Nonparametric Regression

Yuling Yan

University of Wisconsin–Madison, Fall 2024

Nonparametric regression

Setup: we have data $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$ satisfying

$$y_i = f^\star(\boldsymbol{x}_i) + \varepsilon_i$$

- unknown $f^\star \in \mathcal{F}$ where $\mathcal{F}$ is certain function class
- i.i.d. Gaussian noise $\varepsilon_1, \ldots, \varepsilon_n \sim \mathcal{N}(0, \sigma^2)$
- fixed design ($\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are fixed) or random design ($\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{\text{i.i.d.}}{\sim} \rho$)

Goal: estimate $f^\star$ using the data

Error metric: for any estimator f, consider squared L_2 norm

$$\|f - f^\star\|_n^2 := \frac{1}{n} \sum_{i=1}^{n} \big(f(\boldsymbol{x}_i) - f^\star(\boldsymbol{x}_i)\big)^2 \qquad \text{(for fixed design)}$$

$$\|f - f^\star\|_\rho^2 := \mathbb{E}_{\boldsymbol{x}\sim\rho}\big[(f(\boldsymbol{x}) - f^\star(\boldsymbol{x}))^2\big] \qquad \text{(for random design)}$$

Nonparametric least squares

Least squares estimate:

$$\widehat{f} := \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} \big(f(\boldsymbol{x}_i) - y_i\big)^2$$

- this estimator depends on $\mathcal{F}$
- computational: how to compute this least squares estimate?
- statistical: what is the convergence rate of $\widehat{f}$?

Our plan: focus on $\mathcal{F}$ that leads to *computationally feasible* estimate

- **isotonic regression:** $\mathcal{F} = \{$monotone function in $\mathbb{R}\}$
- **convex regression:** $\mathcal{F} = \{$convex function in $\mathbb{R}^d\}$
- **kernel ridge regression:** $\mathcal{F} =$ reproducing kernel hilbert space (RKHS)

Isotonic regression

Isotonic regression: setup

- **Setup:** $\mathcal{F}$ is the set of increasing (or decreasing) function in $\mathbb{R}$
- Suppose without loss of generality that $x_1 < x_2 < \cdots < x_n$
- By solving the following convex optimization problem

$$(\widehat{f}_1, \ldots, \widehat{f}_n) \coloneqq \mathop{\arg\min}_{f_1 \le \cdots \le f_n} \sum_{i=1}^{n} (y_i - f_i)^2,$$

  the least squares estimate is any increasing function $\widehat{f}(x)$ such that

$$\widehat{f}(x_i) = \widehat{f}_i \qquad (i = 1, \ldots, n).$$

- **Key observation:** $f^\star(x)$ is only identifiable for $x \in \{x_1, \ldots, x_n\}$
- It is more reasonable to consider the error under fixed design

Isotonic regression: convergence rate

Theorem 5.1

Consider the class of increasing function with bounded variation

$$\mathcal{F} = \{f : [0,1] \to [0,1] \mid f \text{ is monotonically increasing}\}.$$

Then the isotonic regression estimate $\widehat{f}$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[\big(\widehat{f}(x_i) - f^\star(x_i)\big)^2\Big] \lesssim \Big(\frac{\sigma^2}{n}\Big)^{2/3}$$

Remark: in comparison, without using the monotonic structure, the squared error of MLE does not decrease as n grows:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[\big(y_i - f^\star(x_i)\big)^2\Big] = \sigma^2.$$

Convex regression

Convex regression: setup

- **Setup:** $\mathcal{F}$ is the set of convex function in $\mathbb{R}^d$
- By solving the following convex optimization problem

$$\begin{array}{ll} \underset{f_1,\ldots,f_n\in\mathbb{R},\boldsymbol{g}_1,\ldots,\boldsymbol{g}_n\in\mathbb{R}^d}{\text{minimize}} & \sum_{i=1}^n (y_i - f_i)^2 \\ \text{subject to} & f_j \geq f_i + \boldsymbol{g}_i^\top(\boldsymbol{x}_j - \boldsymbol{x}_i) \quad \text{for all } 1 \leq i,j \leq n \end{array}$$

the least squares estimate is any convex function $\widehat{f}(\boldsymbol{x})$ such that

$$\widehat{f}(\boldsymbol{x}_i) = \widehat{f}_i, \quad \widehat{\boldsymbol{g}}_i \in \partial \widehat{f}(\boldsymbol{x}_i) \qquad (i = 1,\ldots,n).$$

- **Key observation:** $f^\star(\boldsymbol{x})$ is only identifible for $\boldsymbol{x} \in \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\}$
- It is more reasonable to consider the error under fixed design

Convex regression: convergence rate

Theorem 5.2

Consider the class of convex function in $\mathbb{R}$

$$\mathcal{F} = \{f : [0,1] \to [0,1] \mid f \text{ is convex}\}.$$

Then the convex regression estimate $\widehat{f}$ *satisfies*

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(\widehat{f}(x_i) - f^\star(x_i)\right)^2\right] \lesssim \left(\frac{\sigma^2}{n}\right)^{4/5}$$

Remark: for convex regression in $\mathbb{R}^d$, the error is of order $n^{-4/(d+4)}$