Γραμμική Αλγεβρα ΙΙ

Tutorial 6
Τμήμα Στατιστικής
ΟΠΑ

20 Απριλίου 2018

1. Έστω 2 τυχαίες μετάβλητες που δίνονται απο τον παρακάτω πίνακα.

		X		
		4	5	
Y	3	0.1	0.3	0.4
	7	0.4	0.2	0.6
		0.5	0.5	1

Να βρέθει ο πίνακας συνδιάκυμανσης (Variance- covariance matrix) Σ.
Λύση:
Έχουμε οτι

$$Var(X) = \mathbb{E}(X^2) - \Big(\mathbb{E}(X)\Big)^2 \quad \text{και} \quad Var(Y) = \mathbb{E}(Y^2) - \Big(\mathbb{E}(Y)\Big)^2$$

Απο τον πίνακα που μας δόθηκε προκύπτει οτι για την X έχουμε

$$\mathbb{E}(X) = 4 \cdot 0.5 + 5 \cdot 0.5 = 4.5 \quad \text{και} \quad \mathbb{E}(X^2) = 4^2 \cdot 0.5 + 5^2 \cdot 0.5 = 20.5$$

Αρα $Var(X) = 20.5 - 4.5^2 = 0.25$. Ομοια για την Y έχουμε

$$\mathbb{E}(Y) = 3 \cdot 0.4 + 7 \cdot 0.6 = 5.4 \quad \text{και} \quad \mathbb{E}(Y^2) = 3^2 \cdot 0.4 + 7^2 \cdot 0.6 = 33$$

Αρα $Var(Y) = 33 - 5.4^2 = 3.84$. Για την συνδιάκυμανση των 2 τυχαίων μετάβλητων γνώριζουμε οτι

$$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

οπότε για να την υπολογίσουμε θα πρέπει να βρούμε την τιμή της ποσότητας $\mathbb{E}(XY)$. Εχουμε λοιπόν οτι

$$\mathbb{E}(XY) = 4 \cdot 3 \cdot 0.1 + 5 \cdot 3 \cdot 0.3 + 4 \cdot 7 \cdot 0.4 + 5 \cdot 7 \cdot 0.2 = 23.9$$

Αρα εν τέλει $Cov(X,Y) = 23.9 - 4.5 \cdot 5.4 = 0.4$. Ετσι ο πίνακας συνδίανυμανσης θα είναι

$$\Sigma = \begin{bmatrix} 0.25 & -0.4 \\ -0.4 & 3.84 \end{bmatrix}$$

$\triangleleft$

2. Εστω ο πίνακας

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}.$$

Να γίνει η διάσπαση ιδιόμορφων τιμών (SVD).
Λύση:

Έχουμε οτι

$$AA^\top = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$$

με ιδιοτιμές $\lambda_1 = 25$ και $\lambda_2 = 9$ και αντίστοιχα ιδιοδιανύσματα τα

$$\boldsymbol{u_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \boldsymbol{u_2} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

τα οπόια διαιρώντας τα με την νόρμα τους γίνονται

$$\hat{\boldsymbol{u_1}} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \hat{\boldsymbol{u_2}} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Ετσι ο πίνακας U της διάσπασης είναι

$$U = \begin{bmatrix} 1/\sqrt{2} & -1\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Ευκολα μπορούμε να καταλάβουμε οτι ο πίνακας με τις ιδιόμορφες τιμές είναι ο

$$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}.$$

Τέλος, ας βρούμε τον πίνακα V της διάσπασης. Έχουμε ότι

$$A^\top A = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix}$$

με ιδιοτιμές τις $\lambda_1 = 25, \lambda_2 = 9, \lambda_3 = 0$ και αντίστοιχα ιδιοδιάνυσματα τα

$$\boldsymbol{v_1} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{v_2} = \begin{bmatrix} -1 \\ 1 \\ 4 \end{bmatrix} \quad mbox \quad kai \quad \boldsymbol{v_3} = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$$

τα οποία κανονικοποιώντας τα παιρνούμε τις στήλες του V. Εν τέλει λοιπόν καταλήγουμε στην εξής διάσπαση ιδιάζουσων τιμών:

$$A = U\Sigma V^\top \Rightarrow A = \begin{bmatrix} 1/\sqrt{2} & -1\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{18} & 1/\sqrt{18} & -4/\sqrt{18} \\ -2/3 & 2/3 & 1/3 \end{bmatrix}$$

$\triangleleft$

3. Μας δίνόνται τα παρακάτω δεδομένα.

Βάρος (lb)	120	125	125	135	145
Υψος (inch)	61	60	64	68	72

Να βρεθεί ο πίνακας διακυμανσης - συνδιακύμανσης.

Λυση:
Ευκολα μπορούμε να υπολογίσουμε το διάνυσμα των μέσων ως εξης

$$M = \frac{1}{5} \begin{bmatrix} \sum x_i \\ \sum y_i \end{bmatrix} = \begin{bmatrix} 130 \\ 65 \end{bmatrix}$$

Μετά αφαιρώντας τον μέσο απο κάθε x_i έχουμε

$$\hat{x_1} = x_1 - M = \begin{bmatrix} -10 \\ -4 \end{bmatrix}, \quad \hat{x_2} = x_2 - M = \begin{bmatrix} -5 \\ -5 \end{bmatrix}$$

$$\hat{x_3} = x_3 - M = \begin{bmatrix} -5 \\ -1 \end{bmatrix}, \quad \hat{x_4} = x_4 - M = \begin{bmatrix} 5 \\ 3 \end{bmatrix}, \quad \hat{x_5} = x_5 - M = \begin{bmatrix} 15 \\ 7 \end{bmatrix},$$

Αρα τα δεδομένα μας γίνονται ως εξης

$$B = \begin{bmatrix} 10 & -5 & -5 & 5 & 15 \\ -4 & -5 & -1 & 3 & 7 \end{bmatrix}.$$

Απ' τον πίνακα B προκύπτει ο πίνακας διακύμανσης ως εξής

$$\Sigma = \frac{1}{N-1} B B^\top = \begin{bmatrix} 100 & 47.5 \\ 47.5 & 25 \end{bmatrix}$$

οπου παρατηρούμε οτι η συνόλικη διακύμανση είναι ίση με 125. Αρα η μετάβλητη βάρος έχουμε οτι περιγράφει το $\frac{100}{125}\%$ της πληροφορίας που υπάρχει στα δεδομένα ενω η μετάβλητη υψος περιγράφει το $\frac{25}{125}\%$.