06 Friday JANUARY

JAN '23

M	T	W	T	F	S	S
					1	2
9	10	11	12	13	14	15
23	24	25	26	27	28	29

M	T	W	T	F	S	S
3	4	5	6	7	8	
16	17	18	19	20	21	22
30	31					

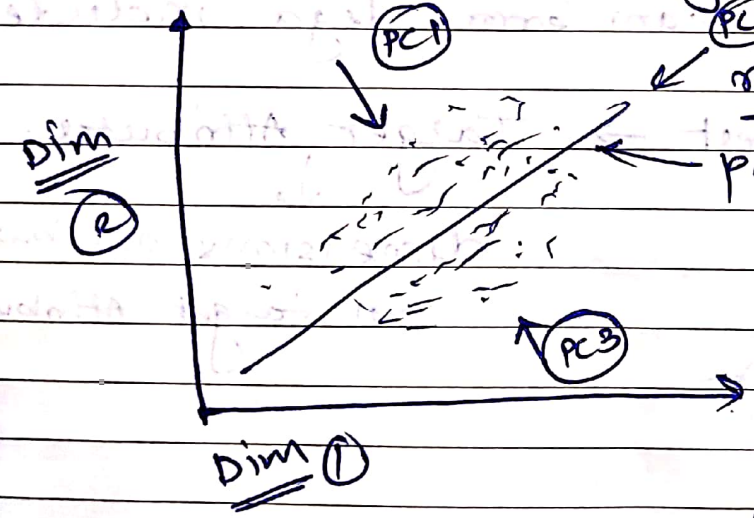
006-359 Wk-01

2023

Principal component Analysis

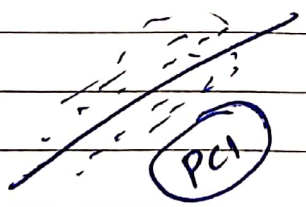
- It is unsupervised learning algo
- Used for dimensionality reduction in ML
- It is a statistical process that converts the observations of correlated features into a set of linear uncorrelated features. With the help of orthogonal transformation
- It is used in
  1. Image processing
  2. Movie recommendation system
  3. optimizing power allocation in various comm. channels.

It is based on 1. Variance & covariance
2. Eigen values & eigen vectors.



model passing through each point

This is overfitting due to extra features so find PC.



PCA
2D → 1D

Reduce dataset from 1D to 2D.

F	Ex1	Ex2	Ex3	Ex4
x_1	4	8	13	7
x_2	11	4	5	14

Step ① find mean

$$\overline{x_1} = \frac{1}{4}(4+8+13+7) = 8$$

$$\overline{x_2} = \frac{1}{4}(11+4+5+14) = 8.5$$

Step ② :- Covariance matrix

$$S = \begin{bmatrix} cov(x_1,x_1) & cov(x_1,x_2) \\ cov(x_2,x_1) & cov(x_2,x_2) \end{bmatrix}$$

$$cov(x_1,x_1) = \frac{1}{N-1}\sum_{k=1}^{N}(x_{1k}-\overline{x_1})(x_{1k}-\overline{x_1})$$

$$= \frac{1}{4-1}\left((4-8)^2+(8-8)^2+(13-8)^2+(7-8)^2\right)$$

$$= 14$$

$$cov(x_1,x_2) = \frac{1}{N-1}\sum_{k=1}^{N}(x_{1k}-\overline{x_1})(x_{2k}-\overline{x_2})$$

$$= \frac{1}{3}\left[(4-8)(11-8.5)+(8-8)(4-8.5)+(13-8)(5-8.5)+(7-8)(14-8.5)\right]$$

$$= -11$$

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

$$= \frac{1}{N-1} \sum_{k=1}^{4} (\quad)(\quad) +$$

$$(\quad)(\quad) + (\quad)(\quad) +$$

$$(\quad)(\quad)$$

$$= -11$$

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

$$= \frac{1}{3} \left((11 - 8.5)^2 + (4 - 8.5)^2 + (5 - 8.5)^2 + (14 - 8.5)^2 \right)$$

$$= 23$$

$$S = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

Step 3 :- Calculate eigen value of covariance matrix

The characteristic eqn of the covariance matrix is

$$0 = \det(S - \lambda I)$$

$$0 = \begin{vmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{vmatrix}$$

Note ⇒ I for 2×2 matrix

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

$$\therefore \lambda I = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$S - \lambda I = \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix}$$

$$0 = (14-\lambda)(23-\lambda) - (-11) \times (-11)$$

$$= \lambda^2 - 37\lambda + 201$$ ← solve in Calci

$$\lambda = \frac{1}{2}\left(37 \pm \sqrt{565}\right)$$

$$= 30.3849, 6.615$$

$$= \lambda_1, \lambda_2$$

Step 4 :- Eigen vector calculation

$$U = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\therefore \begin{bmatrix} 0 \\ 0 \end{bmatrix} = (S - \lambda I)\, U$$

$$= \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$= \begin{bmatrix} (14-\lambda)u_1 & -11u_2 \\ -11u_1 & (23-\lambda)u_2 \end{bmatrix}$$

eqn 1 → $(14-\lambda)u_1 - 11u_2 = 0$

eqn 2 → $-11u_1 + (23-\lambda)u_2 = 0$

consider eqn ①

$(14-\lambda)u_1 = 11u_2$

$\therefore \frac{u_1}{11} = \frac{u_2}{14-\lambda} = t$ ..... (say)

$\therefore u_1 = 11t \quad u_2 = (14-\lambda)t$

put $t=1$

$\therefore U = \begin{bmatrix} 11 \\ 14-\lambda \end{bmatrix}$

Note:- ① for PCA always consider largest value out of two eigen values

$\therefore \lambda_1 = 30.3849 \quad \lambda_2 = 6.6151$ ✗

② If one PCA is asked consider largest value

③ If multiple PC is asked consider 2nd eigen value

$\therefore \quad U_1 = \begin{bmatrix} 11 \\ 14 - \lambda_1 \end{bmatrix}$

To find unit eigen vector, compute the length of U_1

$\therefore \quad \|U_1\| = \sqrt{11^2 + (14 - \lambda)^2}$

$= \sqrt{(11)^2 + (14 - 30.3849)^2}$

$= 19.73.$

$e_1 = \begin{bmatrix} 11 / \|U_1\| \\ (14 - \lambda_1) / \|U_1\| \end{bmatrix}$

↑ 1st eigen vector

$= \begin{bmatrix} \frac{11}{19.73} \\ \frac{(14 - 30.3849)}{19.73} \end{bmatrix} = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$

One more eigen value

$U_2 = \begin{bmatrix} 11 \\ 14 - \lambda_2 \end{bmatrix}$

$\therefore \quad \|U_2\| = \sqrt{(11)^2 - (14 - 6.6151)^2}$

$= \sqrt{\qquad} =$

FEBRUARY MARCH APRIL MAY JUNE

$$e_2 = \begin{bmatrix} \frac{11}{\|U_2\|} \\ \\ \frac{14-\lambda_1}{\|U_2\|} \end{bmatrix}$$

$$e_2 = \begin{bmatrix} \\ \\ \end{bmatrix} = \begin{bmatrix} 0.8303 \\ \\ 0.5574 \end{bmatrix}$$

Step 5 : Computation of first PC

$$e_1^T \begin{bmatrix} x_{1k} - \bar{x}_1 \\ x_{2k} - \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} x_{11} - \bar{x}_1 \\ x_{21} - \bar{x}_2 \end{bmatrix}$$

$$= 0.5574 (x_{11} - \bar{x}_1) - 0.8303 (x_{21} - \bar{x}_2)$$

$$= 0.5574 (4 - 8) - 0.8303 (11 - 8.5)$$

$$= -4.30535$$

Computation of 2nd PC

$$e_2^T \begin{bmatrix} x_{1k} - \bar{x}_1 \\ x_{2k} - \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} x_{12} - \bar{x}_1 \\ x_{22} - \bar{x}_2 \end{bmatrix}$$

$$= 0.5574 (x_{12} - \bar{x}_1) - 0.8303 (x_{22} - \bar{x}_2)$$

$$= 0.65 (8 - 8) - 0.8303 (4 - 8.5)$$

$$=$$

Computation for 3rd component

$$e_1^T \begin{bmatrix} x_{1k} - \bar{x}_1 \\ x_{2k} - \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} x_{13} - \bar{x}_1 \\ x_{23} - \bar{x}_2 \end{bmatrix}$$

Sunday 15

Computation for 4th PC

$$e_1^T \cdot \begin{bmatrix} x_{1k} - \bar{x}_1 \\ x_{2k} - \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} x_{14} - \bar{x}_1 \\ x_{24} - \bar{x}_2 \end{bmatrix}$$

Feature	Ex1	Ex2	Ex3	Ex4
x_1	4	8	13	7
x_2	11	4	5	14
First Principle Components	-4.3052	3.7361	5.6928	-5.1238