Session 5

PCA

Applied Mathematics for AI | Hobot Academy | Zahra Amini

Hobat Aacademy

Telegram: @hobotacademy & Instagram:@hobotacademy & LinkedIn: @zahraamini-ai

Tabular Data

columns = attributes for those observations

Rows = observations

Player	Minutes	Points	Rebounds	Assists
A	41	20	6	5
B	30	29	7	6
C	22	7	7	2
D	26	3	3	9
E	20	19	8	0
F	9	6	14	14
G	14	22	8	3
I	22	36	0	9
J	34	8	1	3

کاهش بعد یا تقلیل ابعاد ← مقدمات ماتریس ها:

ماتریس قطری: ماتریسی که تمام درایه هایش به جز درایه های قطر اصلی، صفر هستند.

$$A=\begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix}$$

ماتریس همانی یا یکه (I): ماتریس مربعی که تمام درایه هایش صفر هستند به جز درایه های قطر اصلی اش که 1 هست.

$$I=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{2\times 2}$$

ماتریس متعامد (orthogonal): ماتریس مربعی است که:

$$A^TA=AA^T=I$$

$$A=\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{5} & -\frac{4}{5} \\ 0 & \frac{4}{5} & \frac{3}{5} \end{bmatrix} \qquad A^T=\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{5} & \frac{4}{5} \\ 0 & -\frac{4}{5} & \frac{3}{5} \end{bmatrix}$$

$$AA^T=\begin{bmatrix} (1\times1+0+0) & 0 & 0 \\ 0 & \underset{1}{\cancel{\left(0+\overset{\frac{9}{25}}{\frac{3}{5}\times\frac{3}{5}}+\overset{\frac{16}{25}}{\frac{-4}{5}\times\frac{-4}{5}}\right)}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

دترمینان ماتریس:

$$A=\begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad |A|=\begin{vmatrix} a & b \\ c & d \end{vmatrix}=ad-bc$$

ex: $A=\begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} \rightarrow |A|=(4\times2)-(1\times1)=7$

ex: $A = \begin{bmatrix} -1 & 4 & 3 \\ 2 & 6 & 4 \\ 3 & -2 & 8 \end{bmatrix} \longrightarrow |A| = ?$

$$|A| = \begin{vmatrix} -1 & 4 & 3 \\ 2 & 6 & 4 \\ 3 & -2 & 8 \end{vmatrix} = -1\big(6\times 8 - (4\times(-2))\big) - 4\big(2\times 8 - 4\times 3\big) + 3\big(2\times(-2) - 3\times 6\big)$$

$$= -138$$

بردار ویژه و مقدار ویژه ی یک ماتریس:

$A.V = \lambda . V$

✓ بردار ویژه ماتریس مربعی A است، اگر ثابت λ وجود داشته باشد به طوری که:

$A = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}_{2\times 2} \qquad V = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

$$A.V = \underbrace{\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 3 \\ 2 \end{bmatrix}}_{V} = \begin{bmatrix} 12 \\ 8 \end{bmatrix} = \underbrace{4}_{\lambda} \underbrace{\begin{bmatrix} 3 \\ 2 \end{bmatrix}}_{V}$$

محاسبه‌ی مقدار ویژه یا λ :

$A.V = \lambda.V$

$\hookrightarrow A.V - \lambda I V = 0 \Longrightarrow (A - \lambda.I).V = 0$

اگر ریشه های دترمینان $(A-\lambda.I)$ را به دست آوریم، مقدار λ به دست می آید.

$|A - \lambda I| = 0$

ex: $A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$ $\quad V, \lambda = ?$

$A.V = \lambda.V \rightsquigarrow$
$|A - \lambda I| = 0$

$\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{bmatrix}$

$$\begin{vmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{vmatrix} = -\lambda(-3-\lambda) - \overset{-2}{\cancel{(+1 \times (-2))}} = +3\lambda + \lambda^2 + 2 = 0 \quad \begin{cases} \lambda_1 = -1 \\ \lambda_2 = -2 \end{cases}$$

فرض: $V_1 = \begin{bmatrix} x \\ y \end{bmatrix}$ $\quad A.V = \lambda.V \xrightarrow{\lambda_1, V_1} AV_1 = \lambda_1 V_1 \Longrightarrow \underbrace{\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{V_1} = \underbrace{(-1)}_{\lambda_1} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{V_1}$

$$\Rightarrow \begin{bmatrix} y \\ -2x-3y \end{bmatrix} = \begin{bmatrix} -x \\ -y \end{bmatrix} \Rightarrow \begin{bmatrix} y \\ -2x-3y \end{bmatrix} - \begin{bmatrix} -x \\ -y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} y-(-x)=0 \\ -2x-3y-(-y)=0 \end{cases} \rightsquigarrow y=-x$$

$$v_1 = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -x \end{bmatrix} \Rightarrow v_1 = x \begin{bmatrix} 1 \\ -1 \end{bmatrix} \rightsquigarrow$$ به ازای تمام xها برقرار است

$A.v_2 = \lambda_2 v_2$ و $\lambda_2 = -2$ $\quad v_2 = \begin{bmatrix} m \\ n \end{bmatrix}$

$$\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}\begin{bmatrix} m \\ n \end{bmatrix} = -2\begin{bmatrix} m \\ n \end{bmatrix} \Rightarrow \begin{bmatrix} n \\ -2m-3n \end{bmatrix} + 2\begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} n+2m \\ -2m-3n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} n+2m=0 \\ -2m-3n=0 \end{cases} \Rightarrow n=-2m \qquad v_2 = \begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} m \\ -2m \end{bmatrix} \Rightarrow v_2 = m\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

مقادیر منفرد (تکین) ماتریس:

مقادیر منفرد ماتریس $A_{m\times n}$ برابر است با جذر مقادیر ویژه ماتریس A^TA یا AA^T:

$$A=\begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \qquad AA^T=\begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}\begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}=\underset{\circledast}{\begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}}$$

برای پیدا کردن مقادیر ویژه کافیه یک λ از قطر اصلی ماتریس $\circledast$ کم کنیم و دترمینان ماتریس حاصل را محاسبه کنیم:

$$\begin{vmatrix} 11-\lambda & 1 \\ 1 & 11-\lambda \end{vmatrix}=0 \Rightarrow (11-\lambda)(11-\lambda)-1=0 \Rightarrow (11-\lambda)^2=1 \Rightarrow (11-\lambda)=\pm 1$$

$$\Rightarrow 11-\lambda=1 \Rightarrow \lambda_1=10 \quad \text{و} \quad 11-\lambda=-1 \Rightarrow \lambda_2=12$$

مقادیر تکین $\longrightarrow \sqrt{\lambda} \Rightarrow \sigma_1=\sqrt{10}$ و $\sigma_2=\sqrt{12}$

SVD (Singular Value Decomposition) ⟵ تجزیه مقدارهای منفرد

$$X = \underset{\text{متعامد}}{U} \; \underset{\text{قطری}}{S} \; \underset{\text{متعامد}}{V^T}$$

$* \text{متعامد} \longrightarrow A \cdot A^T = A^T \cdot A = I$

بررسی روش SVD :

$$X = U S V^T$$

① $X^T X$ را به دست آورید.

② λ را بدست آورید.

③ مقدارهای تکین را به دست آورید و در ماتریس S قرار دهید.

④ V را به دست آورید.

⑤ U را به دست آورید. ⟵ $X V_1 = \sigma_1 u_1$

$$X = \underset{2\times2}{\begin{bmatrix} m & p \\ n & q \end{bmatrix}} \; \underset{\substack{2\times2 \\ \text{قطری}}}{\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}} \; \underset{2\times2}{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}$$

ex: $X = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$ $\qquad X = USV^T$

$$\begin{array}{ccc} U & S & V \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{8} \end{bmatrix} & \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \end{array}$$

X^TX

① $X^TX = \begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$

② $\lambda \rightarrow \begin{vmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{vmatrix} = 0 \Rightarrow (5-\lambda)^2 - 9 = 0 \Rightarrow 5-\lambda = \pm 3 \Rightarrow \lambda_1 = 2 \rightsquigarrow \sigma_1 = \sqrt{2}$ ③

$\lambda_2 = 8 \rightsquigarrow \sigma_2 = \sqrt{8}$

$\hookrightarrow S = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{8} \end{bmatrix}$

④ $\lambda_1 = 2 \rightsquigarrow V_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ $\qquad \lambda_2 = 8 \rightsquigarrow V_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

⑤ $U \rightsquigarrow XV_1 = \sigma_1 U_1 \Rightarrow \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \sqrt{2}\begin{bmatrix} a \\ b \end{bmatrix} \Rightarrow a = 0,\ b = 1 \Rightarrow U_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$XV_2 = \sigma_2 U_2 \Rightarrow \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \sqrt{8}\begin{bmatrix} c \\ d \end{bmatrix} \Rightarrow c = 1,\ d = 0 \Rightarrow U_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

تقلیل ابعاد:

کاربردها ← 1. Data Visualization ← برای ترسیم داده ها در 2 یا 3 بعد

← 2. Data Compression ← برای تسریع الگوریتم ها

1. استخراج ویژگی ← Feature Extraction ← 1. تبدیل خطی یا غیرخطی بر روی فضای ویژگی

2. ایجاد ویژگی های جدید از ویژگی های اصلی

$$\begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \longrightarrow \begin{bmatrix} y_1 \\ \vdots \\ y_{d'} \end{bmatrix} = f\left(\begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}\right)$$

2. انتخاب ویژگی ← Feature Selection

$$\begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \longrightarrow \begin{bmatrix} x_{i_1} \\ \vdots \\ x_{i_{d'}} \end{bmatrix}$$

$$d' < d$$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_{100} \end{bmatrix} \xrightarrow{FS} \begin{bmatrix} x_2 \\ x_5 \\ x_{90} \end{bmatrix} \qquad d = 100, \quad d' = 3$$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_{100} \end{bmatrix} \xrightarrow{FE} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \qquad y_1 = x_2 + 6x_{70}$$

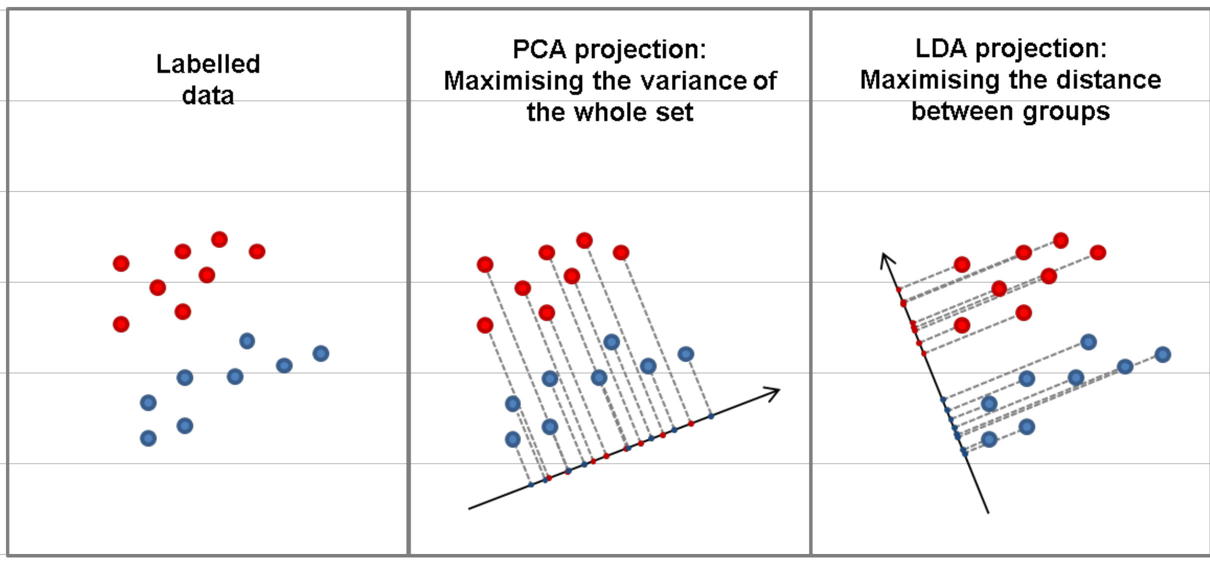
الگوریتم های Feature Selection:

PCA ← Unsupervised

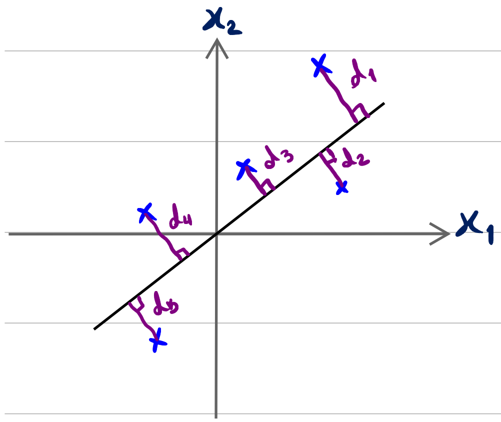
به ما یک خط می دهد که زمانی داده ها روی آن Project یا تصویر می شوند بیشترین واریانس را حفظ می کنند.

LDA ← Supervised

به ما یک خط می دهد که زمانی داده ها روی آن Project یا تصویر می شوند، داده های دو کلاس متفاوت از هم فاصله دارند و داده های هر کلاس به هم نزدیک هستند.



PCA ← در اصل خطی را می‌یابد که داده‌ها کمترین فاصله عمودی را با آن داشته باشند.



minimize $\rightsquigarrow$ $d_1^2 + d_2^2 + \dots + d_5^2$

ex:

x_1	x_2
1	2
3	4
5	6

① محاسبه میانگین:

$$\bar{x}_1 = \frac{1+3+5}{3} = 3$$

$$\bar{x}_2 = \frac{2+4+6}{3} = 4$$

② Centerized

$$\begin{bmatrix} 1-3 & 2-4 \\ 3-3 & 4-4 \\ 5-3 & 6-4 \end{bmatrix}$$

③ Cov Matrix $\rightsquigarrow$ $Cov = \frac{1}{m-1} X^T X = \frac{1}{3-1} \begin{bmatrix} -2 & 0 & 2 \\ -2 & 0 & 2 \end{bmatrix} \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$

④ $\lambda \rightsquigarrow \begin{vmatrix} 4-\lambda & 4 \\ 4 & 4-\lambda \end{vmatrix} = 0 \Rightarrow (4-\lambda) = \pm 4$

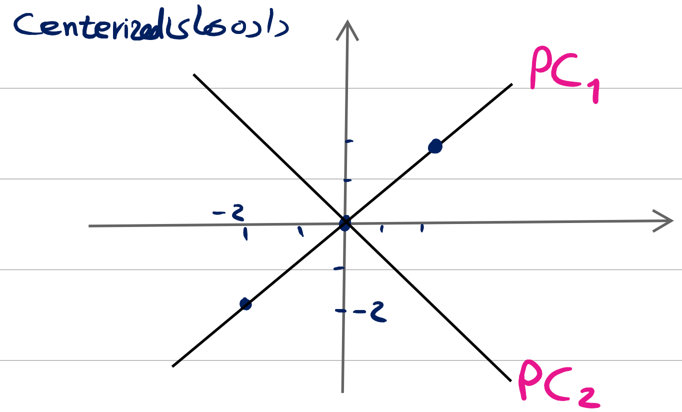
$|v_1| = \sqrt{2}$

$$\Rightarrow \lambda_1 = 0 \longrightarrow v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \xrightarrow{Norm} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0.7 \\ 0.7 \end{bmatrix}$$

$$\Rightarrow \lambda_2 = 8 \longrightarrow v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \xrightarrow{Norm} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -0.7 \\ 0.7 \end{bmatrix}$$

$$PC_1 \leadsto y = mx \Rightarrow y = \frac{0.7}{0.7}x \Rightarrow y = x$$

$$PC_2 \leadsto y = mx \Rightarrow y = \frac{-0.7}{0.7}x \Rightarrow y = -x$$

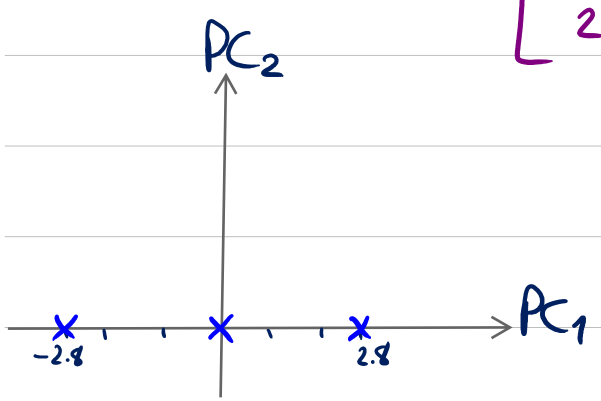


کدام $PC = ?$

PC_1 چون به ما فاصله‌ی کمینه (ه) را می‌دهد.

ماتریس Proj

$$P = X \cdot V \leadsto \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix} \times \begin{bmatrix} 0.7 & -0.7 \\ 0.7 & 0.7 \end{bmatrix} = \begin{array}{c} \begin{bmatrix} -2.8 & 0 \\ 0 & 0 \\ 2.8 & 0 \end{bmatrix} \\ \begin{matrix} PC_1 & PC_2 \end{matrix} \end{array}$$



کافیست که با PC_1 اطلاعات را ذخیره کردیم و خیلی راحت از 2 بعد به 1 بعد رسیدیم.

الگوریتم PCA:

1. $\mu = \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$

2. Centered X

3. $Cov = \frac{1}{m-1} X^T X$

4. λ, v

5. $\lambda = \overset{\text{Sorted}}{[\lambda_1 \dots \lambda_d]}$ $\qquad v = [\underbrace{v_1}_{\text{first PC}}, \dots, \underbrace{v_{d'}}_{d'\text{-th PC}}]$

6. $P = XV$

PCA به کمک SVD:

1. $\mu_j = \frac{1}{m} \sum_{i=1}^{m} x_j^{(i)}$

2. Normalized $\rightsquigarrow x_j^{(i)} - \mu_j$

3. اگر مقیاس feature ها بسیار متفاوت بود، در ضرایبی متناسب با دامنه آنها ضرب شوند.

4. Cov Matrix $\rightsquigarrow \Sigma = \frac{1}{m} \sum_{i=1}^{n} (x^{(i)})(x^{(i)})^T$

5. محاسبه ی λ و v با استفاده از روش SVD

	S1	S2	S3	S4
Math	95 ÷10 → 9.5	88 8.8	93 9.3	75 7.5
Art	9	8	10	7