ADVANCED STATISTICS

BAX 442 HW1

Group :
1) PRIYA IDDALGI
2) RAHUL RAJPUT

I) Linear Algebra Review

1)

We know that for a PSD matrix M,

$v^T M v \geq 0 \quad \forall\ v \in \mathbb{R}^n \quad \rightarrow (1)$

Let λ be an eigenvalue of M

Then,

there exists an eigenvector $v \in \mathbb{R}^n$, s.t.

$Mv = \lambda v \quad \rightarrow (2)$

Multiplying (2) by v^T, we get

$v^T M v = \lambda v^T v$

From (1), we know that $v^T M v \geq 0$

Thus, $\lambda v^T v \geq 0$

Since $v^T v$ is positive for all v, it implies that λ is non-negative.

Hence, Proved.

2)

Given : $M_{ij} = -\rho \quad \forall\ i \neq j$

$M_{ij} = 1 \quad \forall\ i = j$

A ~~positive~~ correlation matrix is a square matrix with all diagonal elements as 1, and all other elements lie between 1 and -1.

Thus $\rho \leq 1$ for M to be a valid correlation matrix.

3)

i) Consider 'n' independent random variables

$Z = [Z_1, Z_2 \ldots\ldots Z_n]$

ii) A correlation matrix can be decomposed using Cholesky Decomposition, where

$C = V^T V$

where, V = lower triangular matrix.

iii) Compute matrix V, where,

Diagonal elements of V:

$$V_{j,j} = \sqrt{C_{i,j} - \sum_{K=0}^{j-1} (V_{j,K})^2}$$

and non-diagonal elements of V:

$$V_{i,j} = \frac{C_{ij} - \sum_{K=1}^{j-1} (V_{j,K})^2}{V_{j,j}}$$

iv) This ~~comp~~ recursive substitution can be easily done by in-built functions (. linalg.cholesky() in python)

v) Compute correlated random variables

$X = VZ$

The random variables X will have the desired correlation matrix C.

4)

a) False.

Correlation is a measure of "linear" association.

Counterexample 1 :

→ Consider the relation between x and y are related through a polynomial or trigonometric relationship

$y = x^3$ or $y = \sin x$

Even though y is strongly dependent on x, their correlation is close to zero.

→ Consider temperature and icecream sales. They're dependent, but ~~[illegible]~~ not linearly, so correlation might be 0.

b) False.

Just because 2 variables are related, does not mean one causes the other.

Counterexample:

Height and weight of people are highly correlated, but one doesn't cause the other, they are simply related

- It is important to look at other factors for causal analysis.

c) False.

A correlation matrix gets affected by scaling X.

Counterexample:

If the original random variable X is ~~changed~~ centered, the mean of each variable will change, and so will C.

~~let~~ If X is changed to X_centered, the new correlation matrix will be

$$C\_centered = \frac{X_{centered}^T X\_centered}{(n-1)}, \quad n = \text{no. of obs.}$$

Similarly, if X is scaled by a factor K, the new correlation matrix will change as standard deviation of each variable will change.

$$C\_scaled = \frac{(KX)^T KX}{n-1} = K^2 C$$

New correlation matrix will change by a factor of K^2.

Vector Calculus

1) $f(x) = \frac{1}{2} x^T A x + b^T x$, A: symmetric
$b \in R^n$

$$\frac{\partial f(x)}{\partial x} = \frac{\partial}{\partial x}\left(\frac{1}{2} x^T A x\right) + \frac{\partial}{\partial x}\left(b^T x\right)$$

$$= \frac{1}{2}\frac{\partial}{\partial x}\left(x^T A x\right) + \frac{\partial}{\partial x}\left(b^T x\right)$$

$$= \frac{1}{2}\frac{\partial}{\partial x}\left(x^T A\right).(x) + b^T \frac{\partial}{\partial x}(x)$$

$$= \frac{1}{2} A^T x^{T^T} + \frac{1}{2} x^T A + b^T$$

$A^T = A$, A is symmetric

$\Rightarrow \quad x^T A + b^T$

or, $\quad Ax + b \quad$ ($A^T = A$, A is symmetric)

$\Rightarrow \quad \nabla f(x) = Ax + b$

& $\nabla^2 f(x) = \frac{\partial}{\partial x}(Ax + b)$

$= A$

2) $f(u,\lambda) = u^T X^T X u - \lambda u^T u$

$$\frac{\partial f}{\partial u} = \frac{\partial}{\partial u}\left(u^T X^T X u\right) - \frac{\partial}{\partial u}\left(\lambda u^T u\right) = 0$$

$$= \frac{\partial}{\partial u}\left(u^T X^T X u\right) - \lambda \frac{\partial}{\partial u}\left(u^T u\right) = 0$$

$$= X^T X \frac{\partial}{\partial u}\left(u^T u\right) - \lambda \frac{\partial}{\partial u}\left(u^T u\right) = 0$$

$$\Rightarrow 2X^T X u = \lambda 2u \qquad \text{(Product Rule for } U^T x\, u\text{)}$$

$$\Rightarrow \cancel{2}\left(X^T X\right) u = \left(\cancel{2}\lambda\right) u$$

$$\underline{\left(X^T X\right)} u = \underline{(\lambda)} u \quad — ①$$

In Equation 1. we see a relationship similar to a matrix & eigenvectors.

$$\underline{Ax = \lambda x}$$

$\therefore$ u is the eigenvector for matrix $X^T X$,
λ is the eigen value.

II) Simple Linear Regression

1)

Prove : Regression line passes through $(\bar{x}, \bar{y})$

Let $Y = a + bX$

The regression line minimizes the least square sum residuals, i.e.,

$$\sum_{i=1}^{n} (Y_i - (a+bX_i))^2 \text{ is minimum.} \longrightarrow ①$$

We know that a minimum occurs when $\frac{\partial}{\partial x}$ & $\frac{\partial}{\partial y} = 0$

Partially differentiating ① w.r.t x and Y we get,

$$\sum_{i=1}^{n} (Y_i - (a+bX_i)) = 0 \quad \& \quad \sum_{i=1}^{n} X_i (Y_i - (a+bX_i)) = 0$$

The first equation can be simplified to

$$\sum_{i=1}^{n} Y_i = a + b \sum_{i=1}^{n} X_i$$

Dividing by n throughout

$$\bar{Y} = a + b\bar{X}$$

Hence, Proved,

2)

Prove : $R^2 = r_{xy}^2$, where $r_{xy} = \text{corr.}(x,y)$

Assume a simple regression model

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Consider estimation using ordinary least squares.

OLS estimates for simple linear regression :

$$\hat{\beta_0} = \bar{y} - \hat{\beta_1}\bar{x}$$

$$\hat{\beta_1} = \frac{S_{xy}}{S_x^2} \longrightarrow ①$$

R^2 = proportion of variation explained in the data

$$= \frac{\text{Explained sum of sq.}}{\text{Total } \cancel{\text{sym}} \text{ sum of sq.}} = \frac{ESS}{TSS}$$

then,

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y_i} - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{\sum_{n=1}^{n} (\hat{\beta_0} + \hat{\beta_1} x_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Applying ②, ①, we get,

$$R^2 = \frac{\sum_{i=1}^{n} (\bar{y} - \hat{\beta_1}\bar{x} + \hat{\beta_1} x_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n} (\hat{\beta_1}(x_i - \bar{x}))^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$= \hat{\beta_1}^2 \frac{\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2}{\frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2} = \hat{\beta_1}^2 \frac{S_x^2}{S_y^2}$$

$$= \left(\frac{S_x}{S_y}\hat{\beta_1}\right)^2 = r_{xy}^2$$

Thus, $R^2 = r_{xy}^2$