STAT DRP Week 2

Thursday, January 18, 2024 1:58 PM

$y = X\theta^* + w$

X ↓ random matrix

Empirical MSE $= \min_{\gamma} \frac{1}{n} \| Z\gamma - y \|_2^2 = \frac{1}{n}\sum_{i=1}^{n} (\gamma^T z_i - y_i)^2$

$\gamma = (Z^T Z)^{-1} (Z^T y)$ $\qquad \Upsilon = y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

$P\left((Z^T Z)^{-1} \text{ exists} \right) = 1$

$\det(Z^T Z) = \boxed{f(Z)}$ random variable

Q: $P(f(z) \neq 0) =$

(i) $n >> d$ $\qquad \text{span}\{z_1, z_2, \dots, z_d\} = \mathbb{R}^s$

n ↑ sample, d ↑ feature

Q $\quad z_{d+1} \in \mathbb{R}^d$? $\qquad \mathbb{R}^n \setminus \mathbb{R}^d$

$e_1, \dots, e_d$

$e_{d+1}, \dots, e_n$

if $z_{d+1} \subset \mathbb{R}^d$

$z_{d+1} = \sum_{i=1}^{n} a_i e_i \Rightarrow a_{d+1} = \dots = a_n = 0$

$\mathbb{P}(a_{d+1} = 0, \dots, a_n = 0) = \mathbb{P}((a_{d+1}, \dots, a_n) = 0) = 0$

$\Rightarrow \mathbb{P}(z_{d+1} \notin \mathbb{R}^d) = 1$

$z_1, \dots, z_n \overset{i.i.d}{\sim} P_z \qquad \mathbb{R}^n \quad (e_1, \dots, e_n)$

$z_n \sim P_z \qquad z_n = (z_n^{(1)}, \dots, z_n^{(d)})$

$z_n^{(1)} \sim P_{z^{(1)}}, \quad z_n^{(d)} \sim P_{z^{(d)}}$

$z_j = \sum_{i=1}^{n} z_j^{(i)} e_i$

$Z^T Z \in \mathbb{R}^{d \times d}$

$y = X\theta^* + w$

Hard sparsity $\quad \theta^* \in \mathbb{R}^d$, $\quad S(\theta^*)$ is defined as

$S(\theta^*) = \{ j \in \{1, \dots, d\} : \theta_j^* \neq 0 \}$

$d = 3$:

$\theta^* = (1, 0, 1)$, $\quad S(\theta^*) = \{1, 3\}$.

↑1 ↑2 ↑3

$\|\theta^*\|_0 = \sum_{j=1}^{d} \mathbb{1}\{\theta_j \neq 0\}$ $\qquad Y = f_\theta(X)$

$\|\theta^*\|_1 = \sum_{j=1}^{d} |\theta_j|$ $\qquad Y = X\theta_1 + X^2\theta_2 + o(X^2)$

(X, Y)

$\boxed{f(t) = \sum_{n=0}^{\infty} \frac{f^{(n)}(t_0)}{n!} (t - t_0)^n}$

$\underbrace{t - t_0 \qquad (t-t_0)^2 \qquad (t-t_0)^3 \qquad \cdots \qquad (t-t_0)^k}_{k \text{ features}}$

$t - t_0$ ↑ feature

$x \sim P_x \qquad n$ samples $\quad (x_1, y_1), \dots, (x_n, y_n)$

data augment $\qquad \left(\begin{pmatrix} x_1 \\ x_1^2 \\ \vdots \\ x_1^d \end{pmatrix}, y_1 \right) \quad \left(\begin{pmatrix} x_n \\ x_n^2 \\ \vdots \\ x_n^d \end{pmatrix}, y_n \right)$

Problem: factors ↓ people health ?

① diet

② sports

③ sleep

⋮

$z_i = (1, t_i, \dots, t_i^d)$

$f_\theta(t) = \theta_1 + \theta_2 t + \dots + \theta_{d+1} t^d$

$\boxed{y = Z\theta} = f_\theta(t)$

---

① $\min \|\theta\|_0$ such that $X\theta = y$

$\|\theta\|_0 = \sum_{j=1}^{d} \mathbb{1}\{\theta_j \neq 0\} \leftarrow 2^d$

② $\min \|\theta\|_1$ such that $X\theta = y$ conditions

$\|\theta\|_1 = \sum_{j=1}^{d} |\theta_j| = \sum_{j=1}^{d} c_j \theta_j$

$X \in \mathbb{R}^{n \times d}, \theta \in \mathbb{R}^d$

$\boxed{\text{Based there exist } \theta^* \text{ for ①}, \ Y = X\theta^*.}$

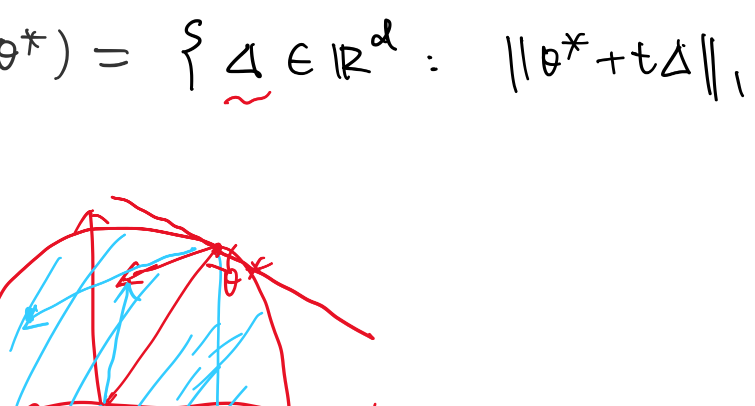
core question: Is θ^* a solution for ② ?

$\boxed{\min \|\theta\|_1 \text{ such that } X\theta = y}$

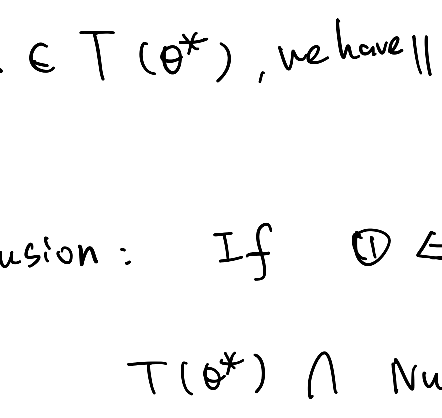
(*) θ^* is a feasible solution

(**) $\underbrace{\theta^* + \text{Null}(X)}_{\text{all feasible solutions}} = \{ \Delta \in \mathbb{R}^d : X\Delta = 0 \}$

$X\theta^* = y, X\Delta = 0, \quad X(\theta^* + \Delta) = y$



$T(\theta^*) = \{ \Delta \in \mathbb{R}^d : \|\theta^* + t\Delta\|_1 \leq \|\theta^*\|_1 \text{ for some } t \}$



$\Delta \in T(\theta^*)$, we have $\|\theta^* + t\Delta\|_1 \leq \|\theta^*\|_1$

Conclusion: If ① $\Leftrightarrow$ ②, we should require

$T(\theta^*) \cap \text{Null}(X) = \{0\}$.

$\Longleftrightarrow$

$\left. \begin{array}{l} \text{feasible} = \theta^* + \text{Null}(X) \\ \text{less } \|\cdot\|_1 = \theta^* + T(\theta^*) \end{array} \right\}$ have only intersection of $\{\theta^*\}$