Nonparametric methods

Parametric methods

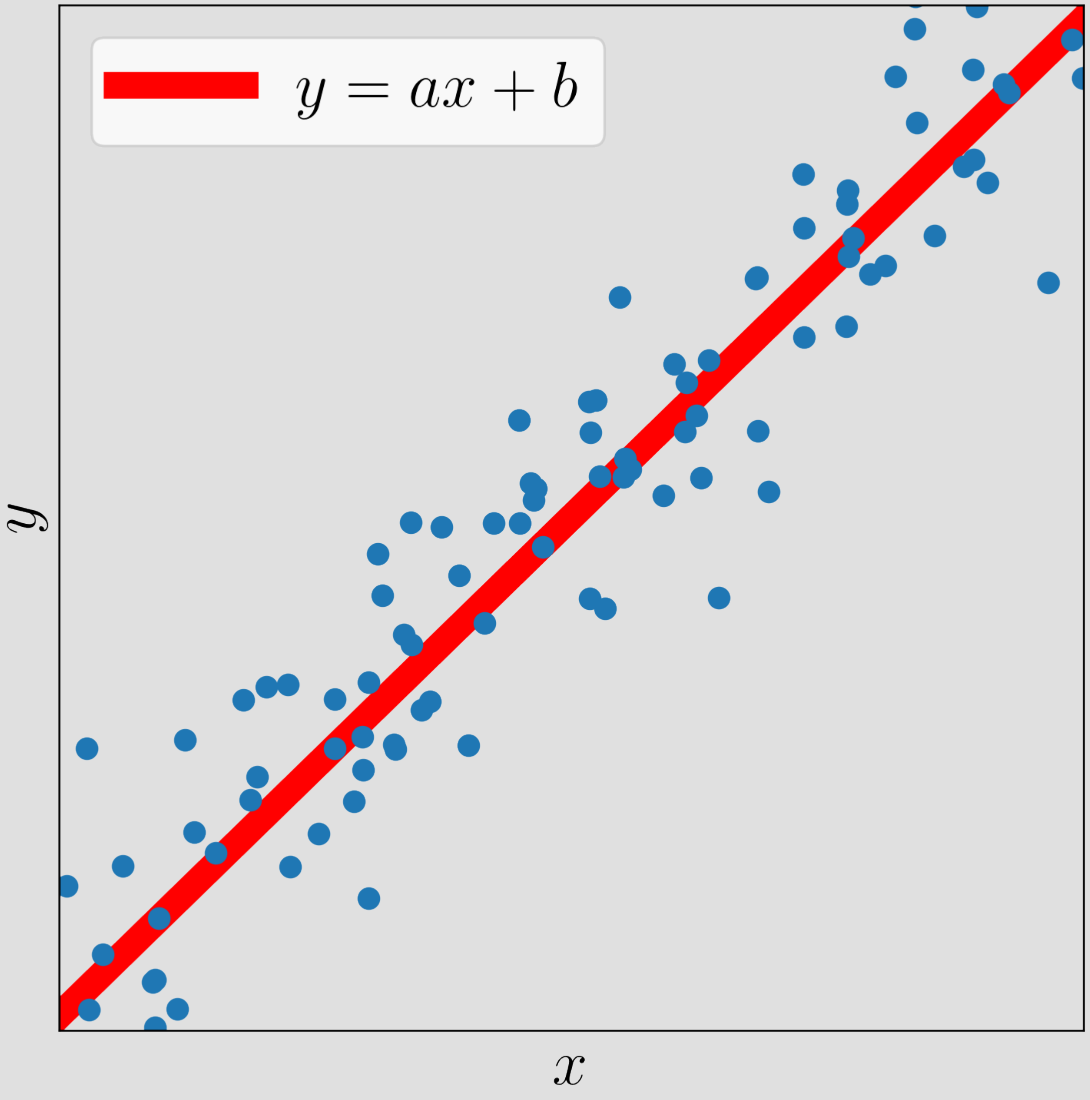
1. Define parametric model:

$$p(y|X, \theta)$$

2. Find best parameters using MAP estimation:

$$p(\theta|y, X) \rightarrow \max_{\theta}$$

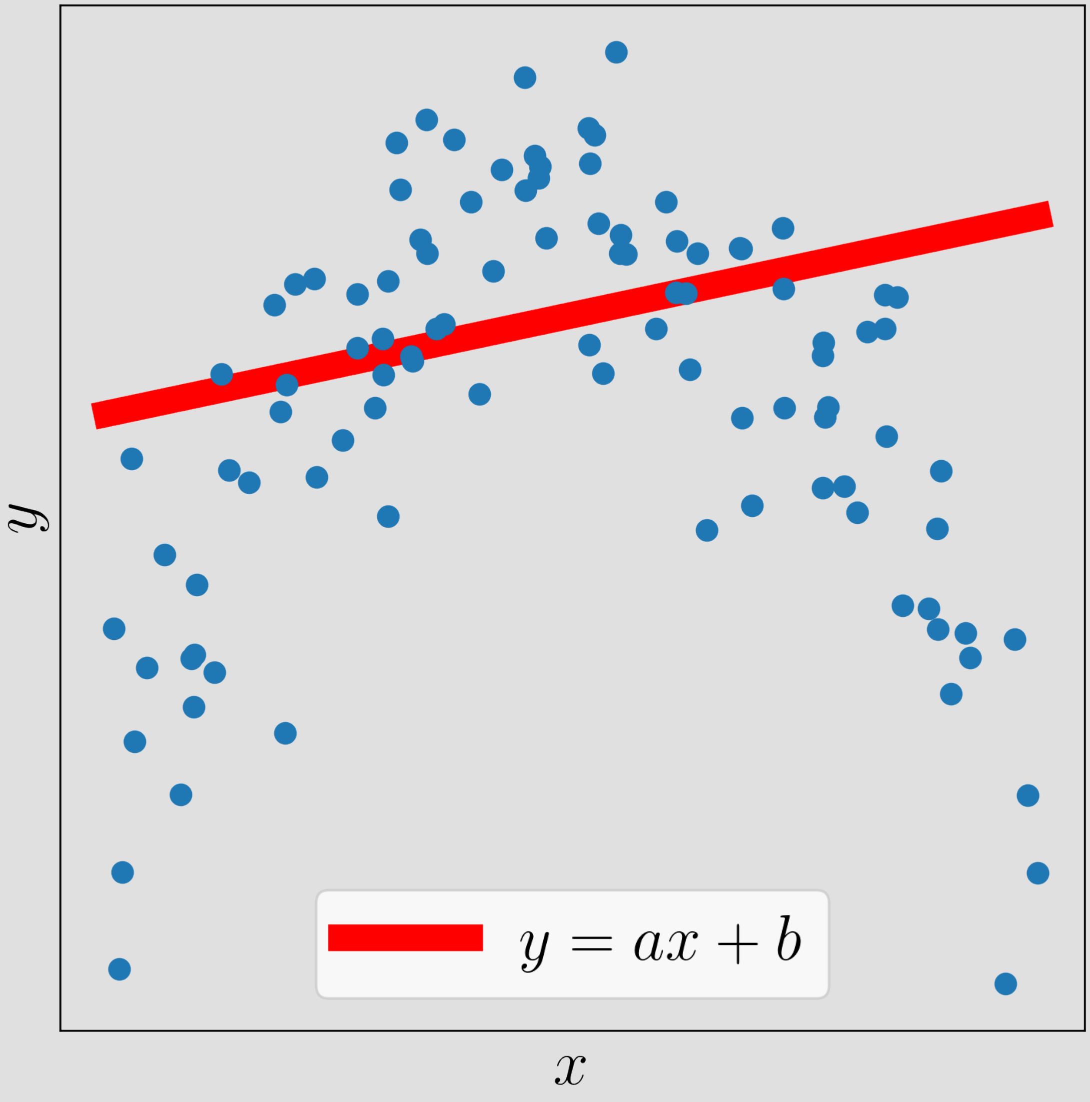
Parametric methods



Parameters:

$$\theta = \{a, b\}$$

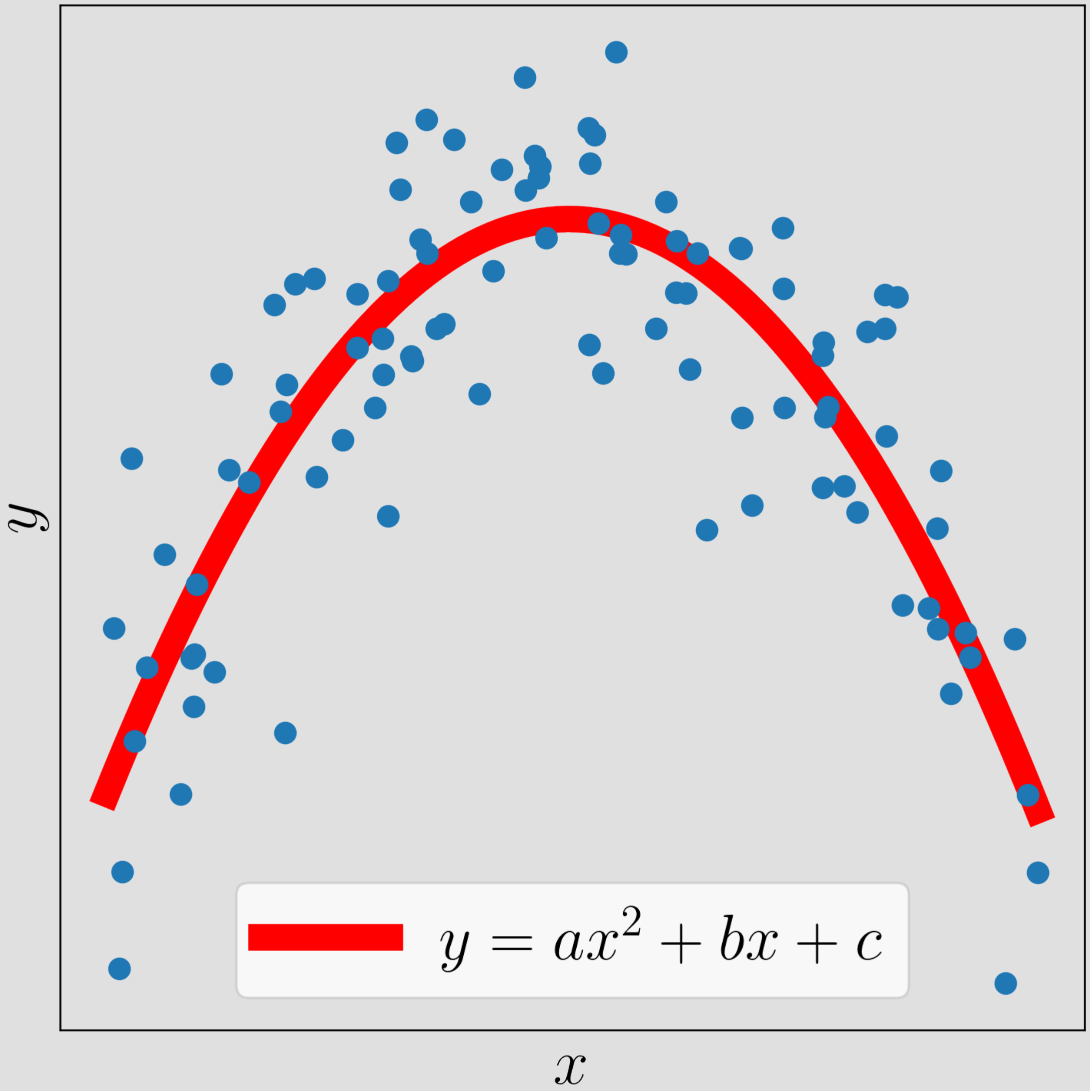
Parametric methods



Parameters:

$$\theta = \{a, b\}$$

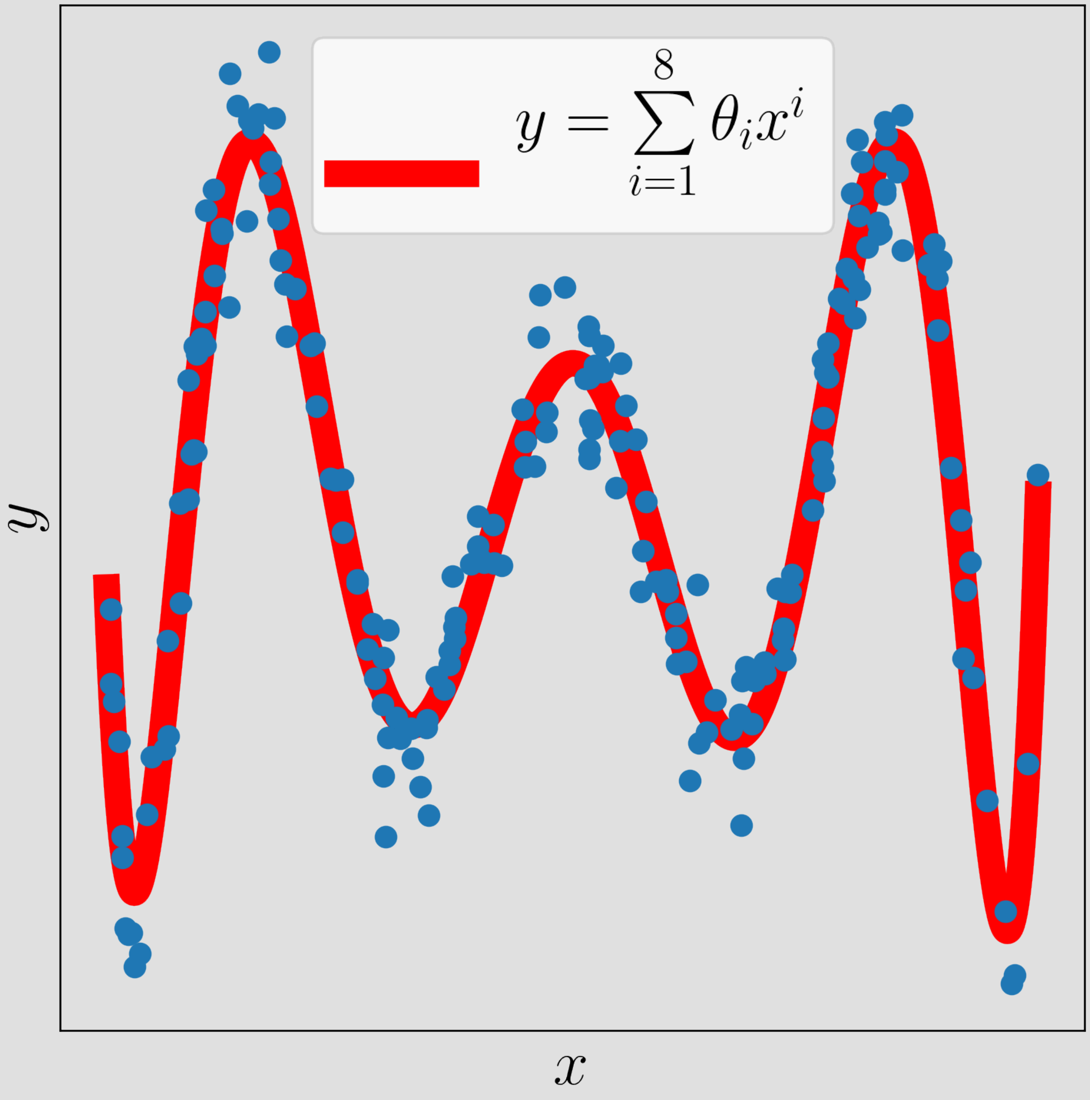
Parametric methods



Parameters:

$$\theta = \{a, b, c\}$$

Parametric methods



Parameters:

$$\theta = \{\theta_1, \theta_2, \ldots, \theta_8\}$$

Non-parametric methods

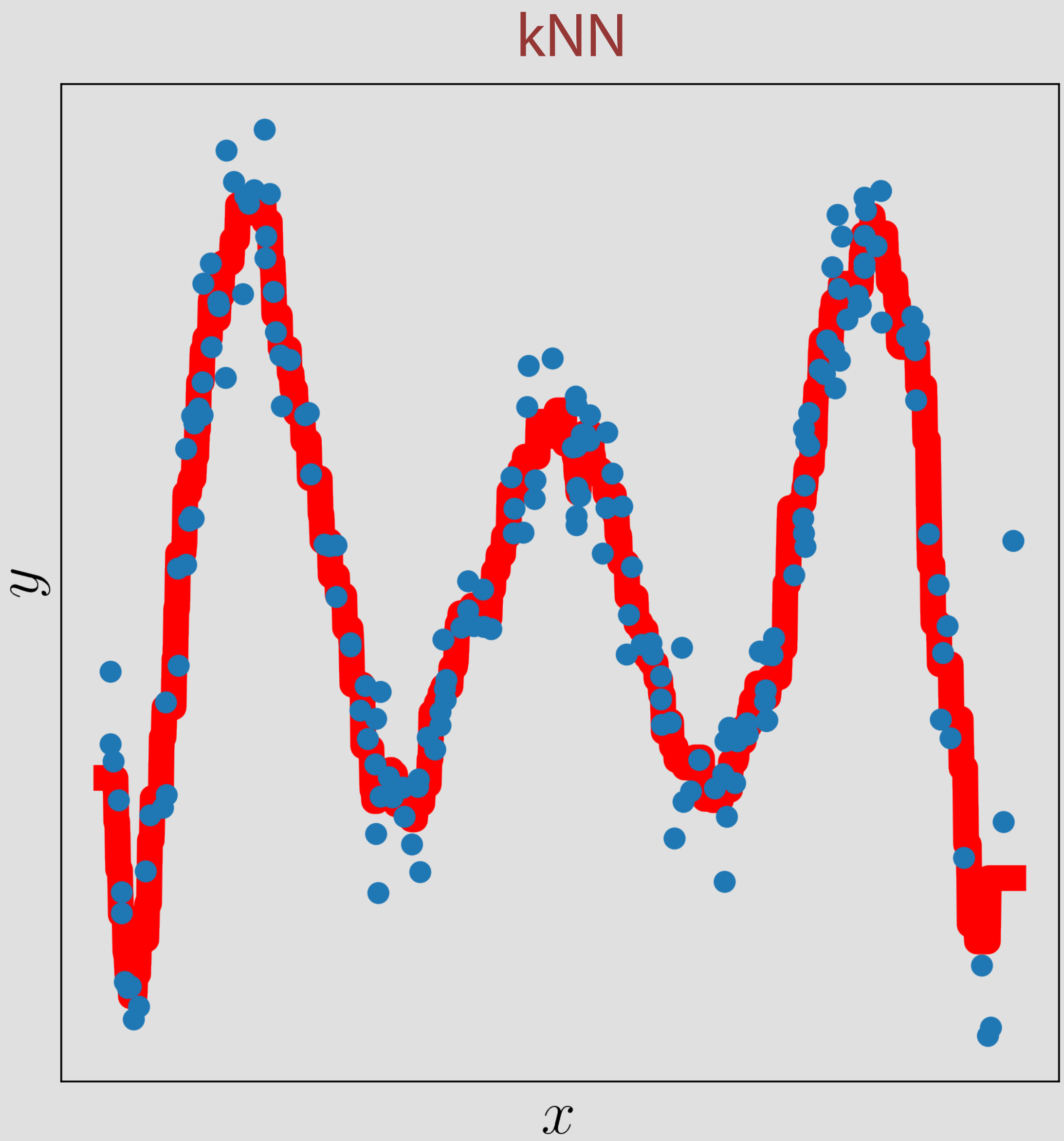
Parametric methods:

- Fixed number of parameters

Non-parametric:

- Number of parameters depends on dataset size

Non-parametric methods

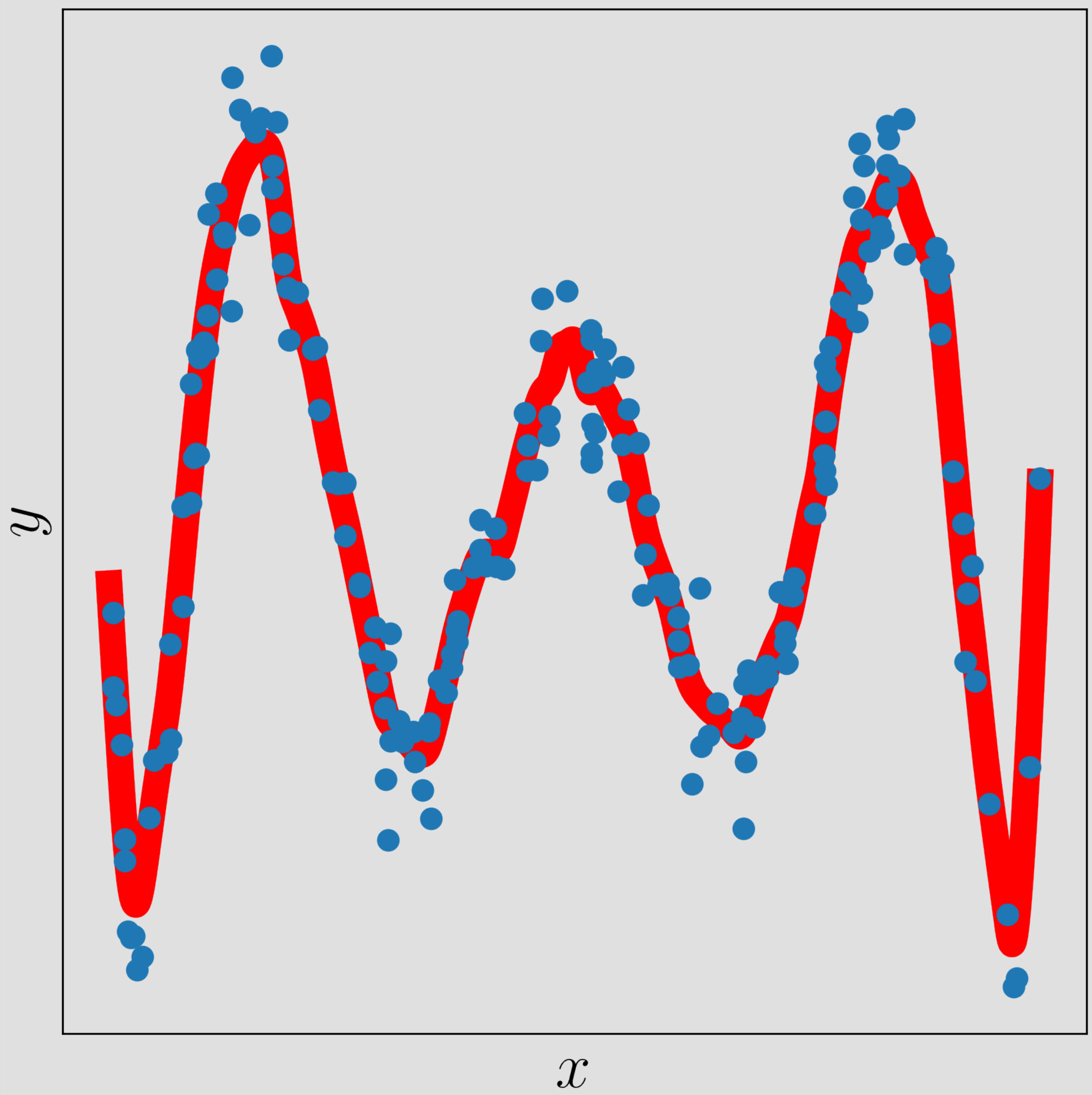


K-nearest neighbors:

$$y = \frac{1}{5}\sum_{i=1}^{5} y_{(i)}$$

Non-parametric methods

Nadaraya-Watson



K-nearest neighbors:

$$y = \frac{1}{5}\sum_{i=1}^{5} y_{(i)}$$

Nadaraya-Watson:

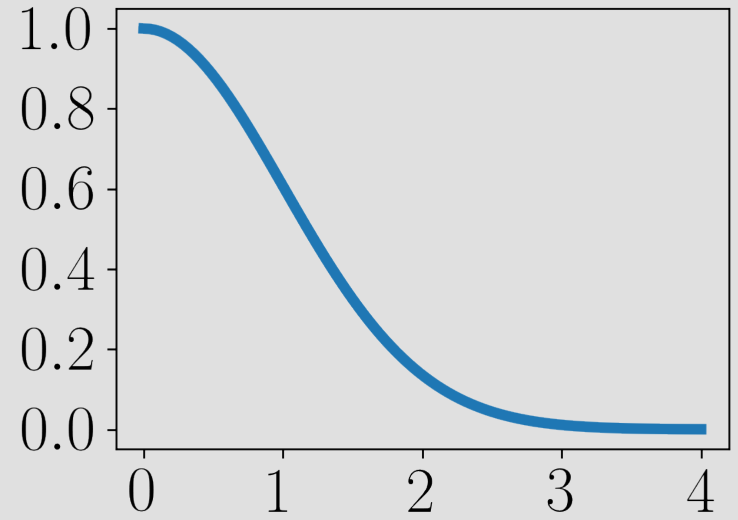
$$y(x) = \sum_{i=1}^{N} w_i(x) y_i$$

$$w_i(x) = \frac{K(x, x_i)}{\sum_{j=1}^{N} K(x, x_j)}$$

Some kernels

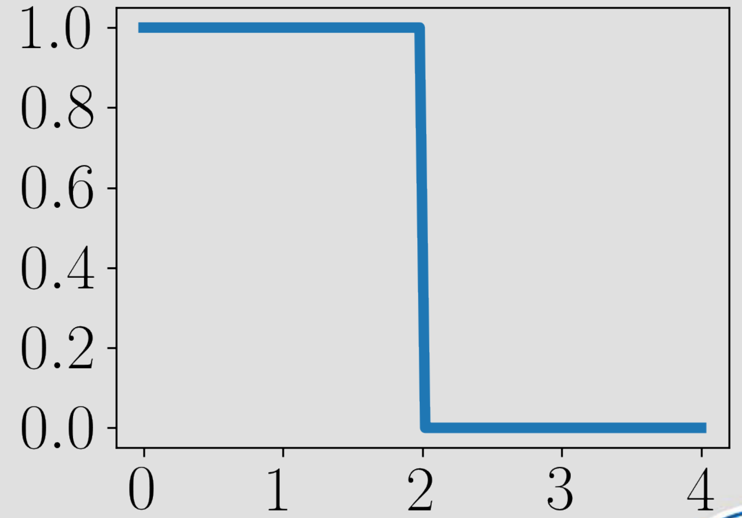
Gaussian

$$K(x_1, x_2) = e^{-\frac{1}{2\sigma^2}\|x_1 - x_2\|^2}$$



Uniform

$$K(x_1, x_2) = \mathbb{I}[\|x_1 - x_2\| < h]$$



Parametric vs Non-parametric

Parametric:

- Limited complexity
- Faster inference
- Slow learning

Non-parametric:

- Arbitrary complex
- Need to process all data for prediction
- Learning: remember all data

Gaussian Processes

Can estimate uncertainty of predictions