

ใบงาน : การลดมิติ ข้อมูล

รายวิชา : ENGCE207 Advanced Topics in computer Engineering
หัวข้อ Dimensionality Reduction

ชื่อ-สกุล นาย ปฐม ศรีสวัสดิ์ รัชชบัณฑิต 68143206017-3

โจทย์ ปัญหา

กำหนด ชุด ข้อมูล 2 มิติ ซึ่งประกอบด้วย จุด ข้อมูล 4 จุด ดังนี้

จุดข้อมูล	X	Y
A	2	1
B	4	3
C	5	5
D	7	5

ขั้นตอนที่ 1 หาค่าเฉลี่ย

$$X = (2+4+5+7)/4 = 18/4 = 4.5$$

$$Y = (1+3+5+5)/4 = 14/4 = 3.5$$

ปรับข้อมูล

$P(x,y)$	X	Y
A'	$2-4.5 = -2.5$	$1-3.5 = -2.5$
B'	$4-4.5 = -0.5$	$3-3.5 = -0.5$
C'	$5-4.5 = 0.5$	$5-3.5 = 1.5$
D'	$7-4.5 = 2.5$	$5-3.5 = 1.5$

ขั้นตอนที่ 2 คำนวณ Covariance Matrix (C)

$$\begin{aligned} \text{Cov}(X_1, X_1) &= \frac{(2-4.5)(2-4.5) + (4-4.5)(4-4.5) + (5-4.5)(5-4.5) + (7-4.5)(7-4.5)}{4-1} \\ &= \frac{(-2.5)(-2.5) + (-0.5)(-0.5) + (0.5)(0.5) + (2.5)(2.5)}{4-1} = \frac{13}{3} \approx 4.33 \end{aligned}$$

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \frac{(2-4.5)(1-3.5) + (4-4.5)(3-3.5) + (5-4.5)(5-3.5) + (7-4.5)(5-3.5)}{4-1} \\ &= \frac{(-2.5)(-2.5) + (-0.5)(-0.5) + (0.5)(1.5) + (2.5)(1.5)}{4-1} = \frac{11}{3} \approx 3.66 \end{aligned}$$

$$\text{Cov}(X_2, X_1) = \frac{(1-3.5)(2-4.5) + (3-3.5)(4-4.5) + (5-3.5)(5-4.5) + (5-3.5)(7-4.5)}{4-1}$$

$$= \frac{(-2.5)(-2.5) + (-0.5)(-0.5) + (1.5)(0.5) + (1.5)(2.5)}{4-1} = \frac{11}{3} \approx 3.66$$

$$\text{Cov}(X_2, X_2) = \frac{(1-3.5)(1-3.5) + (3-3.5)(3-3.5) + (5-3.5)(5-3.5) + (5-3.5)(5-3.5)}{4-1}$$

$$= \frac{(-2.5)(-2.5) + (-0.5)(-0.5) + (1.5)(1.5) + (1.5)(1.5)}{4-1} = \frac{11}{3} \approx 3.66$$

$$\text{Covariance Matrix} = \begin{bmatrix} 4.33 & 3.66 \\ 3.66 & 3.66 \end{bmatrix}$$

ขั้นตอนที่ 3 คำนวณ Eigenvalues (λ)

$$\det \left[\begin{pmatrix} 4.33 & 3.66 \\ 3.66 & 3.66 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] = 0$$

$$(4.33 - \lambda)(3.66 - \lambda) - 13.39 = 0$$

$$\text{กระจายพจน์ } (4.33 - \lambda)(3.66 - \lambda)$$

$$= 4.33 \times 3.66 - 4.33 - 3.66\lambda + \lambda^2$$

$$\text{คูณเลข } 4.33 \times 3.66 = 15.8478$$

$$\text{ดังนั้น } = \lambda^2 - (4.33 + 3.66)\lambda + 15.8478$$

$$= \lambda^2 - 7.99\lambda + 15.8478$$

$$\text{เพื่อหาจุดตัดเส้น } \lambda^2 - 7.99\lambda + 15.8478 - 13.39 = 0$$

$$\lambda^2 - 7.99\lambda + 2.4578 = 0$$

ใช้สูตรกำลังสอง

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \quad a=1, b=-7.99, c=2.4578$$

$$\lambda = \frac{7.99 \pm \sqrt{(-7.99)^2 - 4(1)(2.4578)}}{2}$$

$$= \frac{7.99 \pm \sqrt{63.8401 - 9.8312}}{2}$$

$$= \frac{7.99 \pm \sqrt{54.0089}}{2}$$

$$= \frac{7.99 \pm 7.349}{2}$$

$$\det \left[\begin{pmatrix} 4.33 & 3.66 \\ 3.66 & 3.66 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right] = 0$$

$$\det \left[\begin{pmatrix} 4.33 - \lambda & 3.66 \\ 3.66 & 3.66 - \lambda \end{pmatrix} \right] = 0$$

$(3.66)(3.66) = 13.39$
 $(4.33 - \lambda)(3.66 - \lambda)$

ค่าเฉลี่ยและค่าความแปรปรวนของ

$$\lambda^1 = \frac{7.99 + 7.349}{2} = \frac{15.339}{2} = 7.67$$

$$\lambda^2 = \frac{7.99 - 7.349}{2} = \frac{0.641}{2} \approx 0.32$$

ดังนั้นค่าลักษณะเฉพาะของ Σ คือ Eigen values: $\lambda^1 = 7.67$ (Principal Eigen value)
 $\lambda^2 = 0.32$

หา Eigen vector สำหรับ Eigen value ที่ใหญ่ที่สุด ($\lambda^1 \approx 7.67$)

$$\left[\begin{pmatrix} 4.33 & 3.66 \\ 3.66 & 3.66 \end{pmatrix} - \lambda^1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{pmatrix} -3.34 & 3.66 \\ 3.66 & -4.01 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\left[\begin{pmatrix} 4.33 & 3.66 \\ 3.66 & 3.66 \end{pmatrix} - 7.67 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$= \begin{pmatrix} -3.34x + 3.66y \\ 3.66x - 4.01y \end{pmatrix} = 0$$

$$\left[\begin{pmatrix} 4.33 & 3.66 \\ 3.66 & 3.66 \end{pmatrix} - \begin{pmatrix} 7.67 & 0 \\ 0 & 7.67 \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$-3.34x + 3.66y = 0 \quad 3.66x + 4.01y = 0$$

$$\left[\begin{pmatrix} 4.33 - 7.67 & 3.66 \\ 3.66 & 3.66 - 7.67 \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = 0 \quad \nearrow$$

$$y = \frac{3.34}{3.66}x \approx 0.91x \quad y = \frac{3.66}{4.01}x \approx 0.91x$$

$$y = 0.91x$$

จาก $y = 0.91x$ เลือก $x = 1$ จะได้ $V = [1, 0.91]$

ใช้ทฤษฎีพีทาโกรัส $\|V\| = \sqrt{x^2 + y^2}$ จะได้ $\|V\| = \sqrt{1^2 + 0.91^2} = \sqrt{1 + 0.8281}$
 $= \sqrt{1.8281} \approx 1.352$

ทำให้เป็นเวกเตอร์หนึ่งหน่วย $\hat{u} = V / \|V\|$

- $x_{\text{unit}} = 1 / 1.352 \approx 0.739$ หรือจะได้ว่า $V_1 (PC1) \approx [0.74, 0.67]$

- $y_{\text{unit}} = 0.91 / 1.352 \approx 0.673$

ขั้นตอนที่ 4 & 5: เลือก PC1 และแปลงข้อมูล

เลือก PC1 = $[0.74, 0.67] \times$ ข้อมูลที่ปรับแล้ว

$$A'' = (-2.5)(0.74) + (-2.5)(0.67) \approx -3.53$$

$$B'' = (-0.5)(0.74) + (-0.5)(0.67) \approx -0.71$$

$$C'' = (0.5)(0.74) + (1.5)(0.67) \approx 1.39$$

$$D'' = (2.5)(0.74) + (1.5)(0.67) \approx 2.86$$

ผลลัพธ์ จากข้อมูล 2D (x,y) เหลือเพียง 1D

$$A'' = -3.53, B'' = -0.71, C'' = 1.39, D'' = 2.86$$

โจทย์ 2 ข้อมูลเพิ่มเพิ่ม

จากชุดข้อมูล 2 ชุดที่กำหนดให้ ซึ่งประกอบด้วยข้อมูล 4 ชุด ดังตาราง:

ชุดข้อมูล	X	Y
E	1	5
F	2	3
G	4	2
H	5	1

หาค่าเฉลี่ย ข้อมูล

$$\bar{X} = (1+2+4+5)/4 \approx 12/4 \approx 3$$

$$\bar{Y} = (5+3+2+1)/4 \approx 11/4 \approx 2.75$$

ปรับข้อมูล

ชุดข้อมูล	X	Y
E'	$1-3 = -2$	$5-2.75 = 2.25$
F'	$2-3 = -1$	$3-2.75 = 0.25$
G'	$4-3 = 1$	$2-2.75 = -0.75$
H'	$5-3 = 2$	$1-2.75 = -1.75$

ขั้นตอนที่ 2 คำนวณ Covariance Matrix (C)

$$\text{Cov}(X_1, X_1) = \frac{(1-3)(1-3) + (2-3)(2-3) + (4-3)(4-3) + (5-3)(5-3)}{4-1}$$

$$= \frac{(-2)(-2) + (-1)(-1) + (1)(1) + (2)(2)}{4-1} \approx \frac{10}{3} \approx 3.33$$

$$\text{Cov}(X_1, X_2) = \frac{(1-3)(5-2.75) + (2-3)(3-2.75) + (4-3)(2-2.75) + (5-3)(1-2.75)}{4-1}$$

$$= \frac{(-2)(2.25) + (-1)(0.25) + (1)(-0.75) + (2)(-1.75)}{4-1} = \frac{-9}{3} \approx -3$$

$$\text{Cov}(X_2, X_1) = \frac{(5-2.75)(1-3) + (3-2.75)(2-3) + (2-2.75)(4-3) + (1-2.75)(5-3)}{4-1}$$

$$= \frac{(2.25)(-2) + (0.25)(-1) + (-0.75)(1) + (-1.75)(2)}{4-1} = \frac{-9}{3} \approx -3$$

$$\text{Cov}(X_1, X_2) = \frac{(5-2.75)(5-2.75) + (3-2.75)(3-2.75) + (2-2.75)(2-2.75) + (1-2.75)(1-2.75)}{4-1}$$

$$= \frac{(2.25)(2.25) + (0.25)(0.25) + (-0.75)(-0.75) + (-1.75)(-1.75)}{4-1} = \frac{8.75}{3} \approx 2.92$$

$$\text{Covariance Matrix} = \begin{bmatrix} 3.33 & -3.0 \\ -3.0 & 2.92 \end{bmatrix}$$

ขั้นตอนที่ 3 คำนวณ Eigenvalues (λ)

$$\det \begin{bmatrix} 3.33 & -3.0 \\ -3.0 & 2.92 \end{bmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 0$$

$$\det \begin{bmatrix} 3.33 & -3.0 \\ -3.0 & 2.92 \end{bmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = 0$$

$$\det \begin{pmatrix} 3.33-\lambda & -3.0 \\ -3.0 & 2.92-\lambda \end{pmatrix} = 0$$

$$(3.33-\lambda)(2.92-\lambda) - 9 = 0$$

$$\text{กระจายพจน์ } (3.33-\lambda)(2.92-\lambda)$$

$$= 3.33 \times 2.92 - 3.33\lambda - 2.92\lambda + \lambda^2$$

$$\text{คูณเลข } 3.33 \times 2.92 = 9.7236$$

$$\text{ดังนั้น } = \lambda^2 - (3.33 + 2.92)\lambda + 9.7236$$

$$= \lambda^2 - 6.25\lambda + 9.7236$$

เขียนสมการเพื่อ

$$\lambda^2 - 6.25\lambda + 9.7236 - 9 = 0$$

$$\lambda^2 - 6.25\lambda + 0.7236 = 0$$

ใช้สูตรกำลังสอง $\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, $a=1, b=-6.25, C=0.7236$

$$\lambda = \frac{6.25 \pm \sqrt{(6.25)^2 - 4(1)(0.7236)}}{2}$$

$$\lambda = \frac{6.25 \pm \sqrt{39.0625 - 2.8944}}{2}$$

$$= \frac{6.25 \pm \sqrt{36.1681}}{2}$$

$$= \frac{6.25 \pm 6.014}{2}$$

$$\lambda^1 = \frac{6.25 + 6.014}{2} = \frac{12.264}{2} \approx 6.13$$

$$\lambda = \frac{6.25 - 6.014}{2} = \frac{0.236}{2} \approx 0.12$$

จากการแก้สมการ กำลังสองจะได้ Eigenvalues $\lambda_1 = 6.13$ (Principal Eigenvalue)
 $\lambda_2 = 0.12$

หา Eigen vector สำหรับ Eigen value ที่ใหญ่ที่สุด ($\lambda_1 \approx 6.13$)

$$\begin{bmatrix} 3.33 & -3.0 \\ -3.0 & 2.92 \end{bmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{pmatrix} -2.8 & -3.0 \\ -3.0 & -3.21 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{bmatrix} 3.33 & -3.0 \\ -3.0 & 2.92 \end{bmatrix} - 6.13 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$= \begin{pmatrix} -2.8x & -3.0y \\ -3.0x & -3.21y \end{pmatrix} = 0$$

$$\begin{bmatrix} 3.33 & -3.0 \\ -3.0 & 2.92 \end{bmatrix} - \begin{pmatrix} 6.13 & 0 \\ 0 & 6.13 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$-2.8x + (-3.0y) = 0$$

$$(-3.0x) - 3.21y = 0$$

$$\begin{bmatrix} 3.33-6.13 & -3.0 \\ -3.0 & 2.92-6.13 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$y = \frac{-2.8}{-3.0} x = 0.93x$$

$$y = \frac{-3.0}{3.21} x \approx 0.93x$$

$$y = 0.93x$$

จาก $y = 0.93x$ เลือก $x = 1$ จะได้ $V = [1, 0.93]$

ใช้ทฤษฎีพีทากอรัส $\|V\| = \sqrt{x^2 + y^2}$ จะได้ $\|V\| = \sqrt{1^2 + 0.93^2} = \sqrt{1 + 0.8649}$
 $= \sqrt{1.8649} \approx 1.365$

ทำให้เป็นเวกเตอร์หนึ่งหน่วย $\hat{u} = V / \|V\|$

$$\bullet x\text{-unit} = 1 / 1.365 \approx 0.732$$

$$\bullet y\text{-unit} = 0.93 / 1.365 \approx 0.681$$

$$\text{ดังนั้นจะได้ } V_1(\text{PC1}) \approx [0.73, 0.68]$$

ขั้นตอนที่ 4 & 5 เลือก PC1 และแปลงข้อมูล โดยนำข้อมูลที่ได้ปรับแล้ว $\times PC1 = [0.73, 0.68]$

$$E'' = (-2)(0.73) + (2.25)(0.68) \approx 0.07$$

$$F'' = (-1)(0.73) + (0.25)(0.68) \approx -0.56$$

$$G'' = (1)(0.73) + (-0.75)(0.68) \approx 0.22$$

$$H'' = (2)(0.73) + (-1.75)(0.68) \approx 0.27$$

ผลลัพธ์ จากข้อมูล 2D (x,y) จะเหลือเพียง 1D

$$E'' = 0.07, F'' = -0.56, G'' = 0.22, H'' = 0.27$$