HW 8

Due: April 23trd at 3:00 pm

1. In this exercise we will compare different non-linear regression methods on a simulated data from a given curve.

   a) Use the following code to generate the data:

   ```
   set.seed(26)
   x = runif(100, 0, 1)
   y = x + x^2 + 4*(4*x-2)^3 + 0.1*exp(x) + rnorm(100,0,2)
   ```

   Plot the data to convince yourself the relationship is not linear

   b) Fit a cubic regression model and calculate its RMSE by comparing the predicted values to the *true* values computed from the true curve `x + x^2 + 4*(4*x-2)^3 + 0.1*exp(x)`

   c) Fit a step function using `breaks = c(0, 0.2, 0.8, 1)`. Calculate RMSE.

   d) Fit two piecewise quadratic polynomials using 0.5 as the break point. Calculate RMSE.

   e) Fit a spline regression using the quartiles as break points. Calculate RMSE.

   f) Fit a loess curve using the default span. Calculate RMSE.

   g) Compare and discuss the performance of all the above methods.

2. Use the infrared dataset from Givens and Hoeting's book, that can be read into R with the following command:

```
infrared =
read.table("http://www.stat.colostate.edu/computationalstatistics/datasets/infrared.dat",header=T)
```

   The goal is to estimate the density of the log of the variable labeled F12.

   a) Compute the bandwidth using Silverman's rule of thumb:

$$h = \left(\frac{4}{3n}\right)^{1/5} \hat{\sigma}$$

   b) Use normal kernel and *h* from part a) to obtain a density estimate.

   c) Repeat part b) with a bandwidth smaller than the one from part a) and another one greater than part a). Plot all three results on the same graph and discuss them.

   d) Construct and plot a frequency polygon density estimate using a bin width determined by the formula from the lecture notes.