STP533: Applied Multivariate Analysis

Homework 1

Name: Your name; NetID: Your ID

Due 11:59pm Tuesday February 4 2025

Question 1

The world's 10 largest companies in 2000 yield the following data (in billions):

Company	x_1 =sales	x_2 =profits	x_3 =assets
Citigroup	108.28	17.05	1,484.10
General Electric	152.36	16.59	750.33
American Intl Group	95.04	10.91	766.42
Bank of America	65.45	14.14	1,110.46
HSBC Group	62.97	9.52	1,031.29
ExxonMobil	263.99	25.33	195.26
Royal Dutch/Shell	265.19	18.54	193.83
BP	285.06	15.73	191.11
ING Group	92.01	8.10	1,175.16
Toyota Motor	165.68	11.13	211.15

(a) Plot the pairwise scatter plot. Comment on the appearance between x_1 and x_2.

(b) Compute $\bar{\mathbf{X}} = [\bar{x}_1, \bar{x}_2, \bar{x_3}]$, $\mathbf{S} = [s_{jk}]_{3\times 3}$ and $R = [r_{jk}]_{3\times 3}$. Interpret r_{12}.

Question 2

Define the following distance from the point $P = (x_1, x_2)$ to the origin $O = (0, 0)$ as

$$d_q(P, O) = (|x_1/2|^q + |x_2|^q)^{\frac{1}{q}}$$

(a) Compute the general distance between $P = (1, 2)$ and $Q = (-1, 0)$, $d_q(P, Q) := d_q(P - Q, O)$ for $q = \frac{1}{2}$.

(b) Plot the contours of $d_q(P, O) = 1$ (varying P) for $q = 2, 1$ and $\frac{1}{2}$ on the same graph.

Question 3

Recall that the sample covariance can be represented as

$$\mathbf{S}_{p\times p} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X} - \bar{\mathbf{X}})^T(\mathbf{X} - \bar{\mathbf{X}}) = \frac{1}{n}\mathbf{X}^T(\mathbf{I}_n - \mathbf{J}_n)\mathbf{X}$$

where $\mathbf{J}_n$ is a $n \times n$ matrix with entries all being $\frac{1}{n}$.

Prove:

(a) $\mathbf{J}_n$ is an idempotent matrix, i.e. $\mathbf{J}_n^k = \mathbf{J}_n \cdots \mathbf{J}_n = \mathbf{J}_n$ for $k \in \mathbb{N}$.

(b) $\mathbf{I}_n - \mathbf{J}_n$ is also idempotent.

Question 4

Let $A_{n\times m}$ and $B_{m\times n}$ be two matrices. Prove $\mathrm{tr}(AB) = \mathrm{tr}(BA) = \mathbf{1}_n^T(A \circ B^T)\mathbf{1}_m$ where the Hadamard product between two matrices is by element-wise multiplication.

Note, the last equality indicates that the trace of matrices product can be efficiently computed using `sum`(A*`t`(B)) as in `R`, which takes only $\mathcal{O}(mn)$, rather than $\mathcal{O}(nm^2)$ or $\mathcal{O}(n^2m)$.

Question 5

Consider p-dimensional random vector $X \sim (\mu, \Sigma)$, i.e. $\mathrm{E}[X] = \mu$, $\mathrm{Cov}(X) = \Sigma$. Consider a symmetric matrix Λ and the corresponding random quadratic form $X^T\Lambda X$. Show that $\mathrm{E}[X^T\Lambda X] = \mu^T\Lambda\mu + \mathrm{tr}(\Lambda\Sigma)$.