Statistical Modeling 2
Exercise 3

February 15, 2017

Basic concepts

$$
\begin{aligned}
\mathrm{MSE}(\hat{f}, f) &= E\{[f(x) - \hat{f}(x)]^2\} \\
&= E[f(x)^2 + \hat{f}(x)^2 - 2f(x)\hat{f}(x)] \\
&= f(x)^2 + E[\hat{f}(x)^2] - E[\hat{f}(x)]^2 + E[\hat{f}(x)]^2 - 2f(x)E[\hat{f}(x)] \\
&= \{E[\hat{f}(x)^2] - E[\hat{f}(x)]^2\} + \{f(x)^2 + E[\hat{f}(x)]^2 - 2f(x)E[\hat{f}(x)]\} \\
&= V + B^2
\end{aligned}
$$

Curve fitting by linear smoothing

A

In the least square estimate, we minimize:

$$L = \sum_{i=1}^{n} (\hat{\beta} x_i - y_i)^2$$

We take the derivative and set to 0:

$$
\begin{aligned}
\partial L / \partial \hat{\beta} &= \sum_{i=1}^{n} 2(\hat{\beta} x_i - y_i) x_i = 0 \\
&\implies \hat{\beta}(\sum_{i=1}^{n} x_i^2) - \sum_{i=1}^{n} y_i x_i = 0 \\
&\implies \hat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}
\end{aligned}
$$

We have that the prediction is:

$$\begin{aligned}\hat{f}(x^*) &= \hat{\beta}x^* \\ &= \frac{\sum_{i=1}^{n} x_i x^* y_i}{\sum_{i=1}^{n} x_i^2} \\ &= \sum_{i=1}^{n} w(x_i, x^*) y_i\end{aligned}$$

where

$$w(x_i, x^*) = \frac{x_i x^*}{\sum_{i=1}^{n} x_i^2}$$

This weight function weights examples by the product with predictors, using all the points in the dataset. The K nearest function weights the K nearest neighbors equally and ignore all other points.

B

Code: kernel.r

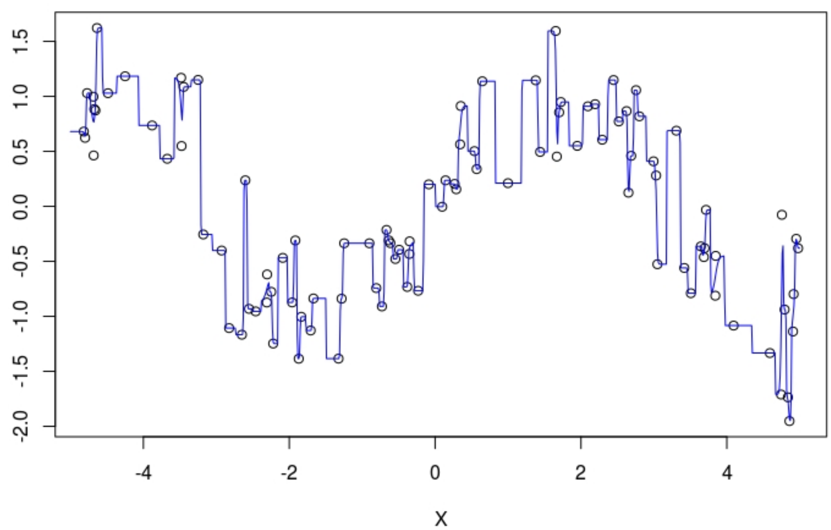

Figure 1: Bandwidth $h = 0.01$

The bandwidth h affects the smoothness of the prediction line. With smaller h, more weight is given to neighbor points and the line fluctuates. With larger

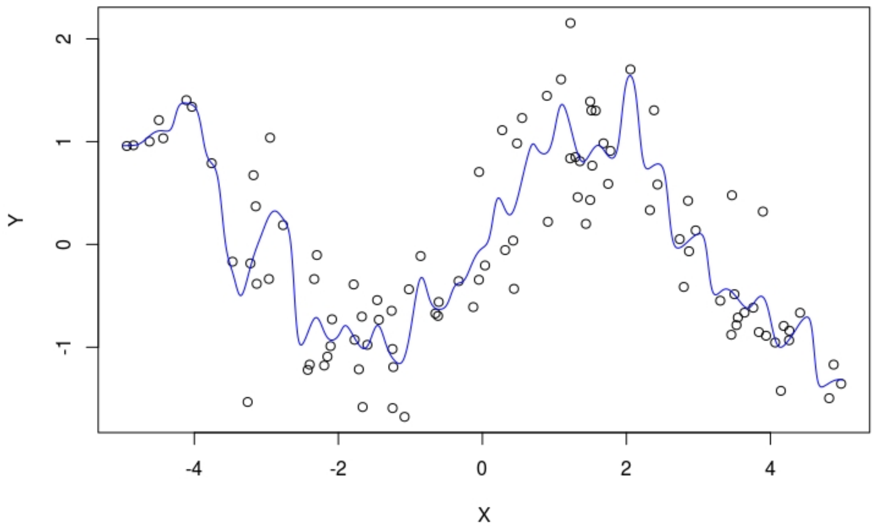

Figure 2: Bandwidth $h = 0.1$

h, more weight is given to points that are further away, resulting in a smoother line.

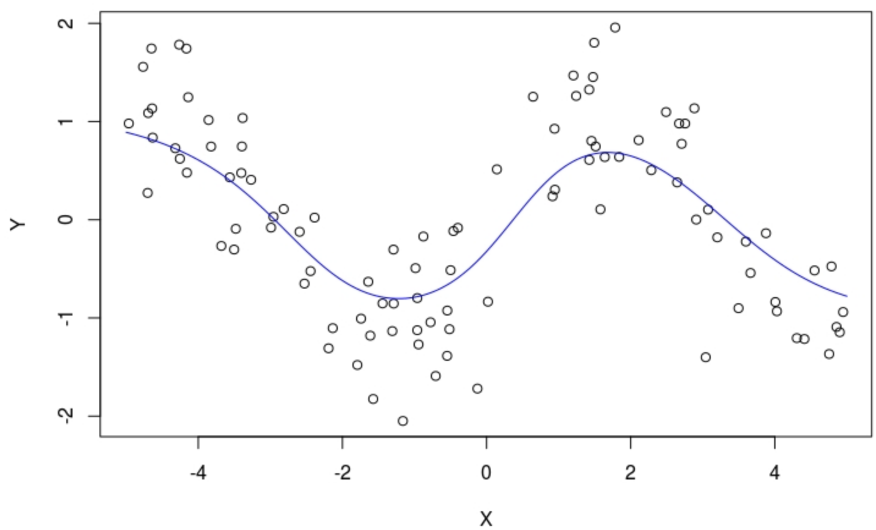

Figure 3: Bandwidth $h = 1$