APTS Assignment: High-Dimensional Statistics

July 2024

Please e-mail `pl128@cam.ac.uk` with any corrections.

Problem 1.

(a) Suppose $X_1, \ldots, X_m$ are zero-mean, sub-Gaussian random variables each with parameter σ. Prove that

$$\mathbb{P}\left(\max_{1\le j\le m} |X_i| \ge \sqrt{2\sigma^2 \log m} + u\right) \le 2\exp\left(-\frac{u^2}{2\sigma^2}\right), \quad \forall u > 0.$$

(b) Using the result of part (a), fill in the following missing detail in the probabilistic analysis of ℓ_2-bounds for the Lasso: Suppose $X \in \mathbb{R}^{n\times p}$ is a fixed design matrix which is column-normalized, so $\max_{1\le j\le p} \frac{\|X_j\|_2}{\sqrt{n}} \le C$, where X_j is the j^{th} column of X. Suppose $\epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Show that

$$\mathbb{P}\left(\left\|\frac{X^T\epsilon}{n}\right\|_\infty \ge C\sigma\left(\sqrt{\frac{2\log p}{n}} + t\right)\right) \le 2e^{-nt^2/2}, \qquad \forall t > 0,$$

so we can choose $t \asymp \sqrt{\frac{\log p}{n}}$ and conclude that the choice $\lambda \ge C'\sigma\sqrt{\frac{\log p}{n}}$ is valid, with high probability.

Problem 2. Consider the primal-dual witness proof for support recovery of the Lasso: Suppose the design matrix satisfies the mutual incoherence condition

$$\max_{j\in S^c} \|(X_S^T X_S)^{-1} X_S^T X_j\|_1 \le \alpha,$$

for some $\alpha \in [0, 1)$, and $X_S^T X_S$ is invertible. Also suppose the regularization parameter satisfies

$$\lambda \ge \frac{2}{1-\alpha}\left\|X_{S^c}^T(I - X_S(X_S^T X_S)^{-1}X_S^T)\frac{\epsilon}{n}\right\|_\infty.$$

Solving the system

$$\frac{1}{n}\begin{pmatrix} X_S^T X_S & X_S^T X_{S^c} \\ X_{S^c}^T X_S & X_{S^c}^T X_{S^c}\end{pmatrix}\begin{pmatrix}\widehat{\beta}_S - \beta_S^* \\ 0\end{pmatrix} - \frac{1}{n}\begin{pmatrix} X_S^T\epsilon \\ X_{S^c}^T\epsilon\end{pmatrix} + \lambda\begin{pmatrix}\hat{z}_S \\ \hat{z}_{S^c}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$

for $\hat{z}_{S^c}$, show that the strict dual feasibility condition $\|\hat{z}_{S^c}\|_\infty < 1$ is satisfied.

Problem 3.

(a) Consider three random variables (X, Y, Z) with strictly positive pdf $p(x, y, z)$. Prove that if $p(x, y, z) = h(x, z)k(y, z)$ for some functions h and k, then $X \perp\!\!\!\perp Y \mid Z$.

(b) Recall that the probability density function of the multivariate Gaussian distribution is given by

$$q(x_1, \ldots, x_p) = \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2} x^T \Theta x\right).$$

Using part (a), show that for any $j \neq k$ with $\Theta_{jk} = 0$, we have $X_j \perp\!\!\!\perp X_k \mid X_{\setminus\{j,k\}}$.

Problem 4. For $X \sim N(0, \Theta^{-1})$, recall the notation

$$X_j = \theta_j^T X_{\setminus\{j\}} + W_j,$$

for each $1 \leq j \leq p$, where θ_j is the best linear predictor for X_j based on $X_{\setminus\{j\}}$. (Hence, $W_j \in \mathbb{R}$ is normally distributed with mean 0, and $W_j \perp\!\!\!\perp X_{\setminus\{j\}}$.)

If we partition the matrix $\Sigma = \Theta^{-1} = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,\setminus\{1\}} \\ \Sigma_{\setminus\{1\},1} & \Sigma_{\setminus\{1\},\setminus\{1\}} \end{pmatrix}$, show that the first column of Θ takes the form $\begin{pmatrix} a_1 \\ -a_1\theta_1 \end{pmatrix}$, for some $a_1 \in \mathbb{R}$ (an analogous statement holds for each value of j). In particular, recovering $\mathrm{supp}(\theta_j)$, for each $1 \leq j \leq p$, allows us to recover $\mathrm{supp}(\Theta)$.

Problem 5. Consider a matrix $A \in \mathbb{R}^{m \times n}$.

(a) Prove that a vector v is an eigenvector of $A^T A$ with eigenvalue λ_1 if and only if v is a right singular vector of A with singular value $\sigma_1 = \sqrt{\lambda_1}$.

(b) Prove that a vector $v_1 \in \mathbb{R}^n$ is a right singular vector of A corresponding to the first singular value σ_1, if and only if

$$v_1 \in \arg\max_{\|v\|_2 = 1} \|Av\|_2^2.$$

Problem 6. Consider a rank-r matrix $A \in \mathbb{R}^{m \times n}$, and recall that for $k \leq r$, we define $A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$ to be the truncated SVD, where $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq 0$ are the ordered singular values.

(a) Prove that for any $1 \leq k \leq r - 1$, we have

$$\|A - A_k\|_2 = \sigma_{k+1}.$$

(b) Suppose $B \in \mathbb{R}^{m \times n}$ has $\operatorname{rank}(B) \leq k$. Prove that $\|A - B\|_2 \geq \sigma_{k+1}$. Hence, conclude the Eckart-Young-Mirsky Theorem:

$$\min_{B \in \mathbb{R}^{m \times n} : \operatorname{rank}(B) \leq k} \|A - B\|_2 = \|A - A_k\|_2 \,,$$

for all $1 \leq k \leq r$.

Problem 7.

(a) Suppose a distribution D over matrices in $\mathbb{R}^{m \times n}$ satisfies the $\left(\frac{\epsilon}{2}, \delta\right)$-distributional JL property. Show that for any $S \subseteq \{1, \ldots, n\}$ such that $|S| = k$, a randomly drawn matrix $\Phi \sim D$ satisfies

$$(1 - \epsilon)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \epsilon)\|x\|_2^2, \quad \forall x \in \mathbb{R}^S, \tag{1}$$

with probability at least $1 - \delta \left(\frac{c}{\epsilon}\right)^k$.

(b) Use the result in part (a) to conclude the following theorem:

Theorem (JL $\implies$ RIP)**.** *Suppose $\epsilon < 1$, and $m \geq c_1(\epsilon) k \log\left(\frac{n}{k}\right)$. If D satisfies the $\left(\frac{\epsilon}{2}, \delta\right)$-distributional JL property with $\delta = e^{-m\epsilon}$, then with probability at least $1 - e^{-\epsilon m/2}$, a randomly drawn matrix $\Phi \sim D$ satisfies (ϵ, k)-RIP.*

Problem 8. Consider a weighted graph $G = (V, E)$, and recall that the Laplacian is defined by $L = D - W$.

(a) Prove that the second smallest eigenvalue of L satisfies $\lambda_2 > 0$ if and only if G is connected.

(b) Prove that the Laplacian L_{K_n} of the complete graph K_n has eigenvalue 0 with multiplicity 1, and eigenvalue n with multiplicity $n - 1$.

(c) Prove that the Laplacian L_{S_n} of the star graph S_n, with edge set $\{(1, u) : 2 \leq u \leq n\}$, has eigenvalue 0 with multiplicity 1, eigenvalue 1 with multiplicity $n - 2$, and eigenvalue n with multiplicity 1.

Problem 9. Recall that a stochastic block model on K communities is defined with respect to a symmetric matrix $B \in \mathbb{R}^{K \times K}$ of connection probabilities between communities. Prove the following result about the expected adjacency matrix $P = \mathbb{E}[A]$:

Lemma. *Suppose B is full-rank. Let $V_0 D_0 V_0^T = P$ be a spectral decomposition, where $V_0 \in \mathbb{R}^{n \times K}$ and $D_0 \in \mathbb{R}^{K \times K}$. Then $V_0 = \Theta X$, where $\Theta \in \mathbb{R}^{n \times K}$ is the membership matrix of the communities (with a 1 in the column corresponding to the community of node i, and all other entries in row i equal to 0), and $X \in \mathbb{R}^{K \times K}$, with $\|X_{i,:} - X_{j,:}\|_2 = \sqrt{\frac{2K}{n}}$, for $i \neq j$, where $\{X_{i,:}\}_{i=1}^K$ denote the rows of X.*