STATS 205: Homework Assignment 6 (Spring 2019)

5/20/2019

Solve problems 1-2 and 4 from the textbook **HWC** available here and problem 3 from **W 2006**.

Send your Rmd and PDF files to pjeganat [at] stanford [dot] edu.

Due on 5/28/2019 (Tuesday) at 1.30 p.m.

1) **HWC** Page 661, Problem 1 (local averaging). The data set `cars` from Ezekiel (1930) contains stopping distances for various speeds. Smooth the data using Friedman's smoother by choosing your own value of the span. Use dist as the dependent (response) variable and speed as the independent (predictor) variable. Using trial and error, what seems to be a reasonable span? Comment on the graphical comparison between the estimate using your choice of span with the estimate using the span chosen by cross-validation.

```
library(datasets)
data(cars)
head(cars)
```

```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
## 6     9   10
```

2) **HWC** Page 661, Problem 2 (local averaging and choosing optimal span).

Consider the data set sunspots from Andrews and Herzberg (1985) as a response variable. For the predictor data x, use

```
data("sunspots")
x = c(1:length(sunspots))
```

Apply Friedman's smoother using trial and error to find a span that seems to work well with the data. Then find an estimate using the span determined by cross-validation. Describe the results (taking into account Comment 5).

```
y = sunspots
```

3) **W 2006** Page 121, Problem 3.

Download and save `forensic_glass_data_W2006.dat` available at Canvas @ Stanford

```
forensic.data = read.table("forensic_glass_data_W2006.dat", header = TRUE)
head(forensic.data)
```

```
##      RI    Na   Mg   Al    Si    K   Ca Ba   Fe type
## 1  3.01 13.64 4.49 1.10 71.78 0.06 8.75  0 0.00 WinF
## 2 -0.39 13.89 3.60 1.36 72.73 0.48 7.83  0 0.00 WinF
## 3 -1.82 13.53 3.55 1.54 72.99 0.39 7.78  0 0.00 WinF
## 4 -0.34 13.21 3.69 1.29 72.61 0.57 8.22  0 0.00 WinF
## 5 -0.58 13.27 3.62 1.24 73.08 0.55 8.07  0 0.00 WinF
## 6 -2.04 12.79 3.61 1.62 72.97 0.64 8.07  0 0.26 WinF
```

Let Y be refractive index and let x be aluminium content (the fourth variable).

```
library(dplyr)
Y = select(forensic.data, RI)
x = select(forensic.data, Al)
```

(1) Perform a nonparametric regression to fit the model $Y = r\left(x\right)+\epsilon$. Use the followi

(i) regressogram,

(ii) kernel,

(iii) local linear kernel regression,

(iv) spline.

In each case, use cross-validation to choose the amount of smoothing.

(2) Estimate the variance.

(3) Construct 95 percent confidence bands for your estimates.

4) **HWC** Page 644 Problem 6 (a) (Wavelet). Let y be the first 512 components in the sunspots data from package datasets.

(a) Use the `mra` command to plot f_4, f_5 and f_6 using the Haar wavelet. Describe the different characteristics of each of these three smooth approximations. For example, use $J = 5$ and either [[6]] or $S5 to find f_4.