ใบงาน: การลดมิติข้อมูล

รายวิชา: ENGCE207 Advanced Topics in Computer Engineering **หัวข้อ:** Dimensionality Reduction

ชื่อ-สกุล: จันทร์ดี ลุงซุ **รหัสนักศึกษา:** 68143206021-5

คำชี้แจง

ให้นักศึกษาทำการลดมิติข้อมูลจาก 2 มิติ ให้เหลือ 1 มิติ โดยใช้ขั้นตอนของ Principal Component Analysis (PCA) กับชุดข้อมูลที่กำหนดให้ต่อไปนี้

โจทย์ปัญหา

กำหนดชุดข้อมูล 2 มิติ ซึ่งประกอบด้วยจุดข้อมูล 4 จุด ดังนี้

จุดข้อมูล	X	Y
A	2	1
B	4	3
C	5	5
D	7	5

โจทย์ปัญหาเพิ่มเติม

จากชุดข้อมูล 2 มิติที่กำหนดให้ ซึ่งประกอบด้วยจุดข้อมูล 4 จุด ดังตาราง:

จุดข้อมูล	X	Y
E	1	5
F	2	3
G	4	2
H	5	1

1. หาค่าเฉลี่ย

จุดข้อมูล	**X**	**Y**
A	2	1
B	4	3
C	5	5
D	7	5
ค่าเฉลี่ย	(2+4+5+7)/4 = 4.5	(1+3+5+5)/4 = 3.5

จุดข้อมูล	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

2. คำนวณ Convariance Matrix(C)

Convariance Matrix = $\begin{bmatrix} Cov(X_1,X_1) & Cov(X_1,X_2) \\ Cov(X_2,X_1) & Cov(X_2,X_2) \end{bmatrix}$

$Cov(X_1,X_1) = \frac{\sum_{k=1}^{4}(X_{1k}- \overline{X}_1)(X_{1k}- \overline{X}_1)}{4-1}$

$= \frac{(X_{11}-\overline{X}_1)(X_{11}-\overline{X}_1)+(X_{12}-\overline{X}_1)(X_{12}-\overline{X}_1)+(X_{13}-\overline{X}_1)(X_{13}-\overline{X}_1)+(X_{14}-\overline{X}_1)(X_{14}-\overline{X}_1)}{4-1}$

$= \frac{(2-4.5)(2-4.5)+(4-4.5)(4-4.5)+(5-4.5)(5-4.5)+(7-4.5)(7-4.5)}{4-1}$

$= \frac{(-2.5)(-2.5)+(-0.5)(-0.5)+(0.5)(0.5)+(2.5)(2.5)}{4-1} = \frac{13}{3} = 4.33$

$$\text{Cov}(X_1, X_2) = \frac{\sum_{k=1}^{4} (X_{1k} - \overline{X}_1)(X_{2k}, \overline{X}_2)}{4-1}$$

$$= \frac{(X_{11} - \overline{X}_1)(X_{21} - \overline{X}_2) + (X_{12} - \overline{X}_1)(X_{22} - \overline{X}_2) + (X_{13} - \overline{X}_1)(X_{23} - \overline{X}_2) + (X_{14} - \overline{X}_1)(X_{24} - \overline{X}_2)}{4-1}$$

$$= \frac{(2-4.5)(1-3.5)+(4-4.5)(3-3.5)+(5-4.5)(5-3.5)+(7-4.5)(5-3.5)}{4-1}$$

$$= \frac{(-2.5)(-2.5)+(-0.5)(-0.5)+(0.5)(1.5)+(2.5)(1.5)}{4-1}$$

$$= \frac{6.25+0.25+0.75+3.75}{3} = \frac{11}{3} = 3.66$$

$$\text{Cov}(X_2, X_1) = \frac{\sum_{k=1}^{4} (X_{2k} - \overline{X}_2)(X_{1k} - \overline{X}_1)}{4-1}$$

$$= \frac{(X_{21} - \overline{X}_2)(X_{11} - \overline{X}_1) + (X_{22} - \overline{X}_2)(X_{12} - \overline{X}_1) + (X_{23} - \overline{X}_2)(X_{13} - \overline{X}_1) + (X_{24} - \overline{X}_2)(X_{14} - \overline{X}_1)}{4-1}$$

$$= \frac{(1-3.5)(2-4.5)+(3-3.5)(4-4.5)+(5-3.5)(5-4.5)+(5-3.5)(7-4.5}{4-1}$$

$$= \frac{(-2.5)(-2.5)+(-0.5)(-0.5)+(1.5)(0.5)+(1.5)(2.5)}{4-1}$$

$$= \frac{6.25+0.25+0.75+3.75}{3} = \frac{11}{3} = 3.66$$

$$\text{Cov}(X_2, X_1) = \frac{\sum_{k=1}^{4} (X_{2k} - \overline{X}_2)(X_{2k} - \overline{X}_2)}{4-1}$$

$$= \frac{(X_{21} - \overline{X}_2)(X_{21} - \overline{X}_2) + (X_{22} - \overline{X}_2)(X_{22} - \overline{X}_2) + (X_{23} - \overline{X}_2)(X_{23} - \overline{X}_2) + (X_{24} - \overline{X}_2)(X_{24} - \overline{X}_2)}{4-1}$$

$$= \frac{(1-3.5)(1-3.5)+(3-3.5)(3-3.5)+(5-3.5)(5-3.5)+(5-3.5)(5-3.5)}{4-1}$$

$$= \frac{(-2.5)(-2.5)+(-0.5)(-0.5)+(1.5)(1.5)+(1.5)(1.5)}{4-1}$$

$$= \frac{6.25+0.25+2.25+2.25}{3} = \frac{11}{3} = 3.66$$

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

3. คำนวณ Eigenvalues (λ)

- หาค่า λ ที่ทำให้สมการ det(c - λI)=0 เป็นจริง

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

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

(3.66)(3.66)=13.395

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

(4.33-λ)$(3.66-\lambda)$

- เขียนเป็นสมการ

(4.33-λ)$(3.66-\lambda)$ - 13.39 = 0

- การคูณพจน์

$4.33 * 3.66 - 4.33\lambda - 3.66\lambda + \lambda^2$

$4.33 * 3.66 = 15.84$

$-4.33\lambda - 3.66\lambda = -7.99$

จะได้

$\lambda^2 - 7.99\lambda + 15.84$

- ลบ 13.39 ออก

$\lambda^2 - 7.99\lambda + 15.84 - 13.39 = \lambda^2 - 7.99\lambda + 2.45$

- แก้สมการกำลังสอง

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$\lambda = \frac{-(-7.99) \pm \sqrt{(-7.99)^2 - 4(1)(2.45)}}{2(1)}$$

$$\lambda_1 = \frac{7.99 + 7.35}{2} = \frac{15.34}{2} = 7.67$$

$$\lambda_2 = \frac{7.99 - 7.35}{2} = \frac{0.64}{2} = 0.32$$

- แก้สมการกำลังสอง ได้ Egivenvalues

$\lambda_1 = 7.67$ Principal Eigenvalue

$\lambda_2 = 0.32$

- หา Eigenvector สำหรับ Eigenvalue ที่ใหญ่ที่สุด ($\lambda_1 = 7.67$)
  -แก้สมการ $(\mathrm{C}-)V_1\lambda_1$=0

$$\left[\begin{pmatrix}4.33 & 3.66\\3.66 & 3.66\end{pmatrix} - \lambda_1\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right]\left(\frac{x}{y}\right) = 0$$

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

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

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

$$\begin{pmatrix}-3.34 & 3.66\\3.66 & -4.01\end{pmatrix}\left(\frac{x}{y}\right) = 0$$

$$\left(\frac{-3.34\mathrm{x} + 3.66\mathrm{y}}{3.66x - 4.01y}\right) = 0$$

$$-3.34\mathrm{x} + 3.66\mathrm{y} = 0 \qquad 3.66\mathrm{x} - 4.01 = 0$$

$$\mathrm{y} = \frac{3.34}{3.66}\mathrm{x} = 0.91\mathrm{x} \qquad \mathrm{y} = \frac{3.66}{4.01}\mathrm{x} = 0.91\mathrm{x}$$

- การแปลงเป็นเวกเตอร์ 1หน่วย y= 0.91x
  -เวกเตอร์ตั้งต้น: จาก $y = 0.91x$ และเลือก x=1จะได้เวกเตอร์ $v = [1, 0.91]$
  -คำนวณขนาด(Magnitude): โดยใช้ทฤษฎีบทพีทาโกรัส $\|V\| = \mathrm{sqrt}(x^2 + y^2)$
  $\|V\| = \mathrm{sqrt}\left(1^2 + 0.91^2\right) = \mathrm{sqrt}\,(1 + 0.8281) = \mathrm{sqrt}\,(1.8281) = 1.352$

-ทำให้เป็นเวกเตอร์ 1หน่วย: นำแต่ละส่วนประกอบมาหารด้วยขนาดของมัน $\hat{u} = \mathrm{v}/\|v\|$

x_unit = 1/1.352 = 0.739

y_unit = 0.91/1.352 = 0.673

ผลลัพธ์: เมื่อปัดเศษจะได้ Eigenvector V_1 (PC1) = [0.73, 0.67]

- เลือก PC1และแปลงข้อมูล

  เลือก Eigenvector ที่มี Eigenvalue สูงสุดเป็น Principal Component 1 (PC1)

  PC1 = [0.73, 0.67]

  - แปลงข้อมูล Project Data: นำข้อมูลที่ปรับแล้ว Centered Data มาคูณ Dot Prodect กับ PC1

  - สูตร: New Coordinate = A' * PC1

    A" = (-2.5)(0.73) + (-2.5)(0.67) = -3.5

    B" =(-0.5)(0.73) + (-0.5)(0.67) = -0.7

    C" = (0.5)(0.73) + (1.5)(0.67) = 1.37

    D" = (2.5)(0.73) + (1.5)(0.67) = 2.83

  - ผลลัพธ์: ได้ลดมิติข้อมูลจาก 2D (x,y) เหลือเพียง 1D

โจทย์ปัญหาเพิ่มเติม

จุดข้อมูล	**X**	**Y**
E	1	5
F	2	3
G	4	2
H	5	1
ค่าเฉลี่ย	(1+2+4+5)/4= 3	(5+3+2+1)/4= 2.75

จุดข้อมูล	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

2.คำนวณ Convariance Matrix

Covariamce Matrix = $\begin{bmatrix} \text{Cov}(X_1,X_1) & \text{Cov}(X_1,X_2) \\ \text{Cov}(X_2,X_1) & \text{Cov}(X_2,X_2) \end{bmatrix}$

Cov(X_1,X_1) $=\frac{\sum_{k=1}^{4}(X_{1k}-\overline{X}_1)(X_{1k}-\overline{X}_1)}{4-1}$

$=\frac{(X_{11}-\overline{X}_1)(X_{11}-\overline{X}_1)+(X_{12}-\overline{X}_1)(X_{12}-\overline{X}_1)+(X_{13}-\overline{X}_1)(X_{13}-\overline{X}_1)+(X_{14}-\overline{X}_1)(X_{14}-\overline{X}_1)}{4-1}$

$$=\frac{(1-3)(1-3)+(2-3)(2-3)+(4-3)(4-3)+(5-3)(5-3)}{4-1}$$

$$=\frac{(-2)(-2)+(-1)(-1)+(1)(1)+(2)(2)}{4-1}$$

$=\frac{4+1+1+4}{4-1}=\frac{10}{3}=3.33$

Cov(X_1,X_2)$=\frac{\sum_{k=1}^{4}(X_{1k}-\overline{X}_1)(X_{2k}-\overline{X}_2)}{4-1}$

$=\frac{(X_{11}-\overline{X}_1)(X_{21}-\overline{X}_2)+(X_{12}-\overline{X}_1)(X_{22}-\overline{X}_2)+(X_{13}-\overline{X}_1)(X_{23}-\overline{X}_2)+(X_{14}-\overline{X}_1)(X_{24}-\overline{X}_2)}{4-1}$

$$=\frac{(1-3)(5-2.75)+(2-3)(3-2.75)+(4-3)(2-2.75)+(5-3)(1-2.75)}{4-1}$$

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

$$=\frac{(-4.5)+(-0.25)+(-0.75)+(-3.5)}{4-1}=\frac{-9}{3}=-3$$

$$\text{Cov}(X_2, X_1) = \frac{\sum_{k=1}^{4} (X_{2k} - \overline{X}_2)(X_{1k} - \overline{X}_1)}{4-1}$$

$$= \frac{(X_{21} - \overline{X}_2)(X_{11} - \overline{X}_1) + (X_{22} - \overline{X}_2)(X_{12} - \overline{X}_1) + (X_{23} - \overline{X}_2)(X_{13} - \overline{X}_1) + (X_{24} - \overline{X}_2)(X_{14} - \overline{X}_1)}{4-1}$$

$$= \frac{(5 - 2.75)(1 - 3) + (3 - 2.75)(2 - 3) + (2 - 2.75)(4 - 3) + (1 - 2.75)(5 - 3)}{4 - 1}$$

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

$$= \frac{(-4.5) + (-0.25) + (-0.75) + (-3.5)}{4 - 1} = \frac{-9}{3} = -3$$

$$\text{Cov}(X_2, X_2) = \frac{\sum_{k=1}^{4} (X_{2k} - \overline{X}_2)(X_{2k} - \overline{X}_2)}{4-1}$$

$$= \frac{(X_{21} - \overline{X}_2)(X_{21} - \overline{X}_2) + (X_{22} - \overline{X}_2)(X_{22} - \overline{X}_2) + (X_{23} - \overline{X}_2)(X_{23} - \overline{X}_2) + (X_{24} - \overline{X}_2)(X_{24} - \overline{X}_2)}{4-1}$$

$$= \frac{(5 - 2.75)(5 - 2.75) + (3 - 2.75)(3 - 2.75) + (2 - 2.75)(2 - 2.75) + (1 - 2.75)(1 - 2.75)}{4 - 1}$$

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

$$= \frac{5.06 + 0.06 + 0.56 + 3.06}{4 - 1} = \frac{8.74}{3} = 2.91$$

$$\text{Convariance Matrix} = \begin{bmatrix} 3.33 & -3 \\ -3 & 2.91 \end{bmatrix}$$

3. คำนวณ Eigenvalues (λ)

-หาค่า λ ที่ทำให้สมการ det(C - λ)=0 เป็นจริง

$$\det\left[\begin{pmatrix} 3.33 & -3 \\ -3 & 2.91 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] = 0$$

$$\det\left[\begin{pmatrix} 3.33 & -3 \\ -3 & 2.91 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right] = 0$$

(-3)(-3) = 9

$$\det\begin{pmatrix} 3.33-\lambda & -3 \\ -3 & 2.91-\lambda \end{pmatrix} = 0$$

(3.33 - λ)(2.91 - λ)

- เขียนเป็นสมการ
  (3.33 - λ)(2.91 - λ) - 9 = 0

- คูณพจน์
  $3.33 * 2.91 - 3.33\lambda - 2.91\lambda + \lambda^2$
  $3.33 * 2.91 = 9.69$
  $-3.33\lambda - 2.91\lambda = -6.24$

จะได้

$\lambda^2 - 6.24\lambda + 9.69$

- ลบ 9 ออก
  $\lambda^2 - 6.24\lambda + 9.69 - 9 = \lambda^2 - 6.24\lambda + 0.69$

- แก้สมการกำลังสอง
  $$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$\lambda = \frac{-(-6.24) \pm \sqrt{(6.24)^2 - 4(1)(0.69)}}{2(1)}$$

$$\lambda_1 = \frac{6.24 \pm 6.01}{2} = \frac{12.25}{2} = 6.12$$

$$\lambda_2 = \frac{6.24 - 6.01}{2} = \frac{0.23}{2} = 0.11$$

- แก้สมการกำลัง ได้ Eigenvalue
  $\lambda_1 = 6.12$ Principal Eigenvalue
  $\lambda_2 = 0.11$

4. หา Eigenvector สำหรับค่าที่ใหญ่ที่สุด ($\lambda_1 = 6.12$)
   -แก้สมการ (C -)V_1,$\lambda_1 = 0$

$$\left[\begin{pmatrix} 3.33 & -3 \\ -3 & 2.91 \end{pmatrix} - \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\left[\begin{pmatrix} 3.33 & -3 \\ -3 & 2.91 \end{pmatrix} - 6.12 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\left[\begin{pmatrix} 3.33 & -3 \\ -3 & 2.91 \end{pmatrix} - \begin{pmatrix} 6.12 & 0 \\ 0 & 6.12 \end{pmatrix}\right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\left[\begin{pmatrix} 3.33 - 6.12 & -3 \\ -3 & 2.91 - 6.12 \end{pmatrix}\right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{pmatrix} -2.79 & -3 \\ -3 & -3.21 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\left(\frac{-2.79x + -3x}{-3y - 3.21y}\right) = 0$$

$$-2.79\mathrm{x} + -3\mathrm{y} = 0 \qquad -3\mathrm{x} - 3.21\mathrm{y} = 0$$

$$\mathrm{y} = \frac{2.79}{3}\mathrm{x} = 0.93\mathrm{x} \qquad \mathrm{y} = \frac{3}{3.21}\mathrm{x} = 0.93\mathrm{x}$$

5. แปลงเวกเตอร์ 1หน่วย y=0.93x
   -เวกเตอร์ตั้งต้น: จาก y=0.93x และเลือก x=1 จะได้เวกเตอร์ $\mathrm{v} = [1, 0.93]$

$$\|\mathrm{v}\| = \mathrm{sqrt}\left(1^2 + 0.93^2\right) = sqrt\,(1 + 0.8649) = sqrt\,(1.8649) = 1.365$$

x_unit = 1/1.365 = 0.732
y_unit = 0.93/1.365 = 0.681

Eigenvector $\mathrm{v}_1\,(\mathrm{PC1}) = [0.73, 0.68]$

6. เลือก PC1 แล้วแปลงข้อมูล

PC1 = [0.73, 0.68]

E'' = (-2)(0.73) + (2.25)(0.68) = 0.07
F'' = (-1)(0.73) + (0.25)(0.68) = -0.90
G'' = (1)(0.73) + (-0.75)(0.68) = 0.56
H'' = (2)(0.73) + (-1.75)(0.68) = 0.27

สรุป

E''	0.07
F''	-0.90
G''	0.56
H''	0.27