APTS High-Dimensional Statistics: Assignments

Yi Yu*

July 8, 2022

1. Assume that
$$Y = X\beta^* + \varepsilon \in \mathbb{R}^n,$$
where $\beta^* \in \mathbb{R}^p$, $p < n$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. Let $\mu_{\min}$ be the smallest eigenvalue of $\Sigma = X^\top X/n > 0$. Let
$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n}\|Y - X\beta\|^2 + \lambda\|\beta\|_1 \right\}$$
be the Lasso estimator with tuning parameter λ. Use basic inequality to show that
$$\mathbb{E}\|\widehat{\beta} - \beta^*\|^2 \le \frac{8\sigma^2 p}{n\mu_{\min}} + \frac{16\lambda^2 p}{\mu_{\min}^2}.$$

2. Define the cone
$$\mathcal{C}(S; \alpha) = \{\theta \in \mathbb{R}^p | \|\theta_{S^c}\|_1 \le \alpha\|\theta_S\|_1\},$$
for $S \subset \{1, \ldots, p\}$ and $\alpha \ge 1$. We say a matrix $\Sigma \in \mathbb{R}^{p\times p}$ satisfies a restricted eigenvalue condition of order k with parameter α, γ, if
$$\theta^\top \Sigma\theta \ge \gamma^2\|\theta\|^2, \quad \forall\theta \in \mathcal{C}(S; \alpha), \forall|S| \le k.$$
We have learned in the course that for any $X \in \mathbb{R}^{n\times p}$ with i.i.d. $\mathcal{N}(0, \sigma)$ rows, there is a universal constant $c > 0$ such that
$$\frac{\|Xv\|}{\sqrt{n}} \ge \frac{1}{4}\|\Sigma^{1/2}v\| - 9\rho(\Sigma)\sqrt{\frac{\log(p)}{n}}\|v\|_1, \quad v \in \mathbb{R}^p,$$
with probability at least $1 - \exp(-cn)$, where $\rho^2(\Sigma) = \max_{j=1}^p(\Sigma_{jj})$.
Suppose that Σ satisfies the restricted eigenvalue condition of order k with parameter (α, γ). For universal constants $c_1, c_2, c_3 > 0$, assume that the sample size satisfies
$$n > c\frac{\rho^2(\Sigma)(1+\alpha)^2}{\gamma^2}k\log(p).$$
Show that the matrix $\widehat{\Sigma} = X^\top X/n$ satisfies the restricted eigenvalue condition with parameters $(\alpha, \gamma/8)$ with probability at least $1 - c_1\exp(-c_2 n)$.

*Department of Statistics, University of Warwick. Email: yi.yu.2@warwick.ac.uk

3. Let (T, d) be a metric space. For every $\varepsilon > 0$, show that

$$\mathcal{N}(T, \varepsilon) \leq \mathcal{M}(T, \varepsilon) \leq \mathcal{N}(T, \varepsilon/2),$$

   where $\mathcal{N}(T, \varepsilon)$ and $\mathcal{M}(T, \varepsilon)$ are the ε-covering number and ε-packing number, respectively.

4. Let $A \in \mathbb{R}^{m \times n}$ be a random matrix with i.i.d. entries $A_{ij} \sim \mathcal{N}(0, 1)$. Then there exists a universal constant $C > 0$ such that for any $t > 0$,

$$\|A\|_{\text{op}} \leq C(\sqrt{m} + \sqrt{n} + t),$$

   with probability at least $1 - 2\exp(-t^2)$.