Denoising

Consider an extremely simple model

$$\mathcal{F}_{k,e} := \{ f : [0,1] \to \mathbb{C} : \exists\, t_0 \leq 0 < t_1 < \ldots < t_k = 1$$

$$\text{st } \forall i \in \{0, \ldots k-1\}\ f|_{(t_i, t_{i+1})} \text{ is}$$

$$\text{a degree } e \text{ polynomial} \}$$

Consider ~~f~~ observing

$$dY_t = f(t)\, dt + \frac{\sigma}{\sqrt{n}}\, dB_t$$

(equivalently $y_i = f(t_i) + \sigma z_i \quad t_i = i/n$ &
for large n. $0 \leq i \leq n$)

- at least k vanish. moment
- compact support

Wavelet decomposition

$$f(t) = \sum_{q=-M}^{M} c_q \phi(t-q) +$$

$$+ \sum_{\ell=0}^{\infty} \sum_{q=-M}^{2^\ell+M} \theta_{\ell,q} \psi_{\ell,q}(t)$$

$$\cancel{Y_{\ell,k} = \theta_{\ell,k} + \frac{\sigma}{\sqrt{n}} g_{\ell,k}}$$

$$Y_{\ell,i} = \theta_{\ell,i} + \frac{\sigma}{\sqrt{n}} g_{\ell,i}$$

- Focus on a specific level $\ell < 0$
- Drop the index ℓ, rename i

$$Y_i = \theta_i + \frac{\sigma}{\sqrt{n}} g_i \qquad 1 \le i \le 2^\ell + 2M =$$

note that (by the wavelet properties)

θ has at most $C \cdot k$ nonzero entries

So we will consider

$$\Theta_{N,k} := \{ \theta \in \mathbb{R}^N : \|\theta\|_0 \le k \}$$

nonzero entries. not a norm

$$y = \theta + \sigma g \qquad g \sim N(0, I_N)$$

(will set $\sigma \to \sigma/\sqrt{n}$ to make contact.

$$\overline{R}_\ell(N,k;\sigma) := \inf_{M} \sup_{\theta \in \Theta_{N,k}} \mathbb{E}\{ \|\theta - My\|_{\ell_2}^2 \}$$

$$\overline{R}_{nl}(N,k;\sigma) := \inf_{\hat{\theta}} \sup_{\theta \in \Theta_{N,k}} \mathbb{E}\{ \|\theta - \hat{\theta}(y)\|_{\ell^2}^2 \}$$

By invariance under scaling

$$\overline{R}_{\ell/nl}(N,k;\sigma) = \sigma^2 R_{\ell/nl}(N,k)$$

will focus on this

consider the linear risk

for a fixed M, θ

$$\sup_{\theta \in \Theta_{N,K}} \mathbb{E}\left(\|\theta - My\|^2\right) = \sup_{\theta \in \Theta_{N,K}} \mathbb{E}\|\theta - M\theta - Mg\|^2 =$$

$$= \sup_{\theta \in \Theta_{N,K}} \|(M-I)\theta\|^2 + \|M\|_F^2$$

If $M \neq I$ $k \geq 2$, $\exists\, \theta_0 \in \Theta_{N,K}$ st $(M-I)\theta_0 \neq 0$

by taking $\theta = c\theta_0$ $c \to \infty$

we obtain $\sup(\ldots) = \infty$

Hence the only finite value is $M = I$.

$\dot{R}_\ell(N,k) = N$ (no denoising!)

For R_N, we will study a specific denoiser

$$\hat{\theta}(y)_i = \eta(y_i; \lambda)$$

soft thresholding fct.

$$\eta(y;\lambda) = \begin{cases} y - \lambda & \text{if } y \geq \lambda \\ 0 & \text{if } |y| < \lambda \\ y + \lambda & \text{if } y \leq -\lambda \end{cases}$$

$$= \operatorname{argmin} \left\{ \frac{1}{2}(x-y)^2 + \lambda |x| \right\}$$

$$\mathbb{E}\left(\|\hat{\theta}(y) - \theta\|^2\right) = \cancel{\sum_{i=1}^{N} \mathbb{E}\{(\eta(\theta_i + g}$$

$$= \sum_{i \in S} \quad = \sum_{i=1}^{N} r_\lambda(\theta_i)$$

$$r_\lambda(\theta) = \mathbb{E}\left\{(\eta(\theta + G_i; \lambda) - \theta)^2\right\}$$

$$r_\lambda(0) = 2\,\mathbb{E}\left[(G_1-\lambda)^2 \mathbb{1}_{G\geq\lambda}\right]$$

$$= 2\int_\lambda^\infty (x-\lambda)^2 e^{-x^2/2}\,\frac{dx}{\sqrt{2\pi}} =$$

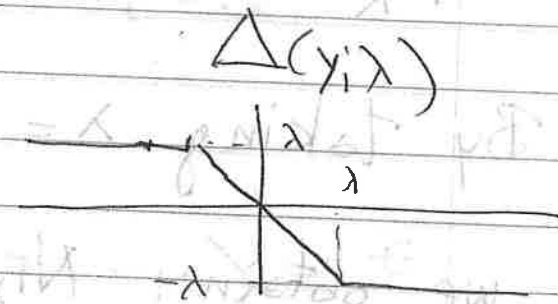
$$= \cancel{2}\sqrt{\frac{2}{\pi}}\, e^{-\frac{\lambda^2}{2}}\int_0^\infty z^2 \exp\left(-\lambda z-\frac{z^2}{2}\right)dz$$

$$\neq = (1+o_\lambda(1))\sqrt{\frac{2}{\pi}}\, e^{-\frac{\lambda^2}{2}}\int_0^\infty z^2 e^{-\lambda z}\,dz$$

$$= (1+o_\lambda(1))\sqrt{\frac{8}{\pi}}\,\frac{1}{\lambda^3}\, e^{-\lambda^2/2}$$

Further $\cancel{y(y)}$

$$\eta(y;\lambda) = y + \Delta(y;\lambda)$$



$$r_\lambda(\theta) = \mathbb{E}\{(G_1+\Delta(\theta+G_1;\lambda))^2\} =$$

$$= 1 + 2\,\mathbb{E}(G_1\Delta(\theta+G_1;\lambda)) + \mathbb{E}\Delta^2(\ldots)$$

$$= 1 + 2\,\mathbb{E}\Delta'(\theta+G_1;\lambda) + \mathbb{E}\Delta^2(\ldots)$$

$$r_\lambda(\theta) \leq 1+\lambda^2 = \lim_{\theta \to \pm\infty} r_\lambda(\theta)$$

$$R_{nl}(N,k) \leq \inf_\lambda \left[(1+\lambda^2)\cdot k + (N-k)\, r_\lambda(0) \right]$$

want to take $\lambda \to$ large

$$\lambda^2 k \approx c N \frac{e^{-\lambda^2/2}}{\lambda^3}$$

$$\frac{N}{k} \approx e^{\lambda^2/2} \qquad \lambda \approx \sqrt{2 \log \frac{N}{k}}$$

By taking $\lambda = \sqrt{2 \log \frac{N}{k}}$

we obtain $N r_\lambda(0) \ll k$

$$R_{nl}(N,k) \leq 2k \log \frac{N}{k} \left(1+o(1)\right)$$

much less than the linear risk

Reinstating the dependence on σ^2, n

$$R_{ne}(N,k;\sigma,n) \le \frac{2k\sigma^2}{n}\log\frac{N}{k}\,(1+o(1))$$

The oracle risk (if we knew the k nonzeros) is $\frac{k\sigma^2}{n}$.

Can we prove a lower bound?

Take $(\theta_i)_{i\le N}$ iid $\theta_i \sim Q$

$$Q(\theta_i = 0) = 1-\epsilon$$

$$Q(\theta_i = +\mu) = Q(\theta_i = -\mu) = \frac{\epsilon}{2}$$

$\epsilon = \frac{ck}{N}$ for some $\cancel{c \ll 1}$ $c\in(0,1)$

$$\cancel{R_B(N, Q, \sigma) = \mathbb{E}\inf_{\hat\theta:\mathbb{R}\to\mathbb{R}} L}$$

$$\cancel{R_B(N; Q,\sigma) = N \cdot \inf_{\hat{\theta}:\mathbb{R}\to\mathbb{R}} \mathbb{E}\{(\theta-\hat{\theta}(y))^2\}}$$

~~wrt~~

For any prior Q, define Bayes error as

$$r_B(Q;\sigma) = \inf_{\hat{\theta}:\mathbb{R}\to\mathbb{R}} \mathbb{E}\{(\theta-\hat{\theta}(Y))^2\}$$

Here expectation is wrt $\theta \sim Q$
$Y = \theta + \sigma G$

~~Then note~~

~~Q⊗~~ Denote by $\cancel{Q^{\otimes}_{N,k}} =$

$$Q_{N,k} := \frac{Q^{\otimes N}\, 1_{\|\theta\|_0 \le k}}{\int Q^{\otimes N}(d\theta)\, 1_{\|\theta\|_0 \le k}}$$

$$R_{nl}(N; k\sigma) \geq \inf_{\hat{\theta}: \mathbb{R}^N \to \mathbb{R}^N} \int \mathbb{E}\left(\|\hat{\theta}(Y) - \theta\|_2^2\right) Q_{N,k}(d\theta)$$

$$\geq N\, r_B(Q; \sigma) + \cdots$$

For $\epsilon = c\frac{k}{N}$.

$$r_B(Q; \sigma) = \mathbb{E}\{(\theta - \mathbb{E}(\theta|Y))^2\}$$

Now take $Q = (1-\epsilon)\delta_0 + \frac{\epsilon}{2}\cancel{(\delta_\mu + \delta_{-\mu})}\ \delta_\mu$

$$\sup_\mu r_B(Q_\mu; \sigma) = \sigma^2 \sup_\mu \underset{\substack{\| \\ r_B(Q_\mu; 1)}}{r_B(\mu)}$$

$$R_{nl}(N; k, \sigma) \gtrsim N\sigma^2 \sup_{\mu \geq 0} r_B(\mu)$$

We next proceed to ~~guess~~ lower bound $\Gamma_B(\mu)$.

Suppose $\hat{\theta}_*: \mathbb{R} \to \{0, \mu\}$ be defined by

$$\hat{\theta}_*(y) = \begin{cases} 0 & \text{if } \hat{\theta}(y) < \mu c \\ \mu & \text{if } \hat{\theta}(y) > \mu c \end{cases}$$

$$\cancel{\Gamma_B(\mu) \geq \ldots}$$

$$\Gamma_B(\mu) = (1-\epsilon)\mathbb{E}_0 \hat{\theta}^2(Y) + \epsilon \mathbb{E}_\mu\left[(\hat{\theta}(Y)-\mu)^2\right]$$

$$\geq (1-\epsilon)\mu^2 c^2 \mathbb{P}_0(\hat{\theta}(Y) > \mu c)$$

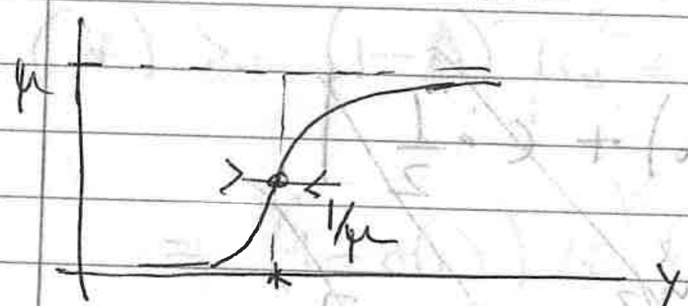
$$+ \epsilon \mu^2 (1-c)^2 \mathbb{P}_\mu(\hat{\theta}(Y) < \mu c)$$

A simple calculation yields

$$\hat{\theta}(y) = \frac{\mu\epsilon\, e^{\mu y - \mu^2/2}}{1-\epsilon+\epsilon e^{\mu y-\mu^2/2}}$$

Hence if $y = \frac{1}{2}\mu^2 + \frac{1}{\mu}\log\left[z\left(\frac{1-\epsilon}{\epsilon}\right)\right]$

we get $\hat{\theta}(y) = \frac{z\mu}{1+z}$



$\bar{y} = \frac{1}{2}\mu^2 + \frac{1}{\mu}\log\frac{1-\epsilon}{\epsilon}$

$$r_B(y) = (1-\epsilon)\,\mathbb{E}_0\hat{\theta}(Y)^2 + \epsilon\,\mathbb{E}_\mu\left[(\hat{\theta}(Y)-\mu)^2\right]$$

$\cancel{\geq (1-\epsilon) \geq \mu(1-\Delta)}$ $(1-\epsilon)$

$$\geq \mu^2(1-\Delta)\left[\mathbb{P}_0\left(Y>\bar{y}+\frac{c}{\mu}\right) + \epsilon\,\mathbb{P}_\mu\left(Y<\bar{y}-\frac{c}{\mu}\right)\right]$$

$$\mu_0^2 = 2\log\frac{1-\epsilon}{\epsilon}$$

$$\bar{y} = \frac{\mu}{2} + \frac{\mu_0^2}{2\mu}$$

$$\cancel{\mu^2 \, 2}$$

$$\geq (1-\Delta)\mu^2\left[\Phi(-\bar{y}) + \epsilon\,\Phi(\bar{y}-\mu)\right]$$

$$\geq (1-\Delta)\mu^2\left[\Phi\left(-\frac{\mu}{2}-\frac{\mu_0^2}{2\mu}\right) + \epsilon\,\Phi\left(-\frac{\mu}{2}+\frac{\mu_0^2}{2\mu}\right)\right]$$

Choose for c a const

$$\cancel{\geq c(1-\Delta)\mu^2\Phi\left(-\frac{\mu}{2\tau}\right)} \quad \cancel{\mu = \mu_0} \quad \mu^2 + \mu c = \mu_0^2$$

$$\cancel{\geq (1-\Delta)\mu_0^2\left[\Phi(-\mu_0) + \epsilon\cdot\frac{1}{2}\right]}$$

$$\cancel{= (1-\Delta)\,\mu_0^2\left[\frac{e^{-\mu_0^2/2}}{\mu_0\sqrt{2\pi}} + \frac{\epsilon}{2}\right]}$$

$$\frac{\mu_0^2}{\mu} = \mu + c \qquad -\frac{\mu}{2} + \frac{\mu_0^2}{2\mu} = \frac{c}{2}$$

$$-\frac{\mu}{2} - \frac{\mu_0^2}{2\mu} = -\mu - \frac{c}{2} = -\sqrt{\mu_0^2 + \left(\frac{c}{2}\right)^2}$$

$$= -\mu_0 \left(1 + \left(\frac{c}{2\mu_0}\right)^2\right)^{1/2} = -\mu_0 + O\left(\frac{1}{\mu_0}\right)$$

$$\Phi\left(-\frac{\mu}{2} + \frac{\mu_0^2}{2\mu}\right) = \Phi\left(\frac{c}{2}\right)$$

$$\Phi\left(-\frac{\mu}{2} - \frac{\mu_0^2}{2\mu}\right) \leq \bar{C}\,\Phi(-\mu_0)$$

$$\leq \frac{\bar{C}}{\mu_0} e^{-\mu_0^2/2} \ll \epsilon$$

$$(*) \geq (1-\Delta)\,\mu_0^2\,\epsilon\,\Phi\left(\frac{c}{2}\right)$$

$$= (1+o(1))\, 2\epsilon \log\frac{1}{\epsilon} + \ldots$$

Generalization

$$\Theta_{N,p}(\rho) := \{\theta \in \mathbb{R}^N : \|\theta\|_p \leq \rho\}$$

p pseudonorm

Theorem [DJ94] Assume $\sigma^2 \log \frac{N}{(\rho/\sigma)^p} \to 0$

$N/(\rho/\sigma)^p \to \infty$. $p \in (0,2)$

$$R_{nl}(N; \rho\,\sigma) = \left[2 \log \frac{N}{(\rho/\sigma)^p}\right]^{1-\frac{p}{2}} \sigma^{2-p} \rho^p \cdot (1+o(1))$$

And is achieved by soft thresholding at level

$$\cancel{\lambda = \sqrt{2 \log N/(\rho/\sigma)^p}}\ (1+o(1))$$

$$\lambda = \sigma \sqrt{2 \log N/(\rho/\sigma)^p}$$

Heuristic $(\sigma = 1)$

$$R_{st} \lesssim \rho^p \lambda^{2-p} + N e^{-\lambda^2/2}$$

soft thresh

$$\lambda_* \doteq \sqrt{2 \log \frac{N}{\rho^2}}$$