Part 1: Theory

The aim of this exercise is to illustrate the behaviour of the expectation and variance of a simple non–parametric estimator of a regression curve. You'll find this exercise easier to carry out if you draw a large sketch. Consider the regression model

$$y_i = m(x_i) + e_i \ , \qquad i = 1, 2, 3$$

with $x_1 = 10,\ x_2 = 17,\ x_3 = 20$ and $e_i \overset{iid}{\sim} (0, \sigma^2)$.

Suppose that we were to estimate $m(x)$ using

$$\hat{m}(x) = \frac{1}{w_.} \sum_{i=1}^{n} w_i y_i \ ,$$

where $w_. = \sum_{i=1}^{n} w_i,\ w_i = w(x - x_i, h)$, and w is a rectangular weighting function:

$$w(z, h) = \begin{cases} 1 & \text{for } |z| \le h \\ 0 & \text{otherwise} \ . \end{cases}$$

(a) Consider the case $h = 2$.

i. For which intervals of $\Re$ is $\hat{m}(x)$ defined?

ii. Give a formula for $\hat{m}(x)$ in terms of $y_1,\ y_2,\ y_3$ for each of the intervals identified in (i).

iii. Find the formula for $E(\hat{m}(x))$ for each of the above intervals and sketch this.

iv. Find the formula for $\mathrm{Var}(\hat{m}(x))$ for each of the above intervals.

(b) Repeat (a) for $h = 5$.

Part 2: Practical

(a) Consider the following data:

```
data<-matrix(scan("D:/kursdaten_IntActDat/strength.txt "),ncol=4,byrow=T))
```

i. Figure out what each of the following three instructions do:

```
colnames(data)<-c("grip", "arm","raing"," sims")
n<-dim(data)[1]
pairs(data)
```

ii. Arrange the data as follows:

```
grip<-data[,1]
arm<-data[,2]
rating<-data[,3]
sims<-data[,4]
```

iii. Figure out what each of the following instructions do:

```
cbind(grip,arm)
cbind(1,arm)
rbind(grip,arm)
diag(1,3)
diag(c(1,3))
```

(b) Write a R function `wls(y,x,w)` that estimates the parmeters of the model

$$y = x\Theta + e$$

using the method of weighted least squares with the weights `w`, where `y` is a vector of observations, `x` is a matrix of covariates and `w` is a vector of weights. [Hint: the function must add in a column of ones to x in the function in order to estimate the intercept parameter. You can use the function `cbind` to do this.]

Perform a regression for the target variable **rating** using the covariates **grip** and **arm** and compare the results with the function `lsfit( )`.

(c) Explore the weighted least squares for the univariate case with target variable **grip** and covariate **arm** in the following ways:

i. Use equal weights and plot the data as well as the fitted line.

ii. Raise the weights of the observations with **100<arm<200** and estimate the parameters. Plot the fitted line in red.

iii. Experiment with the weights, e.g. set some of the weights equal to zero. Estimate the parameters and plot the fitted line for each case.