INDEX

Principal Component Analysis Steps

NAME :

S. No.	Date	Topic/Chapter	Page No.

Step 1: Data - Set

Features	Eg 1	e.g 2	....	example N
X_1	X_{11}	X_{12}	----	X_{1N}
X_2	X_{21}	X_{22}	----	X_{2N}
X_3	X_{31}	X_{32}	----	X_{3N}
X_n	X_{n1}	X_{n2}	----	X_{nN}

Step 2 Compute the means of the variables

Mean of X_i

$$\bar{X}_i = \frac{1}{N}(X_{i1} + X_{i2} + \cdots X_{iN})$$

Step 3: Calculate the covariance matrix

Covariance of all the ordered pairs (X_i, X_j)

$$S = \begin{array}{c c} & \begin{array}{cccc} X_1 & X_2 & \cdots & X_n \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ \vdots \\ X_n \end{array} & \begin{bmatrix} cov(X_1 X_1) & cov(X_1, X_2) & \cdots & cov(X_1 X_n) \\ cov(X_2 X_1) & cov(X_2; X_2) & & cov(X_2 X_n) \\ \vdots & \vdots & & \vdots \\ cov(X_n X_1) & cov(X_n X_2) & \cdots & cov(X_n X_N) \end{bmatrix} \end{array}$$

$$\boxed{Cov(X_i, X_j) = \frac{1}{N-1}\sum_{k=1}^{N}(X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)}$$

INDEX

NAME :

S. No.	Date	Topic/Chapter	Page

Step 4: Calculate the eigenvalues and normalised eigenvectors of the covariance matrix.

(i) ⟹ To find eigen values, solve the equation

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

We get n roots $\lambda_1, \lambda_2 \dots \lambda_n$ which are eigen values such that $\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_n$

(ii) For each eigen values the corresponding eigen vector is a vector

$$U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad n \times 1$$

if $\lambda = \lambda'$ is an eigenvalue, then the corresponding eigenvector is a vector

such that $(S - \lambda' I) U = 0$

Step 5 ⟹ Normalise the eigen vector
Divide the vector U by its length.
i.e Normalised eigen vectors

$$e_i = \frac{U_i}{\|U_i\|} \quad i = 1, 2, \dots n.$$

where $\|U\| = \sqrt{u_1^2 + u_2^2 + \dots u_n^2}$

* The unit eigen vector corresponding to the largest eigen value is the first principal component.

Step 6: Derive the new dataset

New dataset with reduced dimension is

Feature	Example 1	Example 2	---	Example N
PC_1	P_{11}	P_{12}	---	P_{1N}
PC_2	P_{21}	P_{22}	--	P_{2N}
⋮	⋮	⋮	⋮	⋮
PC_n	P_{n1}	P_{n2}	⋮	P_{nN}

such that

$$P_{ij} = e_i^T \begin{bmatrix} x_{1j} - \bar{x}_1 \\ x_{2j} - \bar{x}_2 \\ \vdots \\ x_{nj} - \bar{x}_n \end{bmatrix}$$

Problem Set: Given the data in below table, use PCA to reduce the dimension from 2 to 1

Feature	Example 1	Example 2	Example 3	Example 4
X_1	4	8	13	7
X_2	11	4	5	14

Solution

X_1	X_2	$(X_1-\bar{X})$	$(X_1-\bar{X})^2$	$(X_2-\bar{X})$	$(X_2-\bar{X})^2$
4	11	-4	16	2.5	6.25
8	4	0	0	-4.5	20.25
13	5	5	25	-3.5	12.25
7	14	-1	1	5.5	30.25

$\Sigma X_1 = 32$ $\quad \Sigma X_2 = 34$ $\quad \Sigma(X_1-\bar{X})^2 = 42$ $\quad \Sigma(X_2-\bar{X})^2 = 69$

$\bar{X}_1 = 32/4 = 8$ $\quad \bar{X}_2 = \frac{34}{4} = 8.5$

Step 1 Mean $\bar{X}_1 = 8$ $\quad \bar{X}_2 = 8.5$

Step2: Covariance Matrix

$$S = \begin{array}{c|cc} & X_1 & X_2 \\ \hline X_1 & Cov(X_1, X_1) & Cov(X_1, X_2) \\ X_2 & Cov(X_2, X_1) & Cov(X_2, X_2) \end{array} = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_2) & Var(X_2) \end{bmatrix}$$

We know that

$$Cov(X_1, X_1) = \frac{1}{N-1} \sum_{k=1}^{N} (X_{1k} - \bar{X}_1)^2$$

$$= \frac{1}{3} \times 42 = 14$$

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

$$= \frac{1}{3}\left[(-4 \times 2.5) + (0 \times -4.5) + (5 \times -3.5) + (-1 \times 5.5)\right]$$

$$= \frac{1}{3}\left[-10 + 0 - 17.5 - 5.5\right] = \frac{1}{3} \times 33 = -11$$

$$Cov(X_2, X_1) = Cov(X_1, X_2) = -11$$

$$Cov(X_2, X_2) = \frac{1}{N-1} \sum_{k=1}^{N} (X_{2k} - \bar{X}_2)^2 = \frac{1}{3} \times 69 = 23$$

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

Step3 Eigen value of the covariance matrix

$$\det(S - \lambda I) = 0 \qquad \lambda I = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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

$$(14-\lambda)(23-\lambda) - (-11)\times(-11) = \lambda^2 - 37\lambda + 201$$

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{37 \pm \sqrt{(37)^2 - 4 \times 1 \times 201}}{2} = \frac{37 \pm \sqrt{565}}{2}$$

$$\lambda = \left(\frac{37 + 23.77}{2}, \frac{37 - 23.77}{2}\right)$$

$$30.3849, \ 6.6151$$

$$\lambda_1 = 30.3849$$

$$\lambda_2 = 6.6151$$

Step 4. <u>Computation of the eigenvector</u>

Eigen vector corresponding to $\lambda = \lambda_1$ is a Vector $U = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$

$$(S - \lambda_1 I) U_1 = 0$$

$$\begin{bmatrix} 14-\lambda_1 & -11 \\ -11 & 23-\lambda_1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$(14-\lambda_1) v_1 - 11 v_2 = 0$$
$$-11 v_1 + (23-\lambda_1) v_2 = 0$$

$$\frac{v_1}{11} = \frac{v_2}{14-\lambda_1} = t \Rightarrow v_1 = 11t, \quad v_2 = 14-\lambda\, t$$

$U_1 = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix}$ eigen vector

We know that

$AV = \lambda V$

matrix, eigenvalue, eigen Vector

$$\underset{A}{\begin{bmatrix} 6 & 3 \\ 4 & 5 \end{bmatrix}} \underset{v}{\begin{bmatrix} 1 \\ 4 \end{bmatrix}} = \underset{\lambda}{6} \underset{v}{\begin{bmatrix} 1 \\ 4 \end{bmatrix}}$$

To find a unit eigen vector we compute the length of U_1

$$\|U_1\| = \sqrt{11^2 + (14-\lambda_1)^2} = \sqrt{121 + (14-30.3849)^2}$$
$$= 19.7348$$

$\therefore$ Unit eigen Vector corresponding to $\lambda_1 = 30.3849$ is

$$e_1 = \left\{ \frac{U_1}{\|U_1\|} \right. = \begin{bmatrix} \frac{11}{\|U_1\|} \\ (14-\lambda_1)/\|U_1\| \end{bmatrix} = \begin{bmatrix} 11/19.7348 \\ (14-30.3849)/19.7348 \end{bmatrix}$$

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

~~Also~~ By Carrying out similar computation the unit ~~b~~eigenvector e_2 corresponding to the eigenvalue $\lambda = \lambda_2$ can be shown to be $e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$

$\lambda_2 = 6.6151$

$$(S - \lambda_2 I) U_2 = 0$$

$$\begin{bmatrix} 14-\lambda_2 & -11 \\ -11 & 23-\lambda_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$(14-\lambda_2) v_1 - 11 v_2 = 0$$
$$-11 v_1 + (23-\lambda_2) v_2 = 0$$

$$\frac{v_1}{11} = \frac{v_2}{14-\lambda_2} = t \Rightarrow U_2 = \begin{bmatrix} 11 \\ 14-\lambda_2 \end{bmatrix}$$

~~Unit eigenvector~~

Second Unit eigen vector corresponding to $\lambda_2 = 6.6151$

$$e_i = \frac{U_2}{\|U_2\|} \qquad \|U_2\| = \sqrt{11^2 + (14-\lambda_2)^2} = 13.249$$

$$= \begin{bmatrix} 11/\|U_2\| \\ \frac{14-\lambda_2}{\|U_2\|} \end{bmatrix} = \begin{bmatrix} 11/13.249 \\ \frac{14-6.6151}{13.249} \end{bmatrix} = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$$

<u>Steps</u> Computation of first principal components

$$\text{First PC1} = e_1^T \begin{bmatrix} X_{1k} - \bar{X}_1 \\ X_{2k} - \bar{X}_2 \\ \vdots \\ X_{nk} - \bar{X}_n \end{bmatrix}$$

we have

$$e_1 = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix} \qquad e_1^T = [0.5574 \quad -0.8303]$$

$$\underset{P_{i1}}{PC1} = e_1^T \begin{bmatrix} X_{11} - \bar{X}_1 \\ X_{21} - \bar{X}_2 \end{bmatrix} = [0.5574 \quad -0.8303] \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$$

$$= -4.30535$$

X_1	4	8	13	7
X_2	11	4	5	14
First PC	-4.30535	3.7361	5.6928	-5.1238

$$P_{i2} = e_1^T \begin{bmatrix} X_{12} - \bar{X}_1 \\ X_{22} - \bar{X}_2 \end{bmatrix} = [0.5574 \quad -0.8303] \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix} = 3.7361$$

& so on.