Stat 464, HW 11

1 In Class

In this homework, we will explore non-parametric regression. Because, we cannot actually calculate a function. We will need to set up grids over which to approximate the functions we want to calculate. The following R function will set up a grid and in order to calculate $\hat{r}(x)$ and can be found in `hw11functions.R`

```
kernelregress=function(x,y,h,delta){
xmin=floor(min(x)/delta)-1
xmax=ceiling(max(x)/delta)+1
xgrid=delta*(xmin:xmax)
rx=c()
for (i in 1:length(xgrid)){
rx[i]=nw(xgrid[i],x,y,h)
}
 returnvecs=list(xgrid,rx)
}
```

Recall how we can calculate a traditional regression. First, let's input some data. The following commands will input the data from exercise 10.3. The independent variable is miles per gallon and the dependent variable is speed for different models of cars.

```
temp=read.table("hw11.txt")
x=temp[[1]]
y=temp[[2]]
y=y[order(x,y)]
x=sort(x)
```

Note that these commands will sort the data in terms of x. To calculate a linear fit to the data, we can use the command `fit1=lm(y~x)`. To plot the data along with the fitted line, we may use

```
plot(x,y)
lines(x,lm(y~x)$fit)
```

To fit a qaudratic curve and plot it with the data, try

```
xsq=x^2
fit2=lm(y~x+xsq)
plot(x,y)
lines(x,lm(y~x+xsq)$fit)
```

2 Homework

1. Write the function `nw(xfixed,x,y,h)` where `xfixed` is the point at which you want to evaluate the function $\hat{r}(x)$. The vectors `x` and `y` are the data points and `h` is the bandwidth. The function should calculate $\hat{r}(x)$ at the point `xfixed`; therefore, only one number should be returned.

2. Using the data in `hw11.txt`, try out your function by calling the `kernelregress` function. You can use the `plotnw` function to look at the data with the fitted curve. Try to find a good choice for bandwidth by eye.

3. The function `riskgrid` creates a grid of bandwidth values and calculates and plots the estimated risk. Write the function `nwrisk` to calculate the risk. it should take as arguments the values `x`, `y`, and `h` –the data and the bandwidth. Recall a useful feature of R: If you would like to leave out the ith observation of `x`, then you can put `x[-i]`. Note that you will also need to use your `nw` function to write this new function.

4. Plot the risk. Do you think the minimal risk gives a good fit?

5. You can use the `se` function to calculate the standard error. To plot the data with the error bars, try (after setting `h` to the value you think is correct)

```
plotnw(x,y,h,0.25)
lines(kernelregress(x,y,h,0.25)[[1]],kernelregress(x,y,h,0.25)[[2]]-3*se(x,y,h,0.25)[[2]])
lines(kernelregress(x,y,h,0.25)[[1]],kernelregress(x,y,h,0.25)[[2]]+3*se(x,y,h,0.25)[[2]])
```

6. Find the residuals and plot them. (You may want to write a function to do this.) Do you think this data follows our model?