Nonparametrics.

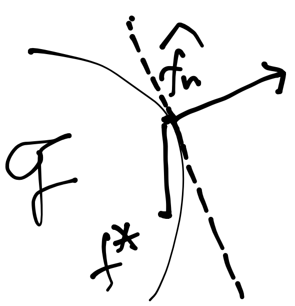
$$\begin{cases} Y_i = f^*(X_i) + \varepsilon_i & (i=1,2,\dots,n) \quad \varepsilon_i \overset{iid}{\sim} N(0,1) \\ X_1,\dots,X_n \overset{iid}{\sim} \mathbb{P} & \frac{d\mathbb{P}}{d\lambda}(x) = p(x) \end{cases}$$

x_i deterministics, $\varepsilon_i \overset{iid}{\sim} N(0,1)$.

MLE: $$\hat{f}_n := \operatorname*{argmin}_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^n (Y_i - f(x_i))^2 \right\}$$

— Can be solved using cvx opt.

Ideal R.O.C



$\forall \beta \in (0,1)$.

$$\frac{1}{n}\sum_{i=1}^n (Y_i - \hat{f}_n(x_i))^2$$

$$\le \frac{1}{n}\sum_{1}^n (Y_i - \beta f^*(x_i) - (1-\beta)\hat{f}_n(x_i))^2$$

$\beta \to 0$:

$$\boxed{\frac{1}{n}\sum_{i=1}^n (Y_i - \hat{f}_n(x_i))(f^*(x_i) - \hat{f}_n(x_i)) \le 0}$$

$$Y_i = f^*(x_i) + \varepsilon_i$$

$$\frac{1}{n}\sum_{i=1}^n (\hat{f}_n(x_i) - f^*(x_i))^2 \le \frac{1}{n}\sum_{i=1}^n \varepsilon_i (\hat{f}_n(x_i) - f^*(x_i)).$$

h: $\|h\|_n^2 := \frac{1}{n}\sum_{i=1}^n h(x_i)^2$. $\qquad \hat{\Delta}_n := \hat{f}_n - f^*$.

$$\|\hat{\Delta}_n\|_n^2 \leq \frac{1}{n}\sum_{i=1}^n \varepsilon_i \hat{\Delta}_n(x_i)$$

$$\leq \sup_{\substack{\|h\|_n \leq \|\hat{\Delta}_n\|_n \\ h \in \mathcal{F}^*}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i h(x_i)$$

$$\mathcal{F}^* := \{f - f^* : f \in \mathcal{F}\}.$$

Idea: $G_n(r) := \mathbb{E}\left[\sup_{\substack{\|h\|_n \leq r \\ h \in \mathcal{F}^*}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i h(x_i)\right]$ "Gaussian complexity"

Let δ_n solve $\delta_n^2 = G_n(\delta_n)$.

Thm, $\forall\, \varepsilon > 0$, $\exists\, C_\varepsilon > 0$, $\|\hat{f}_n - f^*\|_n \leq C_\varepsilon \cdot \delta_n$ w.p. $1-\varepsilon$

($i.e.$ $\|\hat{f}_n - f^*\|_n \lesssim \delta_n$ w.h.p.)

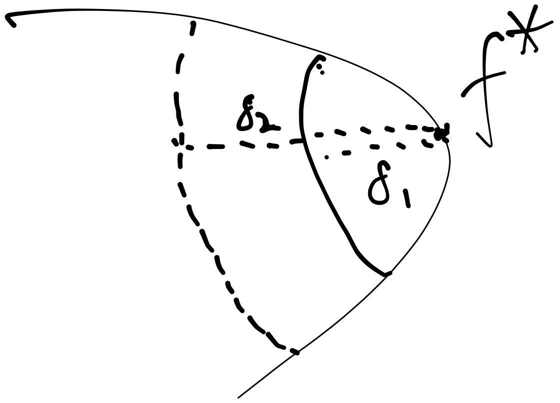
Proof sketch.

$$\mathbb{P}\left(\|\hat{f}_n - f^*\|_n \geq 2^M \delta_n\right) = \sum_{j \geq M+1} \mathbb{P}\left(2^{j-1}\delta_n \leq \|\hat{f}_n - f^*\|_n \leq 2^j \delta_n\right).$$

$$j\text{-th term} \leq \mathbb{P}\left(\sup_{h \in \mathcal{F}^* \cap \mathbb{B}_n(2^j \delta_n)} \frac{1}{n}\sum_{i=1}^n \varepsilon_i h(x_i) \geq 2^{2j-2}\delta_n^2\right).$$

$$\leq \frac{1}{2^{2j-2}\delta_n^2} \cdot G_n(2^j \delta_n) \leq \frac{1}{2^{2j-2}\delta_n^2} \cdot 2^j \cdot G_n(\delta_n).$$

Fact. $\frac{G_n(\delta)}{\delta}$ monotone-decreasing.



$$\frac{G_n(\delta)}{\delta} = \mathbb{E}\left[\sup_{\substack{\|h\|_n \le \delta \\ h \in \mathcal{F}^*}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \frac{h(x_i)}{\delta}\right].$$

$$= \mathbb{E}\left[\sup_{\substack{\|h\|_n \le 1 \\ h \in \mathcal{F}^*/\delta}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i h(x_i)\right].$$

Thm: $$\mathbb{E}\left[\sup_{h \in \mathcal{H}} \left|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i h(x_i)\right|\right]$$

$$\le \frac{C}{\sqrt{n}} \int_0^{\mathrm{diam}(\mathcal{H})} \sqrt{\log N(\delta; \mathcal{H}, \|\cdot\|_n)}\, d\delta.$$

Improved version $\forall \delta_0 > 0$

$$\mathbb{E}\left[\sup_{h \in \mathcal{H}} \left|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i h(x_i)\right|\right] \le c\delta_0 + \frac{c}{\sqrt{n}} \int_{\delta_0}^{\mathrm{diam}(\mathcal{H})} \sqrt{\log N(\delta; \mathcal{H}, \|\cdot\|_n)}\, d\delta$$

$$\left(\mathrm{diam}(\mathcal{H}) := \sup_{h \in \mathcal{H}} \|h\|_n\right).$$

$$\Sigma(\beta) := \left\{ f : [0,1] \to \mathbb{R} : \begin{array}{l} |f^{(k)}(x)| \le 1 \quad \forall x \in [0,1],\ k \le \beta \\ |f^{(l)}(x) - f^{(l)}(y)| \le |x-y|^{\beta - l} \\ \text{where } l = \lfloor \beta \rfloor \end{array} \right\}$$

$$\boxed{f^{(k)} = \partial^k f = \frac{d^k}{dx^k} f}$$

$\beta = 1$: $\quad |f(x)| \le 1, \ |f'(x)| \le 1. \quad (\mathcal{F} = \Sigma(\beta))$.

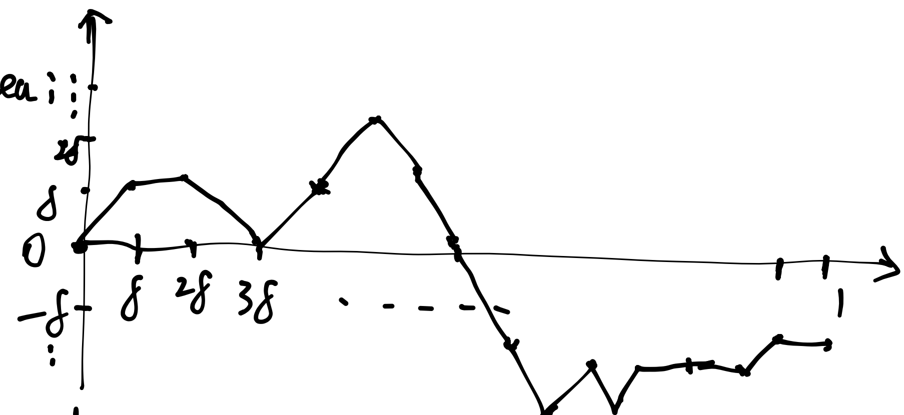
$$N(\delta; \mathcal{F}^* \cap \mathbb{B}_n(r), \|\cdot\|_n) \le N(\delta; \mathcal{F}, \|\cdot\|_\infty).$$

Claim: $\log N(\delta; \Sigma(1), \|\cdot\|_\infty) \le \frac{c}{\delta}$.

[Recall parametric class $\quad N(\delta \cdots) \le \left(\frac{c}{\delta}\right)^d$

$\log N(\text{---}) \le d \cdot \log(1/\delta)$]

Proof idea:



(*)

$f \in \Sigma(1)$, consider

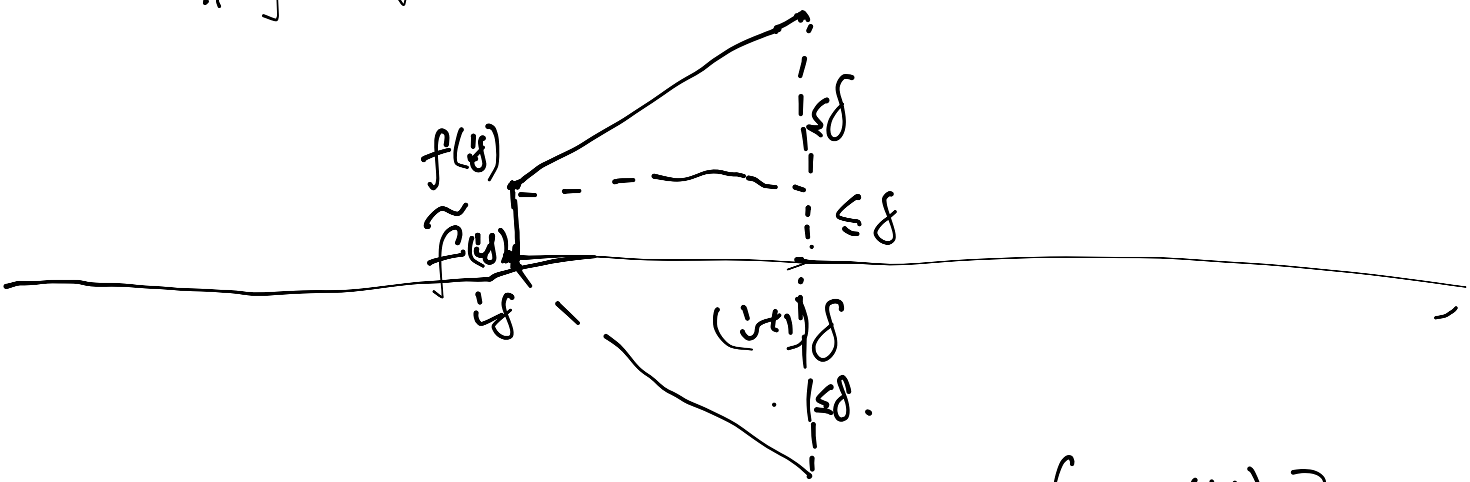
$$\left(\left\lfloor \frac{f(0)}{\delta} \right\rfloor, \left\lfloor \frac{f(\delta)}{\delta} \right\rfloor, \left\lfloor \frac{f(2\delta)}{\delta} \right\rfloor, \cdots \left\lfloor \frac{f(1)}{\delta} \right\rfloor \right)$$

$$\left| \left\lfloor \frac{f((i+1)\delta)}{\delta} \right\rfloor - \left\lfloor \frac{f(i\delta)}{\delta} \right\rfloor \right| \le 1.$$

For any $f \in \Sigma(1)$, $\exists \tilde{f}$ of shape (*)

s.t. $\quad \delta \cdot \left\lfloor \frac{f(i\delta)}{\delta} \right\rfloor = \tilde{f}(i\delta) \quad \forall i.$

$$\|f - \tilde{f}\|_\infty \leq 3\delta.$$



How many functions are of the form $(*)$?

$f(0)$ has $\left\lceil \frac{2}{\delta} \right\rceil$ choices.

Given $f(0), f(\delta), \cdots, f(i\delta)$,

$f((i+1)\delta)$ has 3 choices.

$$\# \text{ choices} \leq \left\lceil \frac{2}{\delta} \right\rceil \cdot 3^{\lceil 1/\delta \rceil}.$$

$$\log N(3\delta; \mathcal{E}(1), \|\cdot\|_\infty) \leq \log(2/\delta) + \frac{1}{\delta} \cdot \log 3$$

$$\leq \frac{4}{\delta}.$$

Corollary: $$\mathbb{E}\left[\sup_{\substack{h \in \mathcal{E}(1) - f^* \\ \|h\|_n \leq r}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i h(x_i)\right] \leq c \cdot \sqrt{\frac{r}{n}}.$$

$$\mathrm{diam}\left((\Sigma(1) - f^*) \cap \mathbb{B}_n(r) \right) \lesssim r.$$

$$\mathbb{E}[\cdots] \lesssim \frac{c}{\sqrt{n}} \int_0^r \sqrt{\log N(\delta; \Sigma(1), \|\cdot\|_\infty)}\, d\delta$$

$$\lesssim \frac{1}{\sqrt{n}} \int_0^r c \cdot \delta^{-1/2}\, d\delta$$

$$\lesssim \sqrt{\frac{r}{n}}.$$

$$r_n^2 = \mathcal{G}_n(r_n) \iff r_n \lesssim n^{-1/3}.$$

Corollary: w.h.p. $\|\hat{f}_n - f^*\|_n \leq c \cdot n^{-1/3}$.

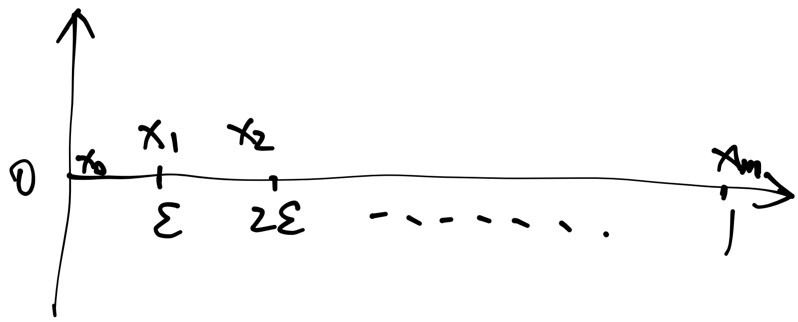
$$\inf_{\hat{f}_n} \sup_{f^* \in \Sigma(1)} \mathbb{E}\left[\|\hat{f}_n - f^*\|^2\right] \gtrsim n^{-1/3}.$$

General Hölder class.

Thm: $\log N(\delta; \Sigma(\beta), \|\cdot\|_\infty) \lesssim \left(\frac{c}{\delta}\right)^{1/\beta}$.

(Indeed, $\log N(\delta; \Sigma_d(\beta), \|\cdot\|_\infty) \lesssim \left(\frac{c}{\delta}\right)^{d/\beta}$).

Proof :



$\varepsilon = \delta^{1/\beta}$

$m \leq 1 + \lceil 1/\varepsilon \rceil.$

$$A : \quad f \longmapsto \left(\left\lfloor \frac{\partial^k f(x_j)}{\varepsilon^{\beta-k}} \right\rfloor \right)_{\substack{0 \leq k \leq \lfloor \beta \rfloor \\ j = 0, 1, \ldots, m}}$$

$$\boxed{\text{Claim : } \quad A(f) = A(g) \implies \| f - g \|_\infty \leq \delta \cdot (\text{c}) \quad (*)}$$

Goal : $\left| \{ Af : f \in \Sigma(\beta) \} \right| \leq ?$

$$|\partial^k f| \leq 1 \implies \left\lfloor \frac{\partial^k f(x_j)}{\varepsilon^{\beta-k}} \right\rfloor \text{ has } \frac{2}{\varepsilon^{\beta-k}} \text{ choices.}$$

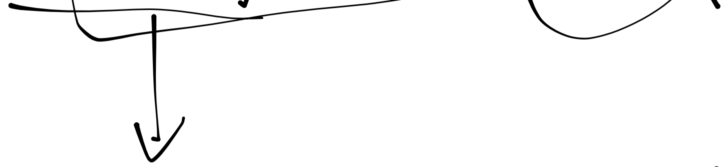
Coarse estimate :

$$\log |\{Af \cdots\}| \leq \log \left[\left(\frac{2}{\varepsilon^\beta} \right)^{(m+1)\cdot(\beta+1)} \right]$$

$$\asymp m \cdot \beta^2 \cdot \log(1/\varepsilon) \asymp \frac{1}{\delta^{1/\beta}} \log(1/\delta).$$

$l = \lfloor \beta \rfloor$

$$\begin{bmatrix} \left\lfloor \frac{f(x_0)}{\varepsilon^\beta} \right\rfloor & \left\lfloor \frac{f(x_1)}{\varepsilon^\beta} \right\rfloor & \cdots & \left\lfloor \frac{f(x_m)}{\varepsilon^\beta} \right\rfloor \\ \left\lfloor \frac{\partial f(x_0)}{\varepsilon^{\beta-1}} \right\rfloor & \vdots & \cdots & \vdots \\ \vdots & & & \\ \left\lfloor \frac{\partial^l f(x_0)}{\varepsilon^{\beta-l}} \right\rfloor & \left\lfloor \frac{\partial^l f(x_1)}{\varepsilon^{\beta-l}} \right\rfloor & \cdots & \left\lfloor \frac{\partial^l f(x_m)}{\varepsilon^{\beta-l}} \right\rfloor \end{bmatrix}$$



First column: $\leq \left(\frac{1}{\varepsilon}\right)^{\beta\cdot(\ell+1)}$ choices.

Given previous columns.

$$\partial^k f(x_{i+1}) = \partial^k f(x_i) + \varepsilon\cdot\partial^{k+1} f(x_i) + \frac{\varepsilon^2}{2}\partial^{k+2} f(x_i) + \cdots + \frac{\varepsilon^{\ell-k}}{(\ell-k)!}\partial^\ell f(x_i+\tau(x_{i+1}-x_i))$$

Hölder: $\left|\partial^\ell f(x_i+\tau(x_{i+1}-x_i)) - \partial^\ell f(x_i)\right| \leq \varepsilon^{\beta-\ell}$.

$\ell = \lfloor \beta \rfloor$

$\forall t$, $\partial^{k+t} f(x_i)$ is determined by the matrix up to error $\varepsilon^{\beta-k-t}$

$\Longrightarrow$ $\partial^k f(x_{i+1})$ is determined by previous columns up to error. $C\cdot\varepsilon^{\beta-k}$

$\Longrightarrow$ Based on previous columns, each entry in $(i+1)$-th column has at most $(2C+1)$ choices.

\# choices $\leq (\varepsilon^{-\beta})^{\ell}\cdot(2C+1)^{\beta\cdot m}$

$$\leq \delta^{-\beta} \cdot \exp(c'/\varepsilon)$$

(c'', c' const depending on β).

$$\leq \exp\left(c''/\delta^{1/\beta}\right).$$

Proof of (*).

$$k=0 \implies |f(x_i)-g(x_i)| \leq \varepsilon^{\beta} = \delta$$

on grid pts.

$$|f(x)-g(x)| \leq |f(x_i)-g(x_i)| + |(x-x_i)\partial(f-g)(x_i)|$$
$$+ \left|\frac{(x-x_i)^2}{2!}\partial^2(f-g)(x_i)\right| + \cdots + \left|\frac{(x-x_i)^l}{l!}\partial^l(f-g)(x_i)\right|$$
$$+ |R(x)|$$

$$|R(x)| \leq |x-x_i|^{\beta} \cdot \text{(Hölder)}$$

$$|f(x)-g(x)| \leq \varepsilon^{\beta} + \varepsilon \cdot \varepsilon^{\beta-1} + \frac{\varepsilon^2}{2} \cdot \varepsilon^{\beta-2}$$
$$+ \cdots + \frac{\varepsilon^l \cdot \varepsilon^{\beta-l}}{l!} + \varepsilon^{\beta}$$
$$\leq (e+1) \cdot \varepsilon^{\beta} = (e+1)\delta$$