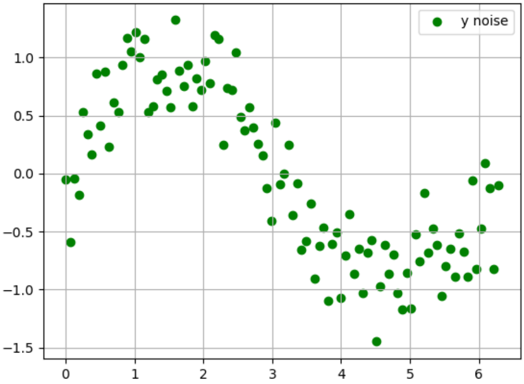
1 - Dimentional Kernel Smoothing Methods

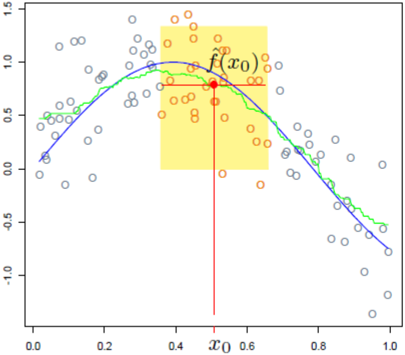
EE17BTECH11023 EE17BTECH11007

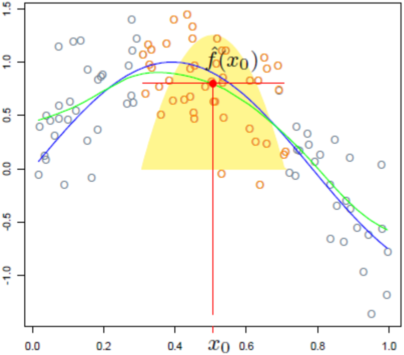
February 28, 2019

Question



Solution


1.5
1.0
0.5
0.0
-0.5
-1.0
0.0
0.2
0.4
x_0
0.6
0.8
1.0
$\hat{f}(x_0)$


1.5
1.0
0.5
0.0
-0.5
-1.0
0.0
0.2
0.4
x_0
0.6
0.8
1.0
$\hat{f}(x_0)$

Kernel Methods

Suppose that we have a dataset available with observations $(x_1, y_1), \ldots, (x_n, y_n)$. A simple kernel-based estimator of $f(x)$ is the *Nadaraya-Watson kernel regression* estimator, defined as

$$\hat{f}_h(x) = \frac{\sum_{i=1}^{n} K_h(x_i - x) y_i}{\sum_{i=1}^{n} K_h(x_i - x)}, \tag{1}$$

with $K_h(\cdot) = K(\cdot/h)/h$ for some kernel function $K(\cdot)$ and bandwidth parameter $h > 0$. The function $K(\cdot)$ is usually a symmetric probability density .

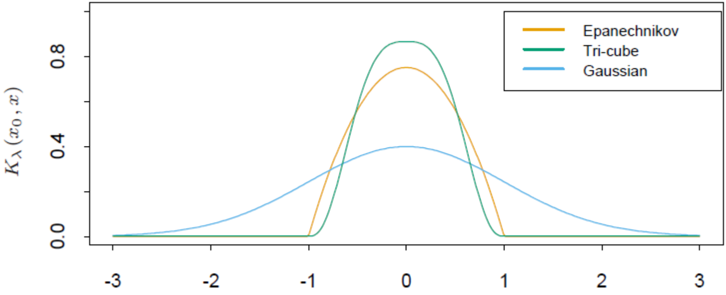
Matrix form:

$$\hat{f}_h(x) = W_x Y, \tag{2}$$

where $Y = (y_1, \ldots, y_n)^T$, $W_x = \text{diag}\{K_h(x_1 - x), \ldots, K_h(x_n - x)\}$

Popular Kernels

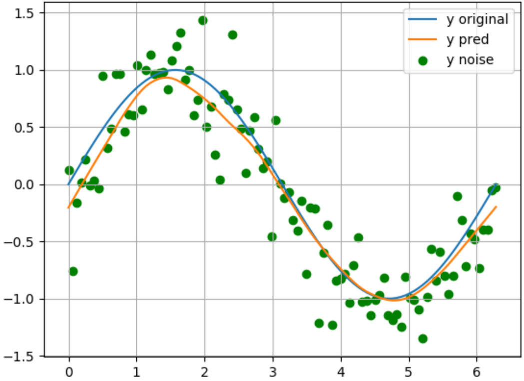
Examples of commonly used kernel functions are:
the Gaussian kernel $K(t) = (\sqrt{2\pi})^{-1} \exp(-t^2/2)$
the *Epanechnikov* kernel $K(t) = \max\{\frac{3}{4}(1 - t^2), 0\}$
the *Tri-cube* kernel $K(t) = \max\{(1 - |t|^3)^3, 0\}$.



Bandwidth of the Kernel

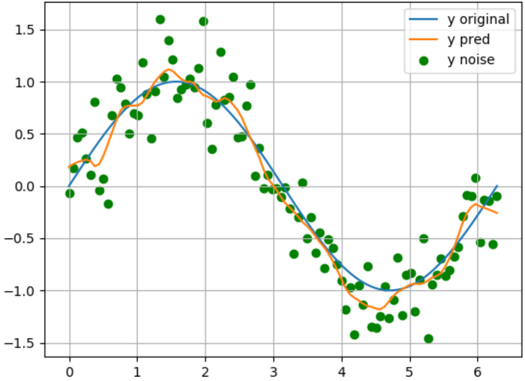
We used the subscript h in $\hat{f}_h(x)$ in (1) to emphasize the fact that the bandwidth h is the main determinant of the shape of the estimated regression, as demonstrated in Figure . When h is small relative to the range of the data, the resulting fit can be highly variable and look "wiggly."

Dependence on Bandwidth:



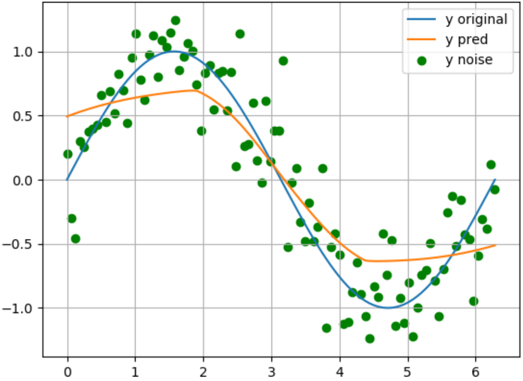
$h = 0.25$

Dependence on Bandwidth::



$h = 0.1$

Dependence on Bandwidth:



$h = 0.5$

Bandwidth selection:

We would like to find a value of h that minimizes the error between the estimated function and the true function.

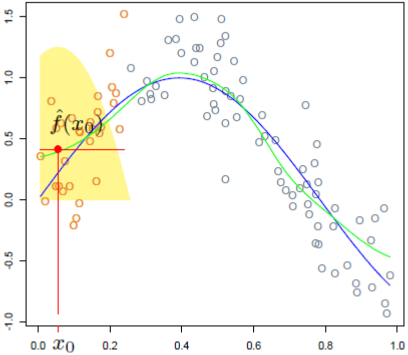
A natural measure is the MSE at the estimation point x, defined by
$E[(\hat{f}_h(x) - f_h(x))^2] = E[(\hat{f}_h(x) - f_h(x))]^2 + var(\hat{f}_h(x)))$.

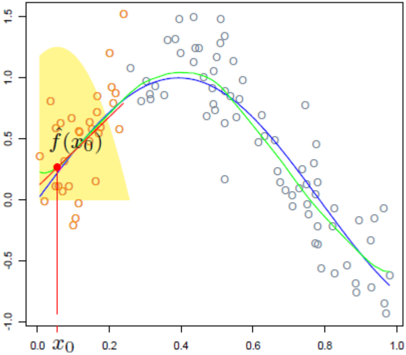
This expression is an example of the **bias-variance tradeoff**:

• The bias of an estimate is the systematic error incurred in the estimation

• The variance of an estimate is the random error incurred in the estimation

Problem


$\hat{f}(x_0)$
1.5
1.0
0.5
0.0
-0.5
-1.0
0.0
x_0
0.2
0.4
0.6
0.8
1.0


$\hat{f}(x_0)$
1.5
1.0
0.5
0.0
-0.5
-1.0
0.0
x_0
0.2
0.4
0.6
0.8
1.0

Local Linear Regression (Kernel Based)

A class of kernel-based estimators that generalizes the Nadaraya-Watson estimator in (1) is referred to as *local linear regression* estimators. At each location x, the estimator $\hat{f}_h(x)$ is obtained as the estimated intercept, $\hat{\beta}_0$, in the weighted least squares fit of a polynomial of degree p,

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1(x_i - x) - \cdots - \beta_p(x_i - x)^p\right)^2 K_h(x_i - x).$$

This estimator can be written explicitly in matrix notation as

$$\hat{f}_h(x) = e^T \left(X_x^T W_x X_x\right)^{-1} X_x^T W_x Y, \tag{3}$$

where the vector e is of length p + 1 and has a 1 in the first position and 0's elsewhere. $Y = (y_1, \ldots, y_n)^T$, $W_x = \text{diag}\{K_h(x_1 - x), \ldots, K_h(x_n - x)\}$ and

$$X_x = \begin{bmatrix} 1 & x_1 - x & \cdots & (x_1 - x)^p \\ \vdots & \vdots & & \vdots \\ 1 & x_n - x & \cdots & (x_n - x)^p \end{bmatrix}.$$

Note that, because the kernel K is symmetric, we could have written the argument of K as $(x_o - x_i)/h$. However, the notation used here emphasizes the fact that the local polynomial regression is a weighted regression using data centered around x_o. The least squares problem is then to minimize the weighted sum-of-squares function

$$(Y - X_x\beta)^T W_x(Y - X_x\beta), \tag{4}$$

$$\hat{\beta} = \left(X_x^T W_x X_x\right)^{-1} X_x^T W_x Y, \tag{5}$$

The quantity $\hat{f}_h(x)$ is then estimated by $\hat{\beta}$ as this defines the position of the estimated local polynomial curve at the point x_o. By varying the value of x_o, we can build up an estimate of the function $\hat{f}_h(x)$ over the range of the data.Thus we get equation (3).
For local linear regression $p = 1$.