Statistical Modeling and Methods: Homework 1

Due by 5pm on Friday, March 13, online through Blackboard

Homework format: all homework must be written in latex. You must turn in both your tex and pdf files. Attach your code and computer output if there is any programming.

1. Let X be a $n \times p$ matrix of rank r, $r \leq p$. Denote $U_\ell = (u_1, \ldots, u_\ell)$ and $V_\ell = (v_1, \ldots, v_\ell)$ for any $\ell \leq r$, where $\{u_j, v_j\}$ is the jth pair of singular component vectors of X with the corresponding singular value d_j, $j = 1, \ldots r$, , i.e., $X = U_r D_r V_r'$, where $D_r = \mathrm{diag}(d_1, \ldots, d_r)$. Show that $X_\ell^* = U_\ell D_\ell V_\ell'$ is the best low rank approximation to X that minimizes $\mathrm{tr}\{(X - Y)(X - Y)'\}$ among all $N \times p$ matrices Y with $\mathrm{rank}(Y) \leq \ell$.

2. Let A be $m \times n$ and A^- is a generalized inverse of A, denote the column space of A by $C(A)$. Show the following facts:
   (a) $\mathrm{rank}(A) \leq \mathrm{rank}(A^-)$, $\mathrm{rank}(A) = \mathrm{rank}(AA^-) = \mathrm{rank}(A^-A)$;
   (b) $C(A) = C(AA'), C(A') = C(A'A)$;
   (c) $A'A(A'A)^-A' = A'$, $A(A'A)^-A'A = A$;
   (d) The matrix $A(A'A)^-A$ does not depend on the choice of the generalized inverse of $A'A$.

3. (a) Consider an $N \times N$ invertible matrix A and a vector $a \in \Re^N$, find explicit expressions for $|A + aa'|$ and $(A + aa')^{-1}$, where $|A|$ denotes the determinant of a square matrix A.
   (b) For two non-singular $p \times p$ matrices A and B, find an equivalent expression of $(A + B)^{-1}$ that contains A^{-1} and B^{-1}.

4. Suppose symmetric matrices A and B are both $(J \times J)$. Denote eigenvalues of A and B as $\{\lambda_j(A)\}$ and $\{\lambda_j(B)\}$, respectively. Please show that:

$$\sum_{j=1}^{J} \{\lambda_j(A) - \lambda_j(B)\}^2 \leq trace\{(A - B)(A - B)'\}$$

5. Consider matrix $X = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 2 & 1 \end{pmatrix}$,

   (a) Find the QR and singular value decomposition of X. What are the two corresponding bases for the column space $C(X)$?

   (b) Use the SVD of X to find the eigen-decomposition of $X^T X$. What are the eigenvalues and eigenvectors?

   (c) Find the best rank=1 and rank=2 approximations to X.

6. Consider a random sample $X_1, \ldots, X_N$ that are uniformly distributed in a unit ball in $\Re^p$, i.e., $\{x \in \Re^p : \|x\| \leq 1\}$.

   (a) Derive the median distance M from the origin to the closest data point. What are the median distances for a sample of size 10^6 and $p = 1, \ldots, 15$, respectively.

   (b) Derive the mean distance D from the origin to the closest data point. What are the mean distances for a sample of size 10^6 and $p = 1, \ldots, 15$, respectively.

7. Let $X \sim N_p(\mu, I_p)$, and A is a $p \times p$ symmetric matrix. Show the following:

   (a) $X'AX \sim \chi^2_r(\lambda)$ with $\lambda = \mu' A \mu$ if and only if A is an idempotent matrix of rank r.

   (b) BX and $X'AX$ are independent if and only if $BA = 0$, where B is a $q \times p$ matrix.

   (c) $X'AX$ and $X'BX$ are independent if and only if $AB = 0$, where B is a $p \times p$ symmetric matrix.

8. Let $X \sim N_p(\mu, I_p)$, and Q, Q_1, Q_2 are quadratic forms in X such that $Q = Q_1 + Q_2$. Assume that $Q \sim \chi^2_r(\lambda)$, $Q_1 \sim \chi^2_{r_1}(\lambda_1)$ and $Q_2 \geq 0$. Show that $Q_2 \sim \chi^2_{r-r_1}(\lambda - \lambda_1)$.