Assignment - 5

AI22MTECH02003 - Shrey Satapara

Q-57 (June 2018) The covariance matrix of four dimensional random vector X is of the form $\begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$, Where $\rho < 0$

If v is the variance of the first principle component then,

1. v cannot exceed 5/4
2. v can exceed 5/4 but cannot exceed 4/3
3. v can exceed 4/3 but cannot exceed 3/2
4. v can exceed 3/2

covariance matrix: In probability theory and statistics, a covariance matrix is a square matrix giving the covariance between each pair of elements of a given random vector

Solution Here, co-variance matrix is given and we've to find maximum value of variance of first principle component. For that, first we have to find principle components. And for that we have to find eigenvalues of covariance matrix.

By calculations we get 4 eigen values $1 - \rho,\ 1 - \rho,\ 1 - \rho,\ 3\rho + 1$ which will be the variance of their respective principle component,

here $\rho > 0$ means $(1 - \rho) > (3\rho + 1)$

So, order of principle components is $1 - \rho,\ 1 - \rho,\ 1 - \rho,\ 3\rho + 1$

here first 3 principle components are same so their variance will be same, let's assume $x = 1 - \rho \; and \; y = 3\rho + 1$ and $x, y \geq 0$ because co-variance matrix is positive semi definite. Also note that variance of data remains same after applying linear transformation through eigan vectors. Hence,

$$\begin{aligned} 3x + y &= 4 \\ \therefore 3x &< 4 \;\; (becausey > 0) \qquad (1) \\ \therefore x &< 4/3 \end{aligned}$$

Means x cannot exceed 4/3 which is option 2

1 Proof that Variance of linearly transformed space is eigan values of principle components

Here in this problem 4 dimensional random vector is given. Let's assume those random vector are X_1, X_2, X_3, X_4 so the matrix will be $\begin{bmatrix} X_1 & X_2 & X_3 & X_4 \end{bmatrix}$

So here, co-variance matrix will be

$$\begin{bmatrix} var(X_1) & cov(X_1, X_2) & cov(X_1, X_3) & cov(X_1, X_4) \\ cov(X_2, X_1) & var(X_2) & cov(X_2, X_3) & cov(X_2, X_4) \\ cov(X_3, X_1) & cov(X_3, X_2) & var(X_3) & cov(X_3, X_4) \\ cov(X_4, X_1) & cov(X_4, X_2) & cov(X_4, X_3) & var(X_4) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix} \qquad \text{—— Eq(A)}$$

now by calculations we found out eigenvectors of all the eigen values and those eigen vectors are $\begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, for $1 - \rho, 1 - \rho, 1 - \rho, 3\rho + 1$.

now since we want to find out variance for first principle component we'll apply linear transformation using eigen vector where eigen value is maximum which is $1 - \rho$ in this case. and respective eigen vector is ev = $\begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$.

In PCA, before applying linear transformation to data first we convert respective eigen vector to unit length vector. so unit eigen vector for ev is

$$uev = \frac{1}{||X||} * ev \tag{2}$$

$$= \frac{1}{\sqrt{(-1)^2 + 0 + 0 + (1)^2}} * \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{3}$$

$$= \frac{1}{\sqrt{2}} * \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ 0 \\ 0 \\ \frac{1}{2} \end{bmatrix} \tag{5}$$

Now let's apply linear transformation on data.

$$\begin{bmatrix} X_1 & X_2 & X_3 & X_4 \end{bmatrix} * \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{-X_1}{\sqrt{2}} + \frac{X_4}{\sqrt{2}} \end{bmatrix}$$

so here we wants to find variance of linearly transformed space.

$$V(\frac{-X_1}{\sqrt{2}} + \frac{X_4}{\sqrt{2}})$$
$$= V(\frac{-X_1}{\sqrt{2}}) + V(\frac{X_4}{\sqrt{2}})$$
$$- 2cov(\frac{-X_1}{\sqrt{2}}, \frac{X_4}{\sqrt{2}})$$
$$= V(\frac{X_1}{\sqrt{2}}) + V(\frac{X_4}{\sqrt{2}}) -$$
$$2cov(\frac{X_1}{\sqrt{2}}, \frac{X_4}{\sqrt{2}})$$
$$(\because V(-X) = V(X))$$
$$= \frac{var(X_1)}{2} + \frac{var(X_4)}{2}$$
$$-\frac{2cov(X_1, X_4)}{\sqrt{2}\sqrt{2}}$$
$$(\because var(cX) = c^2 var(X)$$
$$cov(aX, bY) = ab * cov(X, Y))$$
$$= \frac{1}{2} + \frac{1}{2} - p$$
$$(from\ Eq(A))$$
$$= 1 - p$$

which is the eigen value of first principal component.