Cross-validation and Splines

Vahid Partovi Nia

Advanced Machine Learning: Lecture 05

March 10, 2018

Outline

univariate function approximation

Suppose approximation of a good univariate function over a set of observed $(x_i, y_i), i = 1, \ldots, n$.

$$y_i = f(x_i) + \varepsilon_i \approx \sum_j \beta_j b_j(x_i)$$

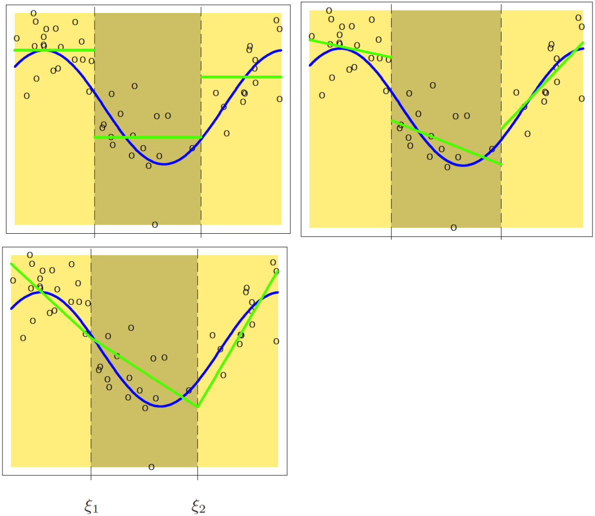
- polynomial base $x \in [-1, 1]$, $b_j(x_i) = x_i^j$
- Fourier base $x \in [-\pi, \pi]$,

$$y_i \approx \sum_{j=1}^{k} \beta_j^{(1)} \sin\left(\frac{2\pi j}{k}\right) + \beta_j^{(2)} \cos\left(\frac{2\pi j}{k}\right)$$

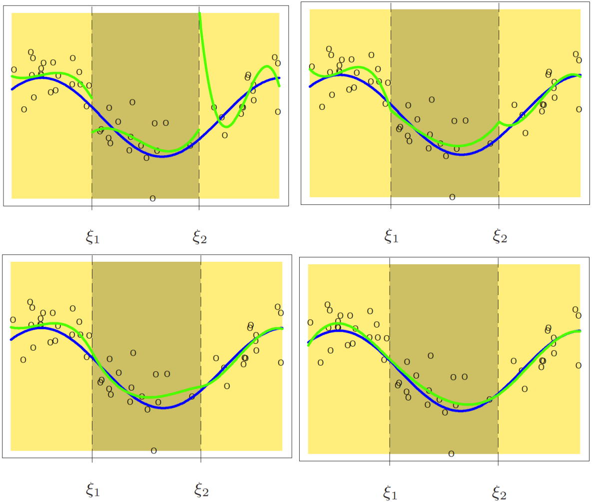
- Wavelet base of resolution k, $x \in [0, 2\pi]$

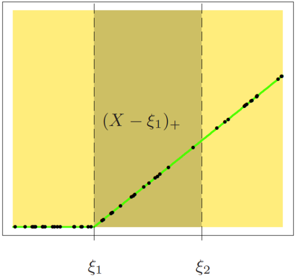
$$y_i \approx \sum_{j=1}^{2^k-1} \beta_j^{(k)} b_j^{(k)}(x_i)$$

Piecewise polynomials



Cubic





Multivariate = Additive Model

$$y_i \approx f(\mathbf{x}_i) = \sum_{j=1}^{p} f_j(x_{ij})$$

B splines of order m are piece-wise polynomial functions

$$\begin{aligned} h_j(x) &= x^j, j = 0, \ldots, m \\ h_{m+l}(x) &= \{\max(0, x - \xi_l)\}^{m-1} \end{aligned}$$

Bsplines

- Bsplines of order $m > 1$ are continuous
- They have continuous derivative up to $m - 2$
- Cubic splines are Bsplines of order $m = 4$
- Natural splines, are cubic splines with $x_i = \xi_i$
- Smoothing splines are natural splines are generalized ridge with penalization $\boldsymbol{\beta}^\top \boldsymbol{\Omega} \boldsymbol{\beta}$
- $\boldsymbol{\Omega} = [\omega_{l,k}]$ is not a diagonal matrix.
- $\omega_{l,k} = \int h''_{m+l}(x) h''_{m+k}(x) dx$

Reproducing Kernel Hilbert Space

- Suppose

$$\hat{f} = \operatorname{argmin} \sum_{i=1}^{n} L(y_i, f(x_i)) + \lambda ||f||$$

- Many function estimation problems fall in this context, like splines, support vector machines, kernel smoothers, local regression, etc.
- Suppose a positive definite kernel function $K(x, x')$ defines the space of the Hilbert space.
- Take some arbitrary values of $x_j \in \mathbb{R}$ and generate a function $f(x) = \sum_j \beta_j K(x, x_j) \in \mathcal{H}_K$

- Suppose $K(x, x') = \sum_{j=1}^{\infty} \gamma_j \phi_j(x) \phi_j(x')$ with $\gamma_j > 0, \sum_j \gamma_j^2 < \infty$
- Then all function in $\mathcal{H}_K$ have an eigen expansion $f(x) = \sum_{j=1}^{\infty} c_j \phi_j(x)$
- Define $||f|| = \sum_{j=1}^{\infty} c_j^2 / \gamma_j$

Solution to

$$\begin{aligned}\hat{f}(x) &= \operatorname{argmin}\ L(y_i, f(x_i)) + \lambda ||f|| \\ &= \operatorname{argmin}\ \sum_{i=1}^{n} L\left(y_i, \sum_{j=1}^{\infty} c_j \phi_j(x_i)\right) + \lambda \sum_{j=1}^{\infty} c_j^2/\gamma_j\end{aligned}$$

has a finite dimensional solution

$$\hat{f}(x) = \sum_{i=1}^{n} \beta_i K(x, x_i)$$

and

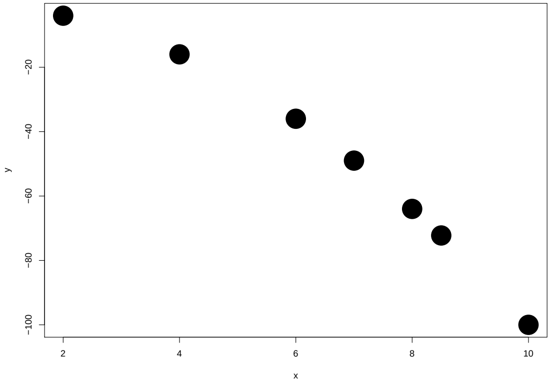
$$||f|| = \sum_{i=1}^{n} \sum_{i'=1}^{n} K(x_i, x_{i'}) \beta_i \beta_i'$$

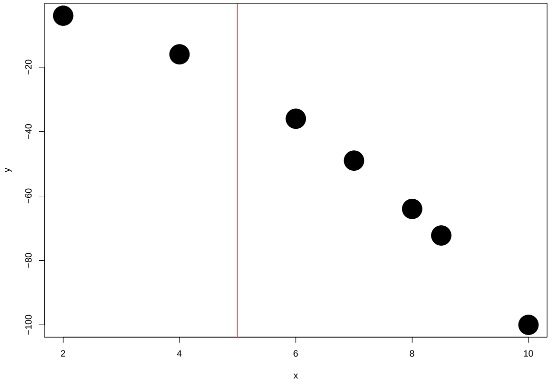
Generalized ridge

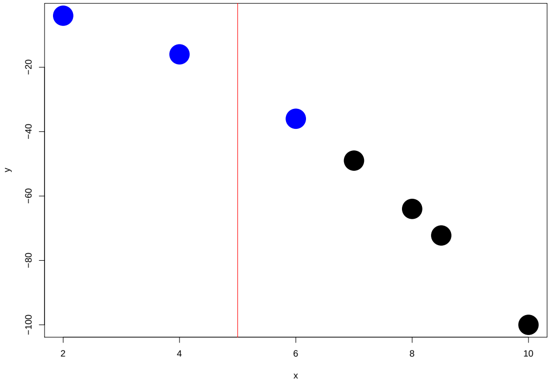
$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \operatorname{argmin}\, (\mathbf{y} - \mathbf{K}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{K}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^\top \mathbf{K}\boldsymbol{\beta} \\
\hat{f}(x) &= \mathbf{K}\hat{\boldsymbol{\beta}}
\end{aligned}
$$

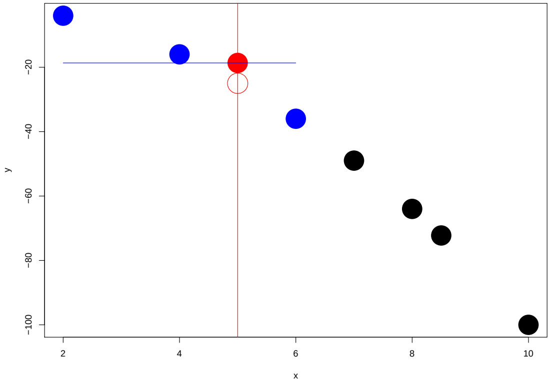
Take $L(y_i, f(x_i)) = \max(0, 1 - y_i f(x_i))$ to produce support vector machines.

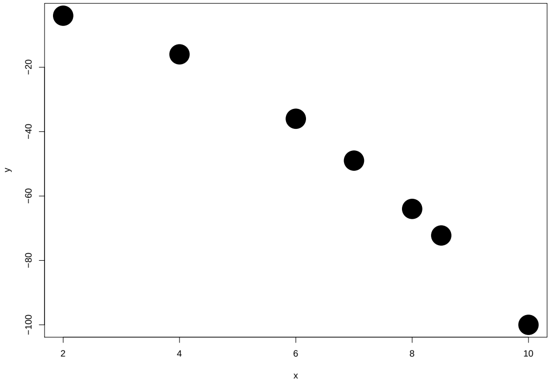
Kernel Smoothing

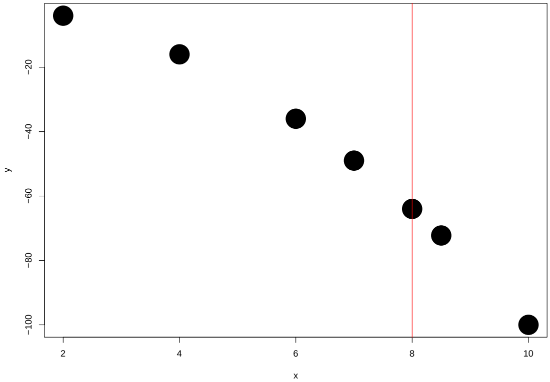

−20
−40
−60
−80
−100
y
2
4
6
8
10
x

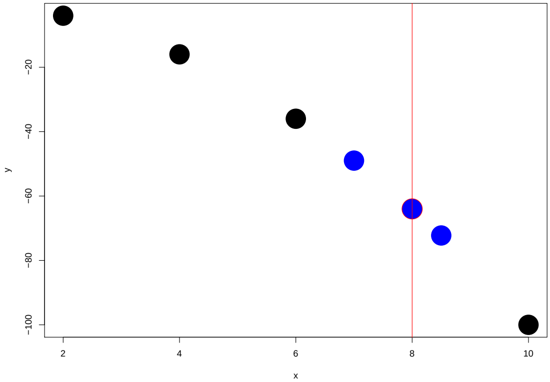

-20
-40
-60
-80
-100
y
2
4
6
8
10
x

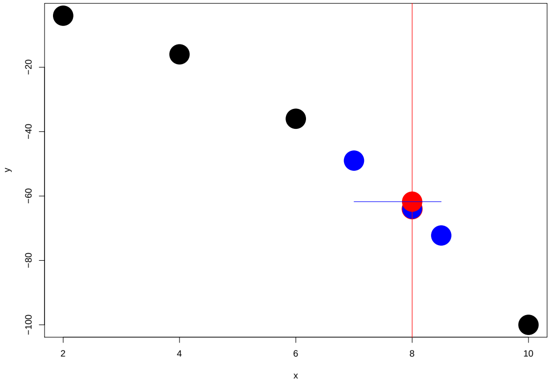

−20
−40
−60
−80
−100
y
2
4
6
8
10
x


−20
−40
−60
−80
−100
y
2
4
6
8
10
x


−20
−40
−60
−80
−100
y
2
4
6
8
10
x


−20
−40
−60
−80
−100
y
2
4
6
8
10
x


−20
−40
−60
−80
−100
y
2
4
6
8
10
x


y
−20
−40
−60
−80
−100
2
4
6
8
10
x

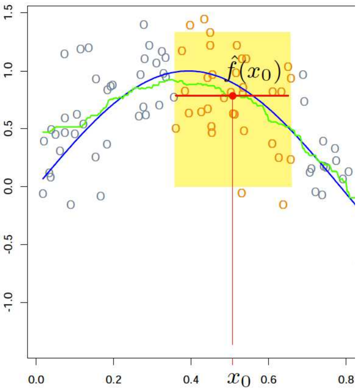
non-parametric function

- Define a neighbourhood of x
- $\hat{f}(x_0) = \frac{1}{||N(x_0)||} \sum_{i \in N(x_0)} y_i$
- $w_i(x_0) = \mathbb{I}_A(x_0), A = \{x \in \mathbb{R}, ||x_0 - x_i|| < \lambda\}$

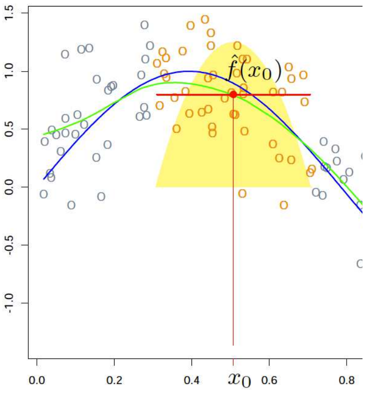
$$\hat{f}(x_0) = \frac{\sum_{i=1}^n w_i(x_0) y_i}{\sum_{i=1}^n w_i(x_0)}$$

- $w_i(x_0) = K(x_0, x_i)$
-
$$\hat{\beta}_0(x_0) = \text{argmin} \sum_{i=1}^n K(x_0, x_i)(y_i - \beta_0)^2$$

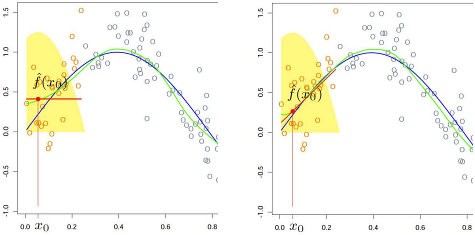
Uniform Kernel



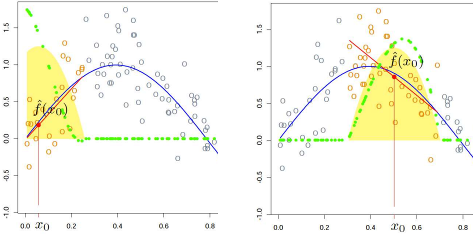
Smooth Kernel



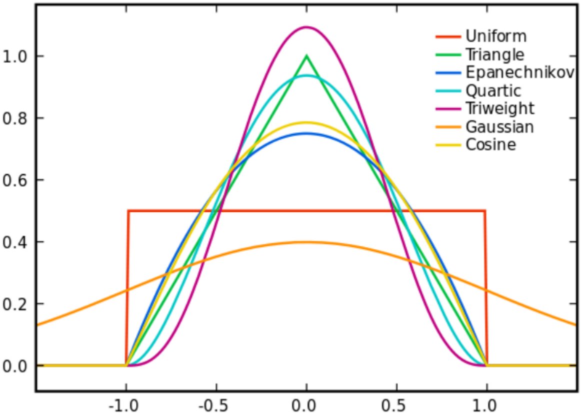
Constant versus linear



transformed kernel



Kernel functions



Smooth histograms