(p1)

$y_i = x_i^T \beta_0 + \varepsilon_i \qquad i = 1, \dots, n$

$n \times p \times 1$

$$\hat\beta = \arg\min \frac{1}{n}\sum_{i=1}^{n} \rho\left(y_i - \langle x_i, \beta\rangle\right) + f(\beta)$$

loss

---

lasso: $\frac{1}{2n}\sum_{i=1}^{n} \|(\quad)\|^2 + \frac{\lambda}{n}\|\beta\|_1$

classical setting: (sparsity).

$\lambda \geq \sigma\sqrt{C_0 \log p}$

$\Rightarrow \|\hat\beta^{lasso} - \beta_0\|^2 \leq C \cdot \|\beta_0\|_0 \lambda^2$

pros:

① near rate-minimax (Information)

② crude bound solid under a wide ranged design X (eg. RE condition)

cons: hard to extend to data-driven λ

proportional H-dim asy.

wisely choose η. looks like de-biased lasso

F.O.C. $\hat\beta(\lambda) = \text{prox}_{\eta\lambda\|\cdot\|_1}\left(\hat\beta(\lambda) + \frac{\eta}{\lambda}X^T(y - X\hat\beta(\lambda))\right)$

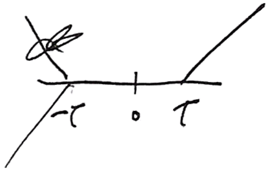
where $\text{prox}_f(x) = \arg\min_z \left\{ \frac{1}{2}|x - z|^2 + f(z) \right\}$

under ① standard gaussian design

② $\frac{p}{n}$ = const

soft-threshold

$\eta_1(x, \tau) = \text{prox}_{\tau|\cdot|_1}(x) = (|x| - \tau)\,\text{sgn}(x)$



$\bar\beta_0 = \frac{1}{p}\sum_{j=1}^{p} \delta_{\beta_{0,j}}$ $\qquad \delta \triangleq \frac{n}{p}$ "aspect ratio"

empirical dist. of β_0 $\quad$ eg $\beta_0 = (0,0,1)$ $\quad \bar\beta_0 = \frac{2}{3}\delta_0 + \frac{1}{3}\delta_1$

$X_{ij} \overset{iid}{\sim} N(0, \frac{1}{n}) \qquad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

"effective noise" σ_*^2 $\qquad \sigma_*, \lambda_*$ unknown

"effective regularization" λ_*

lasso: 1-dim 问题

$$\sigma_*^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\left(\eta_1(\bar\beta_0 + \sigma_* Z; \lambda_*) - \bar\beta_0\right)$$

$$\frac{\lambda}{\lambda_*} = 1 - \frac{1}{\delta}P\left(|\bar\beta_0 + \sigma_* Z| > \lambda_*\right)$$

$\mathbb{E}$ take on $\bar\beta_0 \otimes Z$

where $Z \sim N(0,1)$

lasso sparsity $\approx$ sequence sparsity

$\mathbb{E}\eta_1'(\bar\beta_0 + \sigma_* Z; \lambda_*)$

(β2) de-biased lasso

choose $\eta = \frac{\lambda_*}{\lambda}$ (Miolane & Montanari 2021)

uniformly $\lambda \in [\frac{1}{K}, K]$

(*1) $\hat\beta^*(\lambda) = \hat\beta(\lambda) + \frac{\lambda_*}{\lambda} X^T(y - X\hat\beta(\lambda)) \overset{d}{\simeq} N(\beta_0, \sigma_*^2)$

$\hat\beta(\lambda) \overset{d}{\simeq} \mathrm{prox}_{\lambda_*\|\cdot\|_1}(\beta_0 + \sigma_* Z_p)$

Apply lasso with λ_* for $y = \beta_0 + \sigma_* Z_p$

$\hat r(\lambda) = y - X\hat\beta(\lambda)$

$\hat r(\lambda) \overset{d}{\simeq} \frac{\lambda}{\lambda_*}\left(\varepsilon + \sqrt{\sigma_*^2 - \sigma^2}\, Z_n\right)$ $\quad \varepsilon \perp\!\!\!\perp Z_n$

precisely, for any Lipschitz $g: \mathbb{R}^p \to \mathbb{R}$

$$\sup_{\lambda \in [\frac{1}{K}, K]} \left| g\left(\frac{\hat\beta(\lambda)}{\sqrt{p}}\right) - \mathbb{E}\, g\left(\mathrm{prox}_{\lambda_*\|\cdot\|_1}(\beta_0 + \sigma_* Z_p)/\sqrt{p}\right) \right| \le o(1)$$

Find data-driven estimator for $\frac{\lambda_*}{\lambda}$

$$\frac{\lambda}{\lambda_*} = 1 - \frac{1}{\delta}\underbrace{\mathbb{P}(|\beta_0 + \sigma_* Z| > \lambda_*)}_{\|}$$

$\mathbb{P}(\eta_1(\beta_0 + \sigma_* Z, \lambda_*) \neq 0)$ $\quad$ soft-threshold $\neq 0$

$\frac{1}{p}\sum_{j=1}^{p} \mathbb{E}_Z \mathbb{1}\left(\eta_1(\beta_0 + \sigma_* Z, \lambda_*) \neq 0\right)$

$\Leftrightarrow$ $g(x) = \frac{1}{p}\sum_{j=1}^{p} \mathbb{1}(x \neq 0)$

$\frac{1}{p}\sum_{j=1}^{p} \mathbb{1}(\hat\beta_j(\lambda) \neq 0)$ $\quad \frac{\|\hat\beta\|_0}{p}$

i.e. $\frac{\lambda}{\lambda_*} \approx 1 - \frac{1}{\delta}\frac{\|\hat\beta(\lambda)\|_0}{p} = 1 - \frac{\cancel{p}}{n}\frac{\|\hat\beta(\lambda)\|_0}{\cancel{p}}$

degree-of-freedom adjustment.

$$\hat\beta^d(\lambda) = \hat\beta(\lambda) + \frac{1}{1 - \frac{\|\hat\beta(\lambda)\|_0}{n}} X^T(y - X\hat\beta(\lambda))$$

debiased lasso

$\approx N(\beta_0, \sigma_*^2 I_p)$

Application 1: degree-of-freedom adjusted debiased Lasso

Classical debiased Lasso $\xrightarrow{\text{review}}$ $\hat\beta^d(\lambda) \doteq \hat\beta(\lambda) + X^T(y - X\hat\beta(\lambda))$ $\quad (\Sigma^{-1} = I)$

holds under "strong sparsity regime"

$s^2 = \|\beta_0\|_2^2 \ll \frac{n}{\mathrm{poly}\log p}$

classical regime $\sigma_*^2 \approx \sigma^2 \approx \frac{1}{n}\|y - X\hat\beta(\lambda)\|^2$

$$\hat\sigma_*(\lambda) = \frac{\|y - X\hat\beta(\lambda)\|^2/n}{(1 - \|\hat\beta\|_0/n)^2}$$

prop. $\sup_{\lambda \in [\frac{1}{K}, K]} |\hat\sigma_*^2(\lambda) - \sigma^2(\lambda)| \approx 0$ w.h.p.

P3

$$\hat{\sigma}_*^2(\lambda) = \frac{\|\hat{r}(\lambda)\|^2/n}{(\lambda/\lambda_*)^2} = \frac{(\lambda/\lambda_*)^2}{(\lambda/\lambda_*)^2}\frac{1}{n}\|\varepsilon + \sqrt{\sigma_*^2-\sigma^2}\, z_n\|^2$$

$$= \frac{1}{n}\|\varepsilon + \sqrt{\sigma_*^2-\sigma^2}\, z_n\|^2$$

$\sigma_* \perp\!\!\!\perp z_n$

$$\triangleq \frac{1}{n}\mathbb{E}\|\varepsilon + \sqrt{\sigma_*^2-\sigma^2}\, z_n\|^2$$

$$= \frac{1}{n}\cdot n(\sigma^2+\sigma_*^2-\sigma^2) = \sigma_*^2$$

CI: $[\hat{\beta}_j^d(\lambda) \pm \hat{\sigma}_*(\lambda) z_{\frac{\alpha}{2}}]$ Beller & Zhang

Application Adaptive tuning by SURE

(Stein's unbiased risk estimator)

$$\hat{p}(\lambda) \triangleq \frac{1}{n}\|y - X\hat{\beta}(\lambda)\|^2 + \frac{2\sigma^2}{n}\|\hat{\beta}(\lambda)\|_0$$

$\to$ 对不同λ可计算这个函数 (path)

$$\hat{\lambda} = \underset{\lambda\in[\epsilon,K]}{\operatorname{argmin}}\ \hat{p}(\lambda)$$

Tibshirani & Taylor (2012)

$$\mathbb{E}[\hat{p}(\lambda)\mid X] = \frac{1}{n}\mathbb{E}\|X(\hat{\beta}(\lambda)-\beta_0)\|^2 + \sigma^2$$

$$\sup_\lambda \left|\hat{p}(\lambda) - \frac{1}{n}\|X(\hat{\beta}(\lambda)-\beta_0)\|^2 + \sigma^2)\right| \approx 0 \qquad \sigma^2 = \mathbb{E}\varepsilon_1^2$$

proof:

$$\begin{cases} S_* = P(|\beta_0 + \sigma_* Z| > \lambda_*) \approx \dfrac{\|\hat{\beta}\|_0}{p} \\ \dfrac{\lambda}{\lambda_*} = 1 - \dfrac{S_*}{\delta} \end{cases}$$

$$= \frac{1}{n}\|\hat{r}(\lambda)\|^2 + \frac{2\sigma^2}{n}\|\hat{\beta}(\lambda)\|_0$$

$$= \frac{1}{n}\left(\frac{\lambda}{\lambda_*}\right)^2\|\varepsilon + \sqrt{\sigma_*^2-\sigma^2}\, z_n\|^2 + \frac{2\sigma^2}{\delta}S_*$$

$$= \left(\frac{\lambda}{\lambda_*}\right)^2\sigma_*^2 + 2\sigma^2\left(1-\frac{\lambda}{\lambda_*}\right) \qquad (1)$$

$$\frac{1}{n}\|X(\hat{\beta}(\lambda)-\beta_0)\|^2 = \frac{1}{n}\|\hat{r} - \varepsilon\|^2 = \frac{1}{n}\left\|\left(\frac{\lambda}{\lambda_*}-1\right)\varepsilon + \frac{\lambda}{\lambda_*}\sqrt{\sigma_*^2-\sigma^2}\, z_n\right\|^2$$

$$\approx \left(\frac{\lambda}{\lambda_*}-1\right)\sigma^2 + \left(\frac{\lambda}{\lambda_*}\right)(\sigma_*^2-\sigma^2)$$

$$= \sigma^2 - 2\left(\frac{\lambda}{\lambda_*}\right)\sigma^2 + \left(\frac{\lambda}{\lambda_*}\right)^2\sigma_*^2 \qquad (2)$$

$(1)-(2) = \sigma^2$

$$\frac{1}{n}\|X(\hat{\beta}(\hat{\lambda})-\beta_0)\|^2 = \min_{\lambda\in[\epsilon,K]}\frac{1}{n}\|X(\hat{\beta}(\lambda)-\beta_0)\|^2$$

P4

We have shown $\hat{\sigma}^*$

to give $\hat{\sigma}$ we use $\sigma^{*2} = \sigma^2 + \frac{1}{\delta}\mathbb{E}\left[\eta_1(\beta_0 + \sigma_* Z; \lambda_*) - \beta_0\right]^2$

suppose we have an est.

$$\hat{R}(\lambda) = \frac{1}{p}\|\hat{\beta}(\lambda) - \beta_0\|^2$$

Then $\hat{\sigma}^2 = \hat{\sigma}_*^2 - \frac{1}{\delta}\hat{R}(\lambda)$

$$\hat{R}(\lambda) = \hat{\sigma}_*^2 \cdot \left(2\frac{\|\hat{\beta}\|_0}{p} - 1\right) + \frac{\|X^T(y - X\hat{\beta}(\lambda))\|^2/n}{(1 - \|\hat{\beta}\|_0/n)^2}$$

prop $\sup_\lambda \left|\hat{R}(\lambda) - \frac{1}{p}\|\hat{\beta}(\lambda) - \beta_0\|^2\right| = 0$

$\approx$ use $(*1)$ for numerator

$$\frac{\frac{1}{n}\|(\hat{\beta}^* - \hat{\beta})\frac{\lambda}{\lambda_*}\|^2}{\left(\frac{\lambda}{\lambda_*}\right)^2}$$

$$\left(\frac{1}{p}\|\eta_1(\beta_0 + \sigma_* Z_p; \lambda_*) - \beta_0 - \sigma_* Z_p\|^2\right)$$

$$= \underbrace{\frac{1}{p}\|\eta_1(\beta_0 + \sigma_* Z; \lambda_*) - \beta_0\|^2}_{\frac{1}{p}\|\hat{\beta} - \beta_0\|^2} + \sigma_*^2 - 2\sigma_* \cdot \underbrace{\frac{1}{p}\langle\eta_1(\beta_0 + \sigma_* Z_p) - \beta_0, Z_p\rangle}_{(a)}$$

Stein's lemma

(gaussian integral by parts) $Z \sim N(0,1)$ $\mathbb{E}[Zf(Z)] = \mathbb{E}[f'(Z)]$

$$Z \sim N(0, I_p) \quad \mathbb{E}[\langle Z, f(Z)\rangle] = \sum_{j=1}^p \mathbb{E} Z_j f_j(Z_j) = \sum_{j=1}^p \mathbb{E} f_j'(Z)$$

where $f: \mathbb{R}^p \to \mathbb{R}^p$, $f(Z) = (f_j(Z_j))_{j=1}^p$ $\quad = \mathbb{E}\,\mathrm{div}_f(Z_p)$

$$(a) = \frac{\sigma_*}{p}\sum_{j=1}^p \mathbb{E}\,\eta_1'(\beta_{0,j} + \sigma_* Z; \lambda_*)$$

$$= \sigma_* \mathbb{P}(|\beta_0 + \sigma_* Z| > \lambda_*)$$

$$= \sigma_* \delta\left(1 - \frac{\lambda}{\lambda_*}\right)$$

which is non ignorable in ~~prop~~ propotional regime

$$(\circ) = \frac{1}{p}\|\hat{\beta} - \beta_0\|^2 + \sigma_*^2 - 2\sigma_*^2 \cdot \frac{\|\hat{\beta}\|_0}{p}$$

Application 3 Adaptive tuning by cross-validation k-fold

$$\begin{bmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(k)} \end{bmatrix} \in \mathbb{R}^{\frac{n}{k} \times p}, \quad = X^{(-1)} \text{ which dim is } (n_k \times p)$$

$n - \frac{n}{k}$

$n_k =$

for $i = 1, \dots, k$

use $2 \sim k$ $\rightarrow$ $\hat{\beta}^i(\lambda) = \arg\min\left\{\frac{1}{2n_k}\|y^{(-i)} - X^{(-i)}\beta\|^2 + \frac{\lambda}{n_k}\|\beta\|_1\right\}$

eg. use the first fold $\rightarrow$ $\hat{R}^{cv}(\lambda) = \frac{1}{p}\sum_{i=1}^k \|y^{(i)} - X^{(i)}\hat{\beta}^{(i)}(\lambda)\|^2$

$$\approx \frac{1}{p}\sum_{i=1}^k \mathbb{E}_{(X^{(i)}, y^{(i)})}\|y^{(i)} - X^{(i)}\hat{\beta}^{(i)}\|^2$$

treat as fixed

$$= \frac{1}{p}\sum_{i=1}^k (\hat{\beta}^i - \beta_0)^T \underbrace{\mathbb{E}[X^{(i)T}X^{(i)}]}_{\frac{n}{k}\cdot\frac{1}{n}I}(\hat{\beta}^i - \beta) + \underbrace{\mathbb{E}\|\varepsilon^{(i)}\|^2}_{\frac{n}{k}\cdot\sigma^2}$$

$$= \frac{1}{k}\sum_{i=1}^k \frac{1}{p}\|\hat{\beta}^i - \beta_0\|^2 + \delta\sigma^2$$

k 比较大

$$\approx \frac{1}{p}\|\hat{\beta}(\lambda) - \beta_0\|^2 + \delta\sigma^2$$

$$\hat{\lambda}^{cv} = \arg\min_{\lambda \in [\epsilon, k]} \hat{R}^{cv}(\lambda) \Rightarrow \frac{1}{p}\|\hat{\beta}(\hat{\lambda}^{cv}) - \beta_0\|^2 = \min_{\lambda \in [\epsilon, k]} \frac{1}{p}\|\hat{\beta}(\lambda) - \beta_0\|^2$$

P5

$\hat{\beta} = \arg\min \frac{1}{2n}\sum_{i=1}^{n}(Y_i - X_i^T\beta)^2 + \frac{\lambda}{n}\sum_{j=1}^{p} f(\beta_j)$ or $(+\lambda f(\beta))$

elementwise seperatable

· General fixed point equation (non-rigorious)

$$\begin{cases} \sigma_*^2 = \sigma^2 + \frac{1}{n}\mathbb{E}\|\mathrm{prox}_{\lambda_* f}(\beta_0 + \sigma_* Z_p) - \underbrace{\beta_0}_{p\times 1}\|^2 \\ \frac{\lambda}{\lambda_*} = 1 - \frac{1}{n}\mathbb{E}\,\underbrace{\mathrm{div}_{\mathrm{prox}_{\lambda_* f}}}_{\text{"degree of freedom" (e.g. "sporisity" in lasso)}}(\beta_0 + \sigma_* Z). \end{cases}$$

General Gaussian design $(\Sigma \neq I)$: Celentano, Montanari Wei (2023)

1) Gaussian comparison method

2) Approximate Message Passing

3) leave-one-out method

1) Convex Gaussian Min-Max Thm.

$D_u \subseteq \mathbb{R}^p$, $D_v = \mathbb{R}^n$, are compact sets

$Q: D_u \times D_v \to \mathbb{R}$ continuous function

$G \in \mathbb{R}^{n\times p}$ G_{ij} iid $N(0,1)$

$$C(G) = \min_{u\in D_u}\max_{v\in D_v}\{v^T G u + Q(u,v)\}$$

$g \sim N(0, I_p)$, $h \sim N(0, I_n)$

$$L(g,h) \triangleq \min_{u\in D_u}\max_{v\in D_v}\{\|v\|\langle g,u\rangle + \|u\|\langle h,v\rangle + Q(u,v)\}$$

Then, for $t \in \mathbb{R}$

$$\mathbb{P}(C(G) \leq t) \leq 2\mathbb{P}(L(g,h) \leq t)$$

If D_u, D_v are convex

Q : convex – concave

$$\mathbb{P}(C(G) \geq t) \leq 2\mathbb{P}(L(g,h) \geq t)$$

P6 Regularied least square

$$\hat\beta = \arg\min_\beta \left\{ \frac{1}{2n}\|y-X\beta\|^2 + \frac{1}{n}\sum_{j=1}^{p} f(\beta_j) \right\}$$

$\|a-b\|^2 \geq 0$

$X_{ij} \overset{iid}{\sim} N(0, \frac{1}{n})$? matlab

$$\frac{\|\tilde v\|^2}{2} = \max_u \langle u, \tilde v\rangle - \frac{\|u\|^2}{2}$$

$\tilde v$ here is $y - X\beta$

$$= \min_\beta \max_u \frac{1}{n}\langle u, y - X\beta\rangle - \frac{1}{2n}\|u\|^2 + \frac{1}{n}\sum_{j=1}^{p} f(\beta_j)$$

$$= \min_\beta \max_u \frac{1}{n}\langle u, X(\beta-\beta_0) - \varepsilon\rangle - \frac{1}{2}\|u\|^2 + (\ast\ast)$$

$\beta - \beta_0 = v \Rightarrow \beta = \beta_0 + v$

$$= \min_v \max_u \frac{1}{n}\langle u, Xv\rangle - \frac{1}{n}\langle u, \varepsilon\rangle - \frac{1}{2}\|u\|^2 + \frac{1}{n}\sum_{j=1}^{p} f(v_j + \beta_{0,j})$$

def. $C(G) \triangleq \min_v \max_u \frac{1}{n^{3/2}}\langle u, Gv\rangle - \frac{1}{n}\langle u, \varepsilon\rangle - \frac{1}{2n}\|u\|^2 + \bar f(v)$

$$= \min_v H(v)$$

$$L(g,h) = \min_v \max_u \frac{1}{n^{3/2}}\|u\|\langle g, v\rangle + \frac{1}{n^{3/2}}\|v\|\langle h, u\rangle - \frac{1}{n}\langle u, \varepsilon\rangle - \frac{1}{2n}\|u\|^2 + \bar f(v)$$

$$= \min_v L(v)$$

suppose $\hat\beta \in S_n \subseteq \mathbb{R}^p$, $S_n = \left\{ \beta \in \mathbb{R}^p, \left| \frac{\|\beta - \beta_0\|^2}{p} - C_* \right| \leq o(1) \right\}$

set $V_n = S_n - \beta_0$, WTS $v \in V_n$

1° $\mathbb{P}\left(\min_{v \notin V_n} H(v) \leq t_1\right) \leq 2\mathbb{P}\left(\min_{v \notin V_n} L(v) \leq t_1\right) \leq (\cdots)$

Lower bound for $\min_{v \notin V_n} H(v)$ ← global minimum

2° $\mathbb{P}\left(\min_v H(v) \geq t_2\right) \leq 2\mathbb{P}\left(\min_{v \notin V_n} L(v) \geq t_2\right) \leq (\cdots)$

upper bound for $\min_{v \notin V_n} H(v)$

$$\boxed{\begin{array}{l} ① \min_v L(v) = c_0 \quad \text{w.h.p.} \\ ② \min_{v \notin V_n} L(v) - \min_v L(v) / c_0 \geq \varepsilon \text{ ("or") w.h.p.} \end{array} \Rightarrow \hat v \in V_n \text{ w.h.p.}}$$

$$L(v) = \max_u -\frac{1}{n^{3/2}}\|u\|\langle g, v\rangle + \frac{1}{n^{3/2}}\|v\|\langle h, u\rangle - \frac{1}{n}\langle u, \varepsilon\rangle - \frac{1}{2n}\|u\|^2 + \bar f(v)$$

$$= \max_{b>0} \max_{\|u\|=b} \left(-\frac{1}{n^{3/2}} b\langle g, v\rangle - \frac{1}{2n} b^2 + \bar f(v) + \left\langle u, \frac{1}{n^{3/2}}\|v\| h - \frac{\varepsilon}{n}\right\rangle \right)$$

~~$= \max_{b>0} b($~~

$\| \max(\cdot)$

$b\left\| \frac{1}{n^{3/2}}\|v\| h - \frac{\varepsilon}{n} \right\|$

$$L(v) = \max_{b>0} b\left(\left\| \frac{1}{n^{3/2}}\|v\| h - \frac{\varepsilon}{n} \right\| - \frac{1}{n^{3/2}}\langle g, v\rangle \right) - \frac{1}{2}\left(\frac{b}{\sqrt{n}}\right)^2 + \bar{f}(v)$$

$$\boxed{\frac{b}{\sqrt{n}} = b}$$

$$\max_{b>0} b\left(\left\| \frac{1}{n}\|v\| h - \frac{\varepsilon}{\sqrt{n}} \right\| - \frac{1}{n}\langle g, v\rangle \right) - \frac{1}{2} b^2 + \bar{f}(v)$$

recall that外面还要对 v 作 min

this term

but is non smooth, use $\|t\| = \min_{\gamma} \frac{1}{2}\left(\frac{\|t\|^2}{\gamma} + \gamma\right)$

$$\max_{b>0} \min_{\gamma>0} \frac{b}{2\gamma}\left\| \frac{1}{n}\|v\| h - \frac{\varepsilon}{\sqrt{n}} \right\|^2 + \frac{b\gamma}{2} - \frac{b}{n}\langle g, v\rangle - \frac{b^2}{2} + \bar{f}(v)$$

concentration using Expectation

$$\approx \max_{b>0} \min_{\gamma>0} \frac{b}{2\gamma}\left(\frac{\|v\|^2}{n} + \sigma^2\right) + \frac{b\gamma}{2} - \frac{b}{n}\langle g, v\rangle - \frac{b^2}{2} + \bar{f}(v)$$

$$= \max_{b>0} \min_{\gamma>0} \left(\frac{\sigma^2}{\gamma} + \gamma\right)\frac{b}{2} - \frac{b^2}{2} + \underbrace{\frac{b}{2\gamma}\frac{\|v\|^2}{n} - b\cdot\frac{\langle g, v\rangle}{n} + \bar{f}(v)}_{(*2)}$$

$$\triangleq \max_{b>0} \min_{\gamma>0} \psi(v, b, \gamma)$$

vector scalar

$$\min_v L(v) = \min_v \max_{b>0} \min_{\gamma>0} \psi(v, b, \gamma) = \max_{b>0} \min_{\gamma} \left(\min_v \psi(v, b, \gamma) \right)$$

for given b, γ, $\min_v L(v)$

$$= \min (*2)$$

$$= \min_v \left\{ \frac{b}{2\gamma}\frac{\|v\|^2}{n} - b\frac{\langle g, v\rangle}{n} + \bar{f}(v) \right\}$$

$$= \min_v \frac{b}{2\gamma n}\left(\|v\|^2 - 2\langle \gamma g, v\rangle + \|\gamma g\|^2 \right) - \frac{b\gamma}{2n}\|g\|^2$$

$$= \min_v \left\{ \frac{b}{2\gamma n}\|v - \gamma g\|^2 + \bar{f}(v) \right\} - \frac{b\gamma}{2n}\|g\|^2$$

$$\boxed{\begin{aligned} e_f(x;\tau) &\triangleq \min_z \left\{ \frac{1}{2\tau}\|x - z\|^2 + f(z) \right\} \\ \operatorname{prox}_f(x;\tau) &\triangleq \operatorname{argmin}_z \left\{ \frac{1}{2\tau}\|x - z\|^2 + f(z) \right\} \end{aligned}}$$

v is z $\quad$ γg is x

$$\bar{f}(v) = \frac{1}{n}\sum_{j=1}^{p} f(v_j + \beta_{0j})$$

$$= \frac{1}{n} e_f\left(\beta_0 + \gamma g; \frac{\gamma}{b}\right) - \frac{b\gamma}{2n}\|g\|^2, \quad \hat{v} = \operatorname{prox}_f\left(b_0 + \gamma g, \frac{\gamma}{b}\right)$$

$$\boxed{\min_v L(v) = \max_{b} \min_{\gamma} \left(\frac{\sigma^2}{\gamma} + \gamma\right)\frac{b}{2} - \frac{b^2}{2} + \frac{1}{n} e_f\left(b_0 + \gamma g; \frac{\gamma}{b}\right) - \frac{b\gamma}{2}\frac{\|g\|^2}{n}}$$

对 b, γ 求导

$g \sim N(0, I_p)$

28 lemma 1) $\nabla_x e_f(x;T) = \frac{1}{T}(x - \mathrm{prox}(x;T))$

2) $\frac{\partial}{\partial T} e_f(x;T) = -\frac{1}{2T^2}\|x - \mathrm{prox}_f(x;T)\|^2$

proof: "Variational Analysis" Thm 2.20

辅助函数 $H(v) = e_f(x+\varepsilon;T) - e_f(x,T) - \langle \varepsilon, \frac{1}{T}(x - \underbrace{\mathrm{prox}_f(x;T)}_{w}) \rangle$

WTS $\nabla H(\varepsilon)_{\varepsilon=0} = 0$

$e_f(x) = \frac{1}{2T}\|x-z\|^2 + f(z)$

~~$e_f(x+\varepsilon) =$~~

$e_f(x+\varepsilon;T) \le \frac{1}{2T}\|x+\varepsilon-z\|^2 + f(z)$

$H(\varepsilon) \le \frac{1}{2T}\|x+\varepsilon-z\|^2 - \frac{1}{2T}\|x-z\|^2 - \langle\varepsilon, \frac{1}{T}(x-\mathrm{prox}_f(x;T))\rangle$

$= \frac{1}{2T}(\|\varepsilon\|^2 - 2\langle x-z,\varepsilon\rangle) - \langle \varepsilon, \frac{1}{T}(x-\mathrm{prox}_f(x;T))\rangle$

$\langle e_i, \nabla H(\varepsilon)\rangle|_{\varepsilon=0} = 0$

2) $H(\varepsilon) \triangleq e_f(x;T+\varepsilon) - e_f(x;T) + \frac{\varepsilon}{2T^2}(x-z)^2$ scalar

$e_f(x;T) = \frac{1}{2T}\|x-z\|^2 + f(z)$

$e_f(x;T+\varepsilon) \le \frac{1}{2(T+\varepsilon)}\|x-z\|^2 + f(z)$

$H(\varepsilon) \le \left(\frac{1}{2(T+\varepsilon)} - \frac{1}{2T} + \frac{\varepsilon}{2T^2}\right)\|x-z\|^2$

$= \frac{\varepsilon^2}{2T^2(T+\varepsilon)}$

$\to \lim_{\varepsilon\downarrow 0} H(\varepsilon)/\varepsilon \le 0$ $\quad \lim_{\varepsilon\uparrow}$ 同理

$\min_v L(v) \triangleq \max_\beta \min_\gamma \Phi(\beta,\gamma)$

$-\frac{\gamma}{b^2}$

$\frac{\partial}{\partial b} e_f(\beta_0+\gamma g; \frac{\gamma}{b}) = -\frac{1}{2(\frac{\gamma}{b})^2}\|\beta_0+\gamma g - \mathrm{prox}_f(\beta_0+\gamma g)\|^2 \cdot \frac{\partial}{\partial b}\left(\frac{\gamma}{b}\right)$

$= \frac{1}{2\gamma}\|\beta_0+\gamma g - \mathrm{prox}_f(\beta_0+\gamma g)\|^2$

$= \frac{1}{2\gamma}\|\beta_0+\gamma g\|^2 + \frac{\gamma}{2}\|g\|^2 + \langle\beta_0 - \mathrm{prox}_f(\beta_0+\gamma g), g\rangle$

$\frac{\partial}{\partial\gamma} e_f(\beta_0+\gamma g, \frac{\gamma}{\beta}) = \frac{1}{\gamma/b}\langle \beta_0+\gamma g - \mathrm{prox}_f(\beta_0+\gamma g), g\rangle$

$- \frac{1}{2(\gamma/b)^2}\|\beta_0+\gamma g - \mathrm{prox}_f(\beta_0+\gamma g;\frac{\gamma}{b})\|^2\cdot\frac{1}{b}$

$= \frac{b}{\gamma}\langle\beta_0 - \mathrm{prox}_f(\dots), g\rangle + b\|g\|^2 - \frac{b}{2\gamma^2}\left(\|\beta_0 - \mathrm{prox}(\dots)\|^2 + \gamma^2\|g\|^2 + 2\langle\cdot,\cdot\rangle\right)$

$= \frac{b}{2}\|g\|^2 - \frac{b}{2\gamma^2}\|\beta_0 - \mathrm{prox}_f(\dots)\|^2$

$\nabla\Phi$ of The total term

① $0 = \frac{\partial\Phi}{\partial\beta} = \frac{1}{2}\left(\frac{\sigma^2}{\gamma}+\gamma\right) - \beta + \frac{1}{2n\gamma}\|\beta_0 - \mathrm{prox}_f(\dots)\|^2 + \cancel{\frac{\gamma}{2n}\|g\|^2} + \frac{1}{n}\langle\beta_0-\mathrm{prox}(\dots),g\rangle - \cancel{\frac{\gamma}{2}\frac{\|g\|^2}{n}}$

② $0 = \frac{\partial\Phi}{\partial\gamma} = \left(-\frac{\sigma^2}{\gamma^2}+1\right)\cdot\frac{b}{2} + \cancel{\frac{b}{2n}\|g\|^2} - \frac{b}{2\gamma^2 n}\|\beta_0 - \mathrm{prox}_f(\dots)\|^2 - \cancel{\frac{b}{2n}\|g\|^2}$

② $\Leftrightarrow$ $\gamma^2 = \underbrace{\frac{1}{n}\|\beta_0-\mathrm{prox}_f(\dots)\|^2}$ $+\sigma^2$

$\downarrow$ plugin this term $=\gamma^2-\sigma^2$ to ①

get

$0 = \gamma - \beta + \frac{1}{n}\langle\beta_0 - \mathrm{prox}_f(\dots), g\rangle$ by concentration $\downarrow$

P9

$$\begin{cases} \gamma^2 = \sigma^2 + \frac{1}{n}\mathbb{E}\|\mathrm{prox}_f(\beta_0+\gamma g, \frac{\gamma}{\beta}) - \beta_0\|^2 \\ \beta = \gamma - \frac{1}{n}\mathbb{E}\langle \underbrace{\mathrm{prox}_f(\beta_0+\gamma g; \frac{\gamma}{\beta}) - \beta_0, g}_{\| \;\; \frac{\gamma}{n}\mathbb{E}\,\mathrm{div}_{\mathrm{prox}_f}(\beta_0+\gamma g;\frac{\gamma}{b})}\rangle \end{cases}$$

$\mathbb{E}\langle g, F(g)\rangle = \mathbb{E}\,\mathrm{div}_F(g)$

$$\Leftrightarrow \frac{1}{\gamma/\beta} = 1 - \frac{1}{n}\mathbb{E}\,\mathrm{div}_{\mathrm{prox}_f}(\beta_0+\gamma g; \frac{\gamma}{b})$$

$\gamma \quad \frac{\gamma}{\beta}$

---

phase transition for logistic MLE

iid (x_i, y_i) $x_i \sim N(0, I_p)$

$p(y_i = 1 | x_i) = \sigma(\langle x_i, \beta_0\rangle)$ $\sigma(\cdot) = \frac{1}{1+e^{-x}}$

log-Likelihood

$l(\beta) = \sum^n -\log[1+\exp(-y_i\langle x_i, \beta\rangle)]$

MLE : $\hat\beta \in \mathop{\mathrm{argmax}}_{\beta\in\mathbb{R}^p} l(\beta)$

~~$y_i\langle x_i\beta\rangle$~~ $y_i\langle x_i, \beta\rangle > 0$: correctly classified

< 0 incorrectly

Thm (Candès & Sur, 2019)

$(Y, V) \overset{d}{=} (Y, YX)$ where $X \sim N(0,1)$ $\mathbb{P}(Y=1|X) = \sigma(\|\beta_0\| X)$

Then $Z \sim N(0,1)$ $Z \perp\!\!\!\perp (Y, V)$

$$\begin{cases} \frac{p}{n} \geq \mathbb{E}\min_{t_0,t_1\in\mathbb{R}} (t_0 Y + t_1 V - Z)_+^2 + \varepsilon \Rightarrow \text{MLE } \hat\beta \text{ does not exist w.h.p} \\ \frac{p}{n} \leq \cdots - \varepsilon \Rightarrow \text{MLE } \hat\beta \text{ exists w.h.p.} \end{cases}$$

Gaussian random projection

Thm (Amelunxen, Lotz, M.., Tropp. 2019)

L: fixed subspace $\subseteq \mathbb{R}^n$

K : non-trivial closed convex cone $\subseteq \mathbb{R}^n$

$O_n \subseteq \mathbb{R}^{n\times n}$ Haar measure of orthogonal group $O(n)$

Then

$\dim(L) + \dim(K) \ll n \Rightarrow K \cap O_n L = \{0\}$, w.h.p.

$\dim(L) + \dim(K) \gg n \Rightarrow K \cap O_n L \neq \{0\}$, w.h.p.

$x_i \sim N(0, I_p)$ $\mathbb{P}(y_i = 1 | x_i) = \sigma(\langle x_i, \beta_0\rangle)$

equivently $\mathbb{P}(y_i | x_i) = \sigma(\|\beta_0\| x_{i1})$

so we may assume

$$(\underset{\mathbb{R}}{y_i}, \underset{\substack{\mathbb{R}^p \\ \| \\ (y_i x_{i1}, \cdots, y_i x_{ip})}}{y_i x_i}) \overset{d}{=} (Y, V, x_2, \cdots x_p) \quad \mathbb{P}(Y=1|X_1) = \sigma(\|\beta\| X_1)$$

换坐标
到第一个X_1上 $\longrightarrow$

$V = YX_1$

$X_2, \cdots X_p \overset{iid}{\sim} N(0,1)$

⑩ $K \subseteq \mathbb{R}^n$, closed convex cone

$G \in \mathbb{R}^{m\times n}$ with iid $N(0,1)$ e

$L \in \mathbb{R}^m$: closed convex cone

$L \cap GK = \{0\}$?

Then $\dim(L) + \dim(F) << m \Rightarrow L \cap GK = \{0\}$ w.h.p.

$\dim(L) + \dim(F) >> m \Rightarrow L \cap GK \neq \{0\}$ w.h.p.

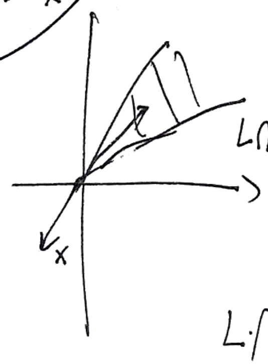
proof: $m=n$, $GL \overset{d}{=} O_n L \overset{d}{=} \mathrm{null}(G_{n\times l, n})$

X gaussian $\overset{d}{=}$ X rotation $\quad \uparrow \mathbb{R}^{(n-l)\times n}$

Gaussian comparison heuristic

Support function $\forall$ compact convex set K_0

$$h_{K_0}(x) = \sup_{\mu \in K} \langle x, \mu\rangle$$



$\longrightarrow +\infty$ or $-\infty$

$L \cap G_K$ $\quad$ so need compactness

$K_n = K \cap B_n(1)$

$L \cap GK \neq \{0\} \Leftrightarrow \sup_{x \in \partial B_n} h_{L\cap GK_n}(x) > 0$

↓単位球上

$$h_{L\cap GK_n}(x) = \sup_{w \in L \cap GK_n} \langle x, w\rangle$$

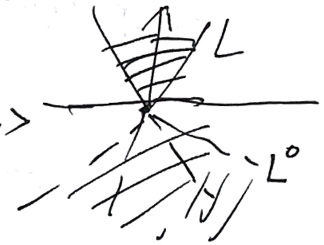
$$= \sup_{\substack{u \in K_n \\ Gu \in L}} \langle x, Gu\rangle$$

$$= \sup_{u \in K_n} \inf_{v \in L^\circ} \langle x, Gu\rangle - \langle v, Gu\rangle$$

$$= \sup_{u \in K_n} \inf_{v \in L^\circ} \langle x - v, Gu\rangle$$

$v \in L \Leftrightarrow \forall v^\circ \in L^\circ, \langle v, v^\circ\rangle \le 0$

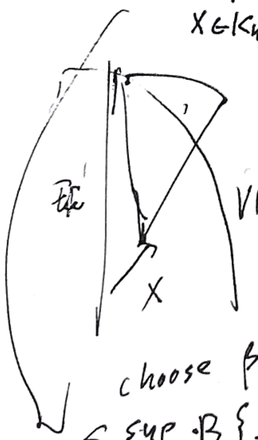
$Gu \quad \Leftrightarrow \sup_{v \in L^\circ} \langle v, v^\circ\rangle = 0$



$$h_{L\cap GK_n}(x) = \sup_{x \in K_n} \inf_{v \in L^\circ} \left\{ \|u\| \langle h, v-x\rangle + \|v-x\| \langle g, u\rangle \right\}$$

$$= \sup_{x \in K_n} \inf_{\beta} \Big\{ \beta \langle g, u\rangle + \|u\| \inf_{\substack{v \in L^\circ \\ \|v-x\|=\beta}} \langle h, v-x\rangle \Big\}$$

$$= \sup_{x \in K_n} \inf_{\beta} \beta \Big\{ \langle g, u\rangle - \|u\| \sup_{\substack{v \in L^\circ \\ \|v-x\|=\beta}} \Big\langle h, \frac{v-x}{\|v-x\|} \Big\rangle \Big\}$$



$$L^\circ\{x;\beta\} \triangleq \left\{ \frac{v-x}{\|v-x\|} : v \in L^\circ, \|v-x\| = \beta \right\} \subseteq \partial B_n$$

$\approx L^\circ \cap \partial B_n \quad , \beta >> 1$

choose $\beta = B >> 1$

$$\le \sup B \{\langle g, u\rangle - \|u\| \sup \langle h, v\rangle\}$$

$$= B\Big(\sup_{u \in K \cap \partial B_n} \langle g, u\rangle - \sup_{v \in L^\circ \cap \partial B_n} \langle h, v\rangle \Big).$$

$$\delta(K) = \mathbb{E}\Big(\sup_{u \in K \cap \partial B_n} \langle g, u\rangle \Big)^2$$

$$= B(\sqrt{\delta(K)} - \sqrt{m - \delta(L^\circ)})_+$$

20220510

(11)

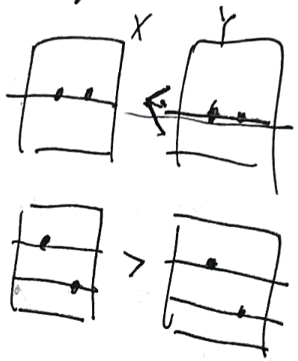
Thm (classical Min-Max Thm)

(X_{ij}) (Y_{ij}) $i\in[n], j\in[m]$: Gaussian vector

① $\mathbb{E} X_{ij} X_{ik} \le \mathbb{E} Y_{ij} Y_{ik} \quad \forall i,j,k$

② $\mathbb{E} X_{ij} X_{lk} \ge \mathbb{E} Y_{ij} Y_{lk} \quad \forall i\neq l, j,k$

③ $\mathbb{E} X_{ij}^2 = \mathbb{E} Y_{ij}^2$



Then for all $\{\lambda_{ij}\} \in \mathbb{R}$

$$\mathbb{P}\Big(\bigcap_{i=1}^{n}\bigcup_{j=1}^{m}\{Y_{ij} > \lambda_{ij}\}\Big) \le \mathbb{P}\Big(\bigcap_{i=1}^{n}\bigcup_{j=1}^{m}\{X_{ij} > \lambda_{ij}\}\Big)$$

$$\min_{i\in[n]}\max_{j\in[m]}(Y_{ij}-\lambda_{ij}) > 0$$

(Ledoux & Talagrand, Prob. in Banach Space, Ch. 3)

Corollary: $D_u \subseteq \mathbb{R}^n$, $D_u \subseteq \mathbb{R}^m$, compact sets

$Q: D_u \times D_v \to \mathbb{R}$

$(X(u,v))$ $(Y(u,v))$ $u\in D_u, v\in D_v$

① 连续性 $(u,v) \mapsto X(u,v)$ $Y(u,v)$ continuous a.s.

② $\mathbb{E} X(u,v) Y(u,v') \le \mathbb{E} Y(u,v) Y(u,v')$ $\quad \mathbb{E} X^2(u,v) = \mathbb{E} Y^2(u,v)$

③ $\mathbb{E} X(u,v) Y(u',v') \ge \mathbb{E} Y(u,v) Y(u',v')$ $\quad u\neq u'$

"$-Q(u,v)$"

$$\mathbb{P}\Big(\min_{u\in D_u}\max_{v\in D_v} Y(u,v) + Q(u,v) \le t\Big) \le \mathbb{P}\Big(\min_{u}\max_{v} X(u,v) + Q(u,v) \le t\Big)$$

recall

$$C(G) = \min_{u\in D_u}\max_{v\in D_v} v^T G u + Q(u,v)$$

$$L(g,h) = \min_{u}\max_{v} \|u\| g^T u + \|u\| h^T v + Q(u,v)$$

令 $X(u,v) = \|u\| g^T u + \|u\| h^T v$

$Y(u,v) = v^T G u + \|u\|\cdot\|v\|\cdot z$

$z \sim N(0,1) \perp\!\!\!\perp (G, g, h)$

计算 $\mathbb{E}[Y(u,v)\, Y(u',v')]$ $\quad$ cross-term cancelled

$= \mathbb{E}\big[(v^T G u)((v')^T G u') + \|u\|\cdot\|v\|\cdot\|u'\|\cdot\|v'\|\cdot z^2\big]$

$= \mathbb{E}\Big(\sum_{ij} v_i G_{ij} u_j\Big)\Big(\sum_{i'j'} v_{i'}' G_{i'j'} u_{j'}'\Big) + (2)$

$= \sum_{(i,j)(i',j')} v_i v_{i'}' u_j u_{j'}' \underbrace{\mathbb{E} G_{ij} G_{i'j'}}_{(i=i',\ j=j')} + (2)$

$= \sum_{i,j} v_i v_i' u_j u_j' + (2) = \langle v, v'\rangle\langle u, u'\rangle + \|u\|\|v\|\cdot\|u'\|\|v'\| z^2$

(P12)

$\mathbb{E}\, X(u,v)\, X(u',v')$

$= \mathbb{E}\, \|v\|\|v'\| (g^T u)(g^T u') + \|u\|\cdot\|u'\| (h^T v)(h^T v')$

$= \mathbb{E}\, \|v\|\|v'\|\, u^T [g g^T] u' + \cdots$

$= \|v\|\|v'\| \langle u, u' \rangle + \|u\|\|u'\| \langle v, v' \rangle$

by $\max_u \min_v H(u,v) \leq \min_v \max_u H(u,v)$ (c)

(d) ② $\Leftrightarrow \left\{ \max_v \min_u \left(-\|v\| g^T u + \|u\| h^T v + Q(u,v) \right) \geq -T \right\}$

(b) $\Leftrightarrow \left\{ T \leq \max_v \min_u (+) \overset{(c)}{\leq} \min_u \max_v \left(\|v\| g^T u + \|u\| h^T v + Q(u,v) \right) \right\}$

$\boxed{\begin{array}{l} \max A \leq b \iff \min -A \geq -b \quad ⓐ \\ \text{(b) 移项} \qquad \text{(c)} \max_u \min_v \leq \min_v \max_u \end{array}}$

$\mathbb{E}\, Y(u,v)\, Y(u',v') - \mathbb{E}\, X(u,v)\, X(u',v')$

$= \left(\|u\|\cdot\|u'\| - \langle u, u' \rangle \right) \left(\|v\|\|v'\| - \langle v, v' \rangle \right) \geq 0$

✓ $u = u'$ Take 0

$$\boxed{\begin{aligned} &\mathbb{P}\left(\min_u \max_v X(u,v) + Q(u,v) \leq T \right) \\ &\geq \mathbb{P}\left(\min_u \max_v v^T G u + \|u\|\cdot\|v\| + Q(u,v) \leq T \right) \\ &\geq \tfrac{1}{2} \mathbb{P}\left(\text{——} \mid z \leq 0 \right) \\ &\geq \tfrac{1}{2} \mathbb{P}\left(\min_u \max_v u^T G v + Q(u,v) \leq T \right) \end{aligned}}$$

flip u, v

$$\mathbb{P}\left(\min_v \max_u v^T G u - Q(u,v) \leq -T \right) \leq 2\,\mathbb{P}\left(\min_v \max_u \|v\| g^T u - \|u\| h^T v - Q(u,v) \leq -T \right)$$

① ②

$$\max_v \min_u v^T G u + Q(u,v) \geq t \overset{\text{duality}}{=} \min_u \max_v (\text{—})$$