Lecture 16: Kernel Regression

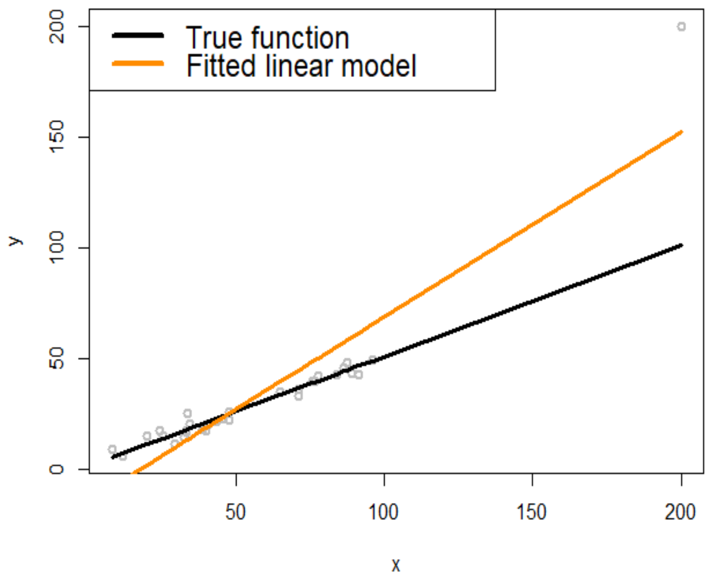
Instructor: Prof. Shuai Huang

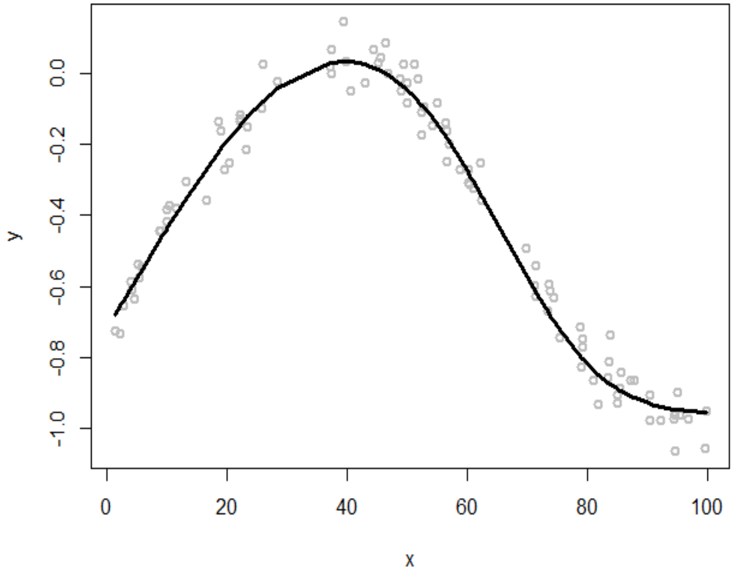
Industrial and Systems Engineering

University of Washington

Limitations of linear regression model

- A global model that generalizes local data points to a central model





How far is it from linear regression to nonlinear regression?

- Let's look at the simple linear regression problem $y = \beta_0 + \beta_1 x$. Let's further simplify it by assuming that we know the mean of y is zero, so is the mean of x. This will lead to the model as $y = \beta_1 x$ and the estimator of β_1 as

$$\beta_1 = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)}{\sum_{i=1}^{n} x_i^2}.$$

- Thus, when we try to make prediction on a new data point with a given x^*, the prediction y^* will be

$$y^* = x^* \frac{\left(\sum_{i=1}^{n} x_i y_i\right)}{\sum_{i=1}^{n} x_i^2}.$$

- This could be further reformed as:

$$y^* = \sum_{i=1}^{n} y_i \frac{x_i}{\sum_{i=1}^{n} x_i^2} x^*,$$

- which is equivalent with

$$y^* = \sum_{i=1}^{n} y_i \frac{x_i x^*}{n S_x^2}.$$

Generalize this insights into development of nonlinear models

- We then pursue a generalized family of model, defined as:

$$y^* = \sum_{n=1}^{N} y_n w(x_n, x^*).$$

- Here, $w(x_n, x^*)$ is the weight that characterizes the similarity between x^* and the existing data points, x_n for $n = 1,2, \ldots, N$.
- Roughly speaking, there are two types of similarity metric.
- One is the **K-nearest neighbor (KNN) smoother**:

$$w(x_n, x^*) = \begin{cases} \frac{1}{k}, & if\, x_n\ is\ one\ of\ the\ k\ nearest\ neighbors\ of\ x^* \\ 0, & if\, x_n\ is\ NOT\ in\ the\ k\ nearest\ neighbors\ of\ x^* \end{cases}.$$

- Another is the **kernel smoother**:

$$w(x_n, x^*) = \frac{K(x_n, x^*)}{\sum_{n=1}^{N} K(x_n, x^*)}.$$

Some examples of the kernel functions

Kernel function	Mathematical form	Parameters
Linear	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^T \boldsymbol{x}_j$	$null$
Polynomial	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^T \boldsymbol{x}_j + 1)^q$	q
Gaussian radial basis	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\gamma \lVert x_i - x_j \rVert^2}$	$\gamma \geq 0$
Laplace radial basis	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\gamma \lVert x_i - x_j \rVert}$	$\gamma \geq 0$
Hyperbolic tangent	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(\boldsymbol{x}_i^T \boldsymbol{x}_j + b)$	b
Sigmoid	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(a\boldsymbol{x}_i^T \boldsymbol{x}_j + b)$	a, b
Bessel function	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \dfrac{bessel_{v+1}^n(\sigma \lVert \boldsymbol{x}_i - \boldsymbol{x}_j \rVert)}{(\lVert \boldsymbol{x}_i - \boldsymbol{x}_j \rVert)^{-n(v+1)}}$	σ, n, v
ANOVA radial basis	$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left(\sum_{k=1}^n e^{-\sigma\left(x_i^k - x_j^k\right)}\right)^d$	σ, d

R lab

- Download the markdown code from course website
- Conduct the experiments
- Interpret the results
- Repeat the analysis on other datasets