Chapter 3 Programming: KRR

AMA4680
Fu Guanxing
guanxing.fu@polyu.edu.hk

Kernel ridge regression

- Given data $\{(x_i, y_i)\}_{i=1}^{n}$ and a (strictly) positive kernel K, KRR proceeds as follows.
  - Step 1: training. The model is determined by the minimizer

$$c^* = \arg\min_{c \in \mathbb{R}^n} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} c_j K(x_j, x_i) - y_i \right)^2 + \lambda c' \mathbb{K} c \right\}.$$

Or in a matrix form,

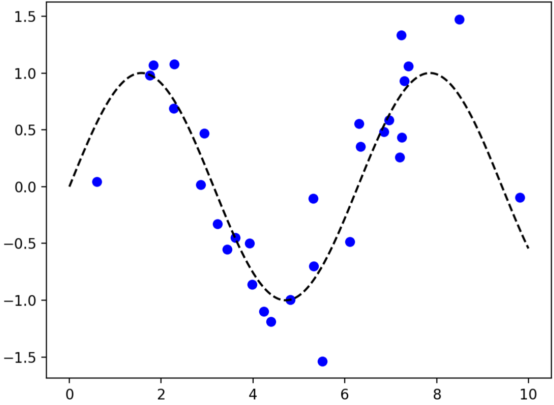
$$\begin{aligned} c^* &= \arg\min_{c \in \mathbb{R}^n} \left\{ \|\mathbb{K}c - y\|^2 + n\lambda c' \mathbb{K} c \right\} \\ &= (\mathbb{K} + \lambda n I)^{-1} y. \end{aligned}$$

  - Step 2: prediction.

$$\hat{f}_{\mathsf{krr}}(x) = \sum_{i=1}^{n} c_i^* K(x_i, x); \qquad c^* = (c_1^*, \cdots, c_n^*)'.$$

Example of KRR with synthetic data

- **Step 1: Data generation:** $\{(x_i, y_i)\}_{i=1}^{30}$
  - Let $f = \sin(x)$, $x \in [0, 10]$, be the true function.
  - Let $y_i = \sin(x_i) + \epsilon_i$, $i = 1, \cdots, 30$, where $x_i \sim$ Uniform$([0, 10])$ and $\epsilon_i \sim$ Normal$(0, 0.09)$.



- **Step 2: Kernel construction:** Gaussian Kernel.

$$K(x, y) = \exp\left(-\frac{(x-y)^2}{\sigma^2}\right),$$

with $\gamma = \frac{1}{\sigma^2} = 1$.

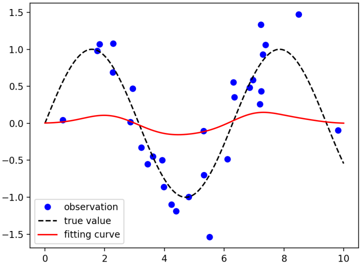
- **Step 3: Training:**[1]

$$c^* = (\mathbb{K} + n\lambda I)^{-1} y \quad \Leftrightarrow \quad (\mathbb{K} + n\lambda I) c^* = y.$$

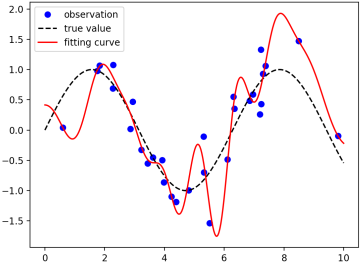
- **Step 4: Prediction** for $x \in [0, 10]$:

$$\hat{f}_{\mathsf{krr}}(x) = \sum_{i=1}^{n} c_i^* K(x, x_i).$$

[1] One can solve c^* by either calculating inverse or by solving a linear equation; the latter one is faster.



1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
2
4
6
8
10
observation
true value
fitting curve



2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
2
4
6
8
10
observation
true value
fitting curve