Fall 2025 Biostat 915/Stat 915/Compsci 762 Homework 1

Option A

Due: 9/30 (Tuesday) to the TA's email (zhengchi.ma@duke.edu)

1. (a) Consider the spiked covariance matrix model

$$Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu, \Sigma).$$

Here, $Y_1, \ldots, Y_n \in \mathbb{R}^p$, $\Sigma = U\Lambda U^\top + \sigma^2 I_p$ is the eigenvalue decomposition; U is an unknown p-by-r matrix with orthonormal columns; Λ is an unknown matrix with non-increasing diagonal entries; and $\sigma^2 > 0$ is the unknown noise level. Consider the principal component analysis on Y and suppose $\hat{U}$ are the leading r PC vectors. Show that $\hat{U}$ is the maximum likelihood estimator of U.

(b) Derive a explicit formula for the maximum likelihood estimates of (U, Λ, σ^2).

(c) Consider the matrix denoising model $Y = X + Z$. Here, Y, X, Z are p_1-by-p_2 matrices, X is a fixed parameter, Z is random and has i.i.d. Gaussian $N(0, \sigma^2)$ entries with unknown variance parameter $\sigma^2 > 0$. Suppose $X = U\Lambda V^\top$ is the singular value decomposition, where U is a p_1-by-r matrix and V is a p_2-by-r matrix, both with orthonormal columns; Λ is diagonal with non-increasing singular values of X. Suppose $\hat{U}, \hat{V}$ are the leading r left and right singular vectors of Y. Show that $\hat{U}, \hat{V}$ are the MLE estimators for U and V, respectively.

2. Suppose A is a positive definite matrix such that both $\mathbb{E}(A)$ and $\mathbb{E}(A^{-1})$ exist. Prove that the matrix $\mathbb{E}(A^{-1}) - [\mathbb{E}(A)]^{-1}$ is a non-negative definite matrix, that is, $u^\top \left(\mathbb{E}(A^{-1}) - [\mathbb{E}(A)]^{-1}\right) u \geq 0$ for any vector u.

By this means, also show that $[\mathbb{E}A^{-1}]_{jj} \geq 1/[\mathbb{E}A_{jj}]$ for any index j. Does this conclusion hold if A is symmetric but not positive definite?

3. The sub-Weilbull distributions are widely used to model light-tailed or heavy-tailed distributed data. We say a random variable X satisfies sub-Weibull distribution with tail parameter θ, if for some positive numbers a, b, θ such that

$$\mathbb{P}(|X| \geq x) \leq a \exp\left(-bx^{1/\theta}\right), \quad \text{for all } x > 0.$$

Prove that the following conditions are all equivalent to the sub-Weibull distributions defined above.

(a) X satisfies

$$\exists K_1 > 0 \text{ such that } \mathbb{P}(|X| \geq x) \leq 2\exp\left(-(x/K_1)^{1/\theta}\right) \text{ for all } x \geq 0.$$

(b) The moment of X satisfy

$$\exists K_2 > 0 \quad \text{such that} \quad \|X\|_k \leq K_2 k^\theta \quad \text{for all } k \geq 1.$$

(c) The moment generating function of $|X|^{1/\theta}$ satisfies

$$\exists K_3 > 0 \quad \text{such that} \quad \mathbb{E}\left[\exp\left((\lambda|X|)^{1/\theta}\right)\right] \leq \exp\left((\lambda K_3)^{1/\theta}\right)$$

for all λ such that $0 < \lambda \leq 1/K_3$.

(d) The moment generating function of $|X|^{1/\theta}$ is bounded at some point, namely

$$\exists K_4 > 0 \quad \text{such that} \quad \mathbb{E}\left[\exp\left((|X|/K_4)^{1/\theta}\right)\right] \leq 2.$$

Also, show that Weibull distributions with $\theta = 1$ and $\theta = 1/2$ are the same as the sub-exponential and sub-Gaussian distributions, respectively.

4. Suppose $X_1, X_2, \ldots, X_n$ be a sequence of sub-Gaussian random variables such that

$$\exists K > 0 \text{ such that } \mathbb{P}(|X_i| \geq x) \leq 2\exp\left(-(x/K)^2\right) \text{ for all } x \geq 0,$$

where K is independent of n. $\{X_i, i = 1, \ldots, n\}$ are not necessarily independent.

(a) Show that for every $n \geq 2$,

$$\mathbb{E}\max_{i \leq n} |X_i| \leq C\sqrt{\log n},$$

where C is a constant independent of n.

(b) Show that the above bound is sharp. Let $X_1, \ldots, X_n$, be independent $N(0,1)$ random variables. Prove that there exists a constant c independent of n such that

$$\mathbb{E}\max_{i \leq n} X_i \geq c\sqrt{\log n}.$$

5. Consider the Lasso problem

$$\min_{\beta} \frac{1}{2n}\|\mathbf{Y} - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_1,$$

where $\lambda > 0$ is a tuning parameter.

(a) If $\hat{\beta}$ and $\hat{\beta}'$ are both minimizers of the Lasso problem, show that they have the same prediction, i.e., $\mathbf{X}\hat{\beta} = \mathbf{X}\hat{\beta}'$.

(b) Let $\hat{\beta}$ be a minimizer of the Lasso problem with jth component $\hat{\beta}_j$. Denote X_j to be the jth column of $\mathbf{X}$. Show that

$$\lambda = \begin{cases} n^{-1}X_j^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) & \text{if } \hat{\beta}_j > 0, \\ -n^{-1}X_j^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) & \text{if } \hat{\beta}_j < 0, \\ \lambda \geq n^{-1}|X_j^T(\mathbf{Y} - \mathbf{X}\hat{\beta})| & \text{if } \hat{\beta}_j = 0. \end{cases}$$

(c) If $\lambda > \|n^{-1}\mathbf{X}^\top\mathbf{Y}\|_\infty$, where $\|\cdot\|_\infty$ denotes the maximum of a vector. Prove that $\hat{\beta}_\lambda = 0$, where $\hat{\beta}_\lambda$ is the minimizer of the Lasso problem with regularization parameter λ.

6. Suppose $\mathbf{\Sigma}$ is the covariance matrix of a set of p variables X, and $\mathbf{A}_{ab}$ indicates the submatrix of $\mathbf{A}$ in the indexes of row or column $a \subset \{1, \cdots, p\}$ and $b \subset \{1, \cdots, p\}$. Consider the partial covariance matrix $\mathbf{\Sigma}_{a.b} = \mathbf{\Sigma}_{aa} - \mathbf{\Sigma}_{ab}\mathbf{\Sigma}_{bb}^{-1}\mathbf{\Sigma}_{ba}$ between the two subsets of variables $X_a = (X_1, X_2)$ consisting of the first two, and X_b the rest. In the Gaussian distribution, $\mathbf{\Sigma}_{a.b}$ is the covariance matrix of the conditional distribution of $X_a \mid X_b$. Define $\rho_{jk|\text{ rest}}$ as the partial correlation between the jth and kth variables conditional on the rest, and $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$.

   (a) Show that $\mathbf{\Sigma}_{a.b} = \Theta_{aa}^{-1}$.

   (b) Show that if we treat $\mathbf{\Theta}$ as if it were a covariance matrix, and compute the corresponding "correlation" matrix
   $$\mathbf{R} = \text{diag}(\mathbf{\Theta})^{-1/2} \cdot \mathbf{\Theta} \cdot \text{diag}(\mathbf{\Theta})^{-1/2}$$
   then $\mathbf{R}_{jk} = -\rho_{jk|\text{ rest}}$.

7. Suppose X is a rank-r p_1-by-p_2 matrix. Partition X to
   $$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}.$$
   Suppose X_{11} is rank-r of shape k_1-by-k_2 ($k_1, k_2 \geq r$ but need not be equal). Prove that X_{22} can be uniquely identified as $X_{22} = X_{21}X_{11}^{\dagger}X_{12}$, where $(X_{11})^{\dagger}$ is the Moore-Penrose inverse of X_{11} (you may check the definition of Moore-Penrose inverse at Wikipedia: `https://en.wikipedia.org/wiki/Moore%E2%80%93Penrose_inverse`).