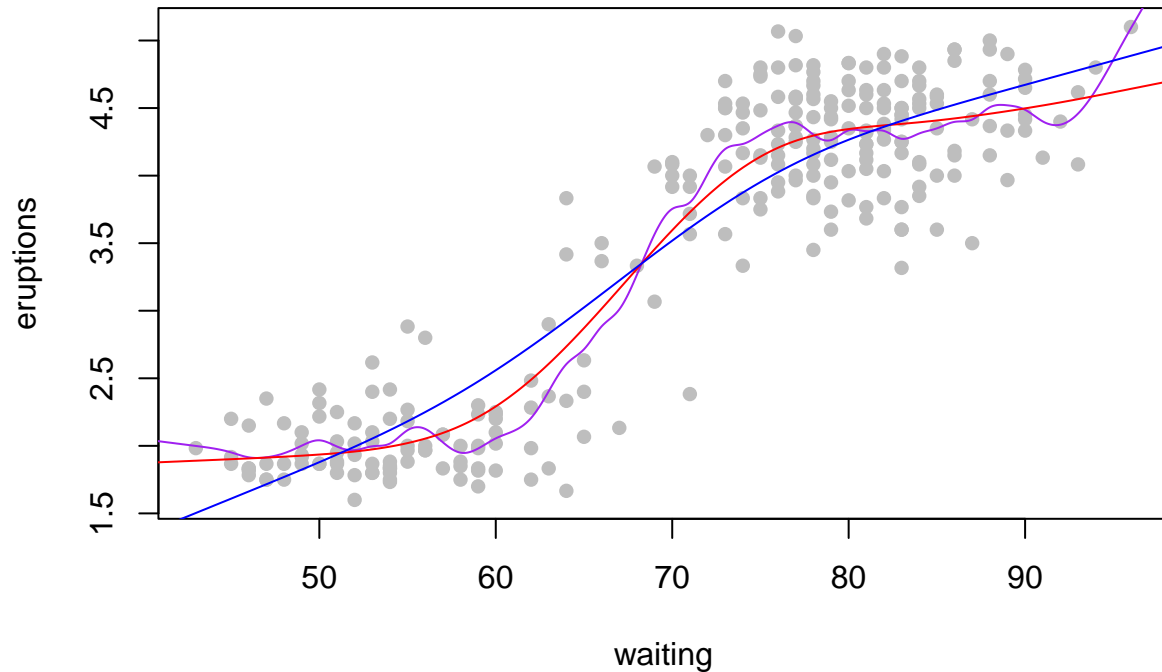
new.waiting = seq(40,100, 0.1)

my.fit = smooth.spline(waiting, eruptions, lambda=1e-5)
my.fitted = predict(my.fit, new.waiting)
lines(my.fitted, col="purple")

my.fit = smooth.spline(waiting, eruptions, lambda=1e-2)
my.fitted = predict(my.fit, new.waiting)
lines(my.fitted, col="red")

my.fit = smooth.spline(waiting, eruptions, lambda=1e-1)
```

```
my.fitted = predict(my.fit, new.waiting)
lines(my.fitted, col="blue")
```



A small value of the smoothing parameter, for example $\lambda = 1e^{-5}$ creates a spline with too much “wiggle”, even though it is not interpolating the data. The intermediate $\lambda = 1e^{-2}$ describes the changes well, it shows the general trend without following being effected by small frequent fluctuations. The largest, $\lambda = 1e^{-1}$, gives a spline which is a little stiff and over simplifies the pattern – it is getting close to the asymptotic straight line fit.

Of these, $\lambda = 1e^{-2}$ gives the best description – though slightly more experimentation which suggest a better nearby value.

End of Solutions to Exercises 4