

ใบงาน: การลดมิติข้อมูล

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

ชื่อ-สกุล: ปาระ ยุตสาธร วินogradoff รหัสนักศึกษา: 68143206012-4

คำ解釋

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

จุดที่ 1: นำค่าเฉลี่ย/ปรับข้อมูล (center the data)

P(x,y)	X	Y
A	2	1
B	4	3
C	5	5
D	7	5

mean (x, y) $(\bar{x}, \bar{y}) = (2+4+5+7)/4 = 4.5$ $(1+3+5+5)/4 = 3.5$

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'	$7 - 4.5 = 2.5$	$5 - 3.5 = 1.5$

2: Find Covariance Matrix (C)

$$\text{Cov}(x_1, x_1) = \frac{\sum_{k=1}^4 (x_{1k} - \bar{x}_1)(x_{1k} - \bar{x}_1)}{4-1}$$

$$= \frac{(x_{11} - \bar{x}_1)(x_{11} - \bar{x}_1) + (x_{12} - \bar{x}_1)(x_{12} - \bar{x}_1) + (x_{13} - \bar{x}_1)(x_{13} - \bar{x}_1) + (x_{14} - \bar{x}_1)(x_{14} - \bar{x}_1)}{4-1}$$

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

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

$$\text{Cov}(x_1, x_2) = \frac{\sum_{k=1}^4 (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2)}{4-1}$$

$$= \frac{(x_{11} - \bar{x}_1)(x_{21} - \bar{x}_2) + (x_{12} - \bar{x}_1)(x_{22} - \bar{x}_2) + (x_{13} - \bar{x}_1)(x_{23} - \bar{x}_2) + (x_{14} - \bar{x}_1)(x_{24} - \bar{x}_2)}{4-1}$$

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

$$= \frac{(-2.5)(-2.5) + (-0.5)(-0.5) + (0.5)(1.5) + (2.5)(1.5)}{3} = \frac{11}{3} \approx 3.67$$

$$\text{Cov}(x_2, x_1) = \frac{\sum_{k=1}^4 (x_{2k} - \bar{x}_2)(x_{1k} - \bar{x}_1)}{4-1}$$

$$= \frac{(x_{21} - \bar{x}_2)(x_{11} - \bar{x}_1) + (x_{22} - \bar{x}_2)(x_{12} - \bar{x}_1) + (x_{23} - \bar{x}_2)(x_{13} - \bar{x}_1) + (x_{24} - \bar{x}_2)(x_{14} - \bar{x}_1)}{4-1}$$

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

$$= \frac{(-2.5)(-2.5) + (-0.5)(-0.5) + (1.5)(0.5) + (1.5)(2.5)}{3} \approx 3.67$$

$$\text{Cov}(x_2, x_2) = \frac{\sum_{k=1}^4 (x_{2k} - \bar{x}_2)(x_{2k} - \bar{x}_2)}{4-1}$$

$$= \frac{(x_{21} - \bar{x}_2)(x_{21} - \bar{x}_2) + (x_{22} - \bar{x}_2)(x_{22} - \bar{x}_2) + (x_{23} - \bar{x}_2)(x_{23} - \bar{x}_2) + (x_{24} - \bar{x}_2)(x_{24} - \bar{x}_2)}{4-1}$$

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

$$= \frac{(-2.5)(-2.5) + (-0.5)(-0.5) + (1.5)(1.5) + (1.5)(1.5)}{3} \approx 3.67$$

$$\text{Covariance Matrix} = \begin{bmatrix} 4.33 & 3.67 \\ 3.67 & 3.67 \end{bmatrix}$$

ቻውን መሠረት የ Eigenvalues (λ)

$$\det \begin{bmatrix} 4.33 & 3.67 \\ 3.67 & 3.67 \end{bmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 0$$

$$\det \begin{bmatrix} 4.33 & 3.67 \\ 3.67 & 3.67 \end{bmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = 0$$

$$\det \begin{bmatrix} 4.33 - \lambda & 3.67 \\ 3.67 & 3.67 - \lambda \end{bmatrix} = 0$$

$\xrightarrow{\quad}$

$$(4.33 - \lambda)(3.67 - \lambda) = 0$$

$$(4.33 - \lambda)(3.67 - \lambda) - 13.47 = 0$$

$$\lambda^2 - 8\lambda + 15.89 - 13.47 = 0$$

$$\lambda^2 - 8\lambda + 2.42 = 0$$

$$(\lambda - 7.685)(\lambda - 0.315) = 0$$

የ ንብረቱ በኋላ ስራ የ Eigenvalues

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

$$\lambda = \frac{-(-8) \pm \sqrt{(-8)^2 - 4(1)(2.42)}}{2(1)}$$

$$\lambda = \frac{8 \pm \sqrt{64 - 9.68}}{2}$$

$$\lambda = \frac{8 \pm \sqrt{54.32}}{2}$$

$$\lambda = \frac{8 \pm 7.37}{2}$$

$$\lambda = \frac{8 + 7.37}{2} = \frac{15.37}{2} \approx 7.685$$

$$\lambda = \frac{8 - 7.37}{2} = \frac{0.63}{2} \approx 0.315$$

~~✓~~ $\lambda_1 = 7.685$ (Principal Eigenvalues)

$$\lambda_2 = 0.315$$

47 Eigenvalues និងអនុវត្តន៍ ($\lambda_1 = 7.685$) និងកំសាលា $(C - \lambda_1 I)v_1 = 0$

$$\left[\begin{pmatrix} 4.33 & 3.67 \\ 3.67 & 3.67 \end{pmatrix} - \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\left[\begin{pmatrix} 4.33 & 3.67 \\ 3.67 & 3.67 \end{pmatrix} - 7.685 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\left[\begin{pmatrix} 4.33 & 3.67 \\ 3.67 & 3.67 \end{pmatrix} - \begin{pmatrix} 7.685 & 0 \\ 0 & 7.685 \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{pmatrix} 4.33 - 7.685 & 3.67 \\ 3.67 & 3.67 - 7.685 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{pmatrix} -3.355 & 3.67 \\ 3.67 & -4.015 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{pmatrix} -3.355x + 3.67y \\ 3.67x - 4.015y \end{pmatrix} = 0$$

$$-3.355x + 3.67y = 0$$

$$y = \frac{-3.355}{3.67}x \approx 0.91x$$

$$3.67x - 4.015y = 0$$

$$y = \frac{3.67}{4.015}x \approx 0.91x$$

“ກຳນົດຂໍ້ມູນ ໂດຍ ລາຍງານ ອະນຸຍາກ (Normalization)

- **Exemplo:** $\gamma \approx 0.91x$, então $x=1 \Rightarrow \gamma \approx 0.91$ e podemos ver que $V = [1, 0.91]$

- ຕ່າງໝານຫົວໜ້າ (Magnitude): ອັກ ພົມງ ໂກຮ້າ

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

ការរួមចំណាំ និងការប្រើប្រាស់សាខាដែលត្រូវបានរាយការណ៍ គឺជាការរួមចំណាំ និងការប្រើប្រាស់សាខាដែលត្រូវបានរាយការណ៍

$$x_unit = 1 / 1.352 \approx 0.739$$

$$Y - Y_{init} = 0.91 / 1.352 \approx 0.673$$

\therefore 第一个主成分的 eigenvector v_1 (PC1) $\approx [0.74, 0.67]$

ចំណាំទី 4 & 5 នៃពាក្យ PC7 នេះ នូវលក្ខណៈ

- ເວັນ Eigenvector ສໍາເລັດ Eigenvalue ຂີ່ກົງກົງ ດັວກ ປະເທດ Principal component (PC1)

$$\text{PC1} = [0.74, 0.67]$$

- ແກ່ປົງຈົວລູກ ຂໍາ ພົມເຄົ້າ ປົ່ງປະ ພົມ ພັດຖາ ກົມ PC1

$$S_{MS} = \partial \zeta^2_{\alpha\beta} = A' \cdot PC1$$

$$A'' = (-2.5)(0.74) + (-2.5)(0.67) \approx -3.525$$

$$\beta'' = (-0.5)(0.74) + (-0.5)(0.67) \approx -0.705$$

$$C'' = (0.5)(0.74) + (1.5)(0.67) \approx 1.375$$

$$D'' = (2.5)(0.74) + (1.5)(0.67) \approx 2.955$$

ମେ ଏହି ପାତାରେ କିମ୍ବା କିମ୍ବା କିମ୍ବା

โจทย์ปัญหาเพิ่มเติม

จากชุดข้อมูล 2 มิติที่กำหนดให้ ซึ่งประกอบด้วยจุดข้อมูล 4 จุด ดังตาราง:

จุดข้อมูล	X	Y
E	1	5
F	2	3
G	4	2
H	5	1

P(x,y)	X	Y
E	1	5
F	2	3
G	4	2
H	5	1
mean(ค่าเฉลี่ย)	$(1+2+4+5) \div 4 = 3$	$(5+3+2+1) \div 4 = 2.75$

P'(x,y)	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: Find Covariance Matrix (c)

$$\text{Cov}(x_1, x_1) = \frac{\sum_{k=1}^4 (x_{1k} - \bar{x}_1)(x_{1k} - \bar{x}_1)}{4-1}$$

$$= \frac{(x_{11} - \bar{x}_1)(x_{11} - \bar{x}_1) + (x_{12} - \bar{x}_1)(x_{12} - \bar{x}_1) + (x_{13} - \bar{x}_1)(x_{13} - \bar{x}_1) + (x_{14} - \bar{x}_1)(x_{14} - \bar{x}_1)}{4-1}$$

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

$$= \frac{(-2)(-2) + (-1)(-1) + (1)(1) + (2)(2)}{3} = \frac{8}{3} \approx 2.73$$

$$\text{Cov}(x_1, x_2) = \frac{\sum_{k=1}^4 (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2)}{4-1}$$

$$= \frac{(x_{11} - \bar{x}_1)(x_{21} - \bar{x}_2) + (x_{12} - \bar{x}_1)(x_{22} - \bar{x}_2) + (x_{13} - \bar{x}_1)(x_{23} - \bar{x}_2) + (x_{14} - \bar{x}_1)(x_{24} - \bar{x}_2)}{4-1}$$

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

$$= \frac{(-2)(2.5) + (-1)(0.25) + (1)(-0.75) + (2)(-1.75)}{3} = -\frac{9}{3} \approx -3$$

$$\text{Cov}(x_2, x_1) = \frac{\sum_{k=1}^4 (x_{2k} - \bar{x}_2)(x_{1k} - \bar{x}_1)}{4-1}$$

$$= (x_{21} - \bar{x}_2)(x_{11} - \bar{x}_1) + (x_{22} - \bar{x}_2)(x_{12} - \bar{x}_1) + (x_{23} - \bar{x}_2)(x_{13} - \bar{x}_1) + (x_{24} - \bar{x}_2)(x_{14} - \bar{x}_1)$$

$$= \frac{(5-2.75)(1-3) + (3-2.75)(2-3) + (2-2.75)(4-2) + (1-2.75)(9-3)}{4-1}$$

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

$$\text{Cov}(x_2, x_2) = \frac{\sum_{k=1}^4 (x_{2k} - \bar{x}_2)(x_{2k} - \bar{x}_2)}{4-1}$$

$$= \frac{(x_{21} - \bar{x}_2)(x_{21} - \bar{x}_2) + (x_{22} - \bar{x}_2)(x_{22} - \bar{x}_2) + (x_{23} - \bar{x}_2)(x_{23} - \bar{x}_2) + (x_{24} - \bar{x}_2)(x_{24} - \bar{x}_2)}{4-1}$$

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

$$= \frac{(2.25)(2.25) + (0.25)(0.25) + (-0.75)(-0.75) + (-1.75)(-1.75)}{3} = \underbrace{\frac{8.74}{3}}_{3} \approx 2.91$$

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

ቻውን መሠረት የ λ ንዑስ የ Eigenvalues (λ)

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

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

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

$\xrightarrow{\quad}$

$$(3.33 - \lambda)(2.91 - \lambda) = 0$$

$$(3.33 - \lambda)(2.91 - \lambda) - 9 = 0$$

$$\lambda^2 - 6.24\lambda + 9.69 - 9 = 0$$

$$\lambda^2 - 6.24\lambda + 0.69 = 0$$

$$(\lambda - 6.127)(\lambda - 0.112) = 0$$

የ ንዑስ በ ማስቀመጥ የ λ ንዑስ የ Eigenvalues

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

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

$$\lambda = \frac{6.24 \pm \sqrt{38.9376 - 2.76}}{2}$$

$$\lambda = \frac{6.24 \pm \sqrt{36.1776}}{2}$$

$$\lambda = \frac{6.24 \pm 6.0148}{2}$$

$$\lambda = \frac{6.24 + 6.0148}{2} = \frac{12.2548}{2} \approx 6.127$$

$$\lambda = \frac{6.24 - 6.0148}{2} = \frac{0.2252}{2} \approx 0.112$$

~~ቻ~~ $\lambda_1 = 6.127$ (Principal Eigenvalues)

$$\lambda_2 = 0.112$$

47 Eigenvalues និងអនុវត្តន៍របស់ $(\lambda_1 = 6.127)$ និងការសម្រាប់ $(C - \lambda_1 I)v_1 = 0$

$$\left[\begin{pmatrix} 8.37 & -3 \\ -3 & 2.91 \end{pmatrix} - \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\left[\begin{pmatrix} 8.37 & -3 \\ -3 & 2.91 \end{pmatrix} - 6.127 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\left[\begin{pmatrix} 8.37 & -3 \\ -3 & 2.91 \end{pmatrix} - \begin{pmatrix} 6.127 & 0 \\ 0 & 6.127 \end{pmatrix} \right] \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{pmatrix} 8.37 - 6.127 & -3 \\ -3 & 2.91 - 6.127 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{pmatrix} -2.797 & -3 \\ -3 & -3.217 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

$$\begin{pmatrix} -2.797x & -3y \\ -3x & -3.217y \end{pmatrix} = 0$$

$$-2.797x - 3y = 0$$

$$y = \frac{-2.797}{3}x \approx 0.93x$$

$$-3x - 3.217y = 0$$

$$y = \frac{3}{3.217}x \approx 0.93x$$

“បញ្ចប់ប្រុងបែកពេលវេលា និងសមត្ថភាព (Normalization)

- ចែកចែកតាត់ទីនេះ: $y \approx 0.93x$, ដូច $x=1$ នៅរឹងចែកគោរ៌ $v = [1, 0.93]$

- ការអាមុនុយ (Magnitude): នឹងផ្តល់វិវឌ្ឍន៍

$$\|v\| = \sqrt{1^2 + 0.93^2} = \sqrt{1 + 0.8649} = \sqrt{1.8649} \approx 1.365$$

ការរួមចែកចែកនិងសមត្ថភាព $\hat{v} = v / \|v\|$

$$x_unit = 1 / 1.365 \approx 0.732$$

$$y_unit = 0.93 / 1.365 \approx 0.681$$

\therefore ចំណាំនឹងជាប្រភព Eigenvector v_1 (PC1) $\approx [0.73, 0.68]$

ចំណាំនឹង 4 & 5 នៃពាក្យ PC1 នៅលើខែងធន្លឹង

- នៃពាក្យ Eigenvalue នឹងជាប្រភព Principal component 1 (PC1)
 $PC1 = [0.73, 0.68]$

- បានលាងខ្លួន នៅខែងធន្លឹង ត្រូវអាតុនុយ ក្នុង PC1

$$\text{ខ្លួន} \quad \text{នៅខែងធន្លឹង} = A' \cdot PC1$$

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

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

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

$$D'' = (2)(0.73) + (-2.75)(0.68) \approx 0.27$$

\therefore នាមីតិ នៅខ្លួន ជាភាសា 2D \Rightarrow 1D