Nonparametric Regression



Yuling Yan
University of Wisconsin–Madison, Fall 2024

Nonparametric regression

Setup: we have data $(x_1, y_1), \ldots, (x_n, y_n)$ satisfying

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

- unknown $f^* \in \mathcal{F}$ where \mathcal{F} is certain function class
- i.i.d. Gaussian noise $\varepsilon_1, \dots, \varepsilon_n \sim \mathcal{N}(0, \sigma^2)$
- ullet fixed design $(x_1,\ldots,x_n$ are fixed) or random design $(x_1,\ldots,x_n\stackrel{\mathsf{i.i.d.}}{\sim}
 ho)$

Goal: estimate f^* using the data

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

$$\begin{split} \|f - f^\star\|_n^2 &\coloneqq \frac{1}{n} \sum_{i=1}^n \left(f(\boldsymbol{x}_i) - f^\star(\boldsymbol{x}_i) \right)^2 \qquad \text{(for fixed design)} \\ \|f - f^\star\|_\rho^2 &\coloneqq \mathbb{E}_{\boldsymbol{x} \sim \rho} \big[(f(\boldsymbol{x}) - f^\star(\boldsymbol{x}))^2 \big] \qquad \text{(for random design)} \end{split}$$

Nonparametric least squares

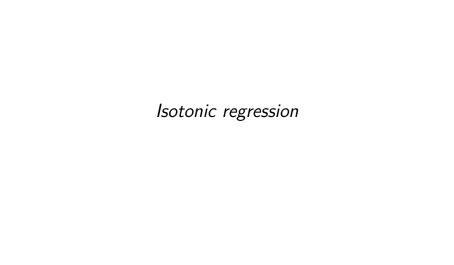
Least squares estimate:

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

- ullet this estimator depends on ${\cal F}$
- 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: $\mathcal{F} = \{\text{monotone function in } \mathbb{R}\}$
- convex regression: $\mathcal{F} = \{\text{convex function in } \mathbb{R}^d\}$
- ullet kernel ridge regression: $\mathcal{F}=$ reproducing kernel hilbert space (RKHS)



Isotonic regression: setup

- ullet 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

$$(\widehat{f}_1,\ldots,\widehat{f}_n) := \underset{f_1 \leq \cdots \leq f_n}{\operatorname{arg\,min}} \sum_{i=1}^n (y_i - f_i)^2,$$

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

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

- Key observation: $f^*(x)$ is only identifible for $x \in \{x_1, \dots, x_n\}$
- 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

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

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

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



Convex regression: setup

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

$$\begin{array}{ll} \underset{f_1,\ldots,f_n\in\mathbb{R},g_1,\ldots,g_n\in\mathbb{R}^d}{\text{minimize}} & \sum_{i=1}^n (y_i-f_i)^2 \\ \text{subject to} & f_j\geq f_i+\boldsymbol{g}_i^\top(\boldsymbol{x}_j-\boldsymbol{x}_i) & \text{for all } 1\leq i,j\leq n \end{array}$$

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

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

- ullet Key observation: $f^\star(x)$ is only identifible for $x\in\{x_1,\ldots,x_n\}$
- 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}

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

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 $n^{-4/(d+4)}$