PCA Manual Calculations

Authored by - Premanand

consider dataset,

Feature	sample1	sample2	sample3	sample4
a	4	8	13	7
b	11	4	5	14

Step1:

No. of features, n = 2 (a,b)

No. of samples, N = 4 (sample1, sample2, sample3, sample4)

Step2:

calculating mean,

$$\bar{a} = \frac{4+8+13+7}{4} = 8$$

$$\bar{b} = \frac{11+4+5+14}{4} = 8.5$$

Step3:

calculating covariance matrix, between features,
In the given dataset, ordered features are as,
(a,a), (a,b), (b,a), (b,b)

$$\text{cov}(a,a) = \frac{1}{N-1}\sum_{k=1}^{N}(a_i-\bar{a})(a_i-\bar{a})$$

$$= \frac{1}{N-1}\sum_{k=1}^{N}(a_i-\bar{a})^2 \rightarrow \text{for same feature}$$

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

$$= \frac{4^2+0+5^2+1^2}{3} = \frac{16+0+25+1}{3}$$

$$= \frac{42}{3} = \underline{\underline{14}}.$$

$$\text{cov}(a,b) = \frac{1}{N-1}\sum_{k=1}^{N}(a_i-\bar{a})(b_i-\bar{b})$$

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

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

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

$$\operatorname{cov}(b,a) = \frac{1}{N-1}\sum_{k=1}^{N}(b_i-\bar{b})(a_i-\bar{a})$$

$$= \operatorname{cov}(a,b)$$

$$= -11$$

$$\operatorname{cov}(b,b) = \frac{1}{N-1}\sum_{k=1}^{N}(b_i-\bar{b})\,(b_i-\bar{b})$$

$$= \frac{1}{N-1}\sum_{k=1}^{N}(b_i-\bar{b})^2$$

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

$$= \frac{1}{3}\left[(2.5)^2+(-4.5)^2+(-3.5)^2+(5.5)^2\right]$$

$$= \frac{1}{3}\left[6.25+20.25+12.25+30.25\right]$$

$$= \frac{69}{3} = \underline{\underline{23}}$$

Hence covariance matrix can be

$$S = \begin{bmatrix} \operatorname{cov}(a,a) & \operatorname{cov}(a,b) \\ \operatorname{cov}(b,a) & \operatorname{cov}(b,b) \end{bmatrix} = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

Step 4:-

calculate Eigen value, Eigen vector, Normalized Eigen vector.

Inorder calculate Eigen value,

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

$$I(\text{Identity matrix}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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

$$\det\left(\begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = 0$$

$$\det\begin{pmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{pmatrix} = 0.$$

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

$$322-14\lambda-23\lambda+\lambda^2-121=0$$

$201 - 37\lambda + \lambda^2 = 0.$

After rearranging,

$$\lambda^2 - 37\lambda + 201 = 0.$$

'λ' can be calculated by quadratic equ,

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad a = 1,\ b = -37,\ c = 201$$

$$= \frac{-(-37) \pm \sqrt{(-37)^2 - 4(1)(201)}}{2(1)}$$

$$= \frac{37 \pm \sqrt{1369 - 804}}{2} = \frac{37 \pm \sqrt{565}}{2}$$

$$= \frac{37 \pm 23.76}{2} \Rightarrow \frac{37 + 23.76}{2}, \frac{37 - 23.76}{2}$$

$$= \frac{60.76}{2}, \frac{13.24}{2}$$

Eigen values. $\boxed{\lambda_1 = 30.38,\ \lambda_2 = 6.62}$

So, while arranging in descending order,

$$\lambda_1 > \lambda_2 > \ldots$$

Hence, $\lambda_1 = 30.38$

$\lambda_2 = 6.62$

We are going to find out Eigenvectors for Eigen value, $\lambda = 30.38$.

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

Covariance matrix

30.38

Identity matrix

Eigen vector of λ_1

$$\left(\begin{pmatrix} 14 & -11 \\ -11 & 23 \end{pmatrix} - 30.38\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) U_1 = 0$$

Assume $U_1 = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$

Hence,

$$\begin{pmatrix} 14 & -11 \\ -11 & 23 \end{pmatrix} - \begin{pmatrix} 30.38 & 0 \\ 0 & 30.38 \end{pmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{pmatrix} 14-30.38 & -11 \\ -11 & 23-30.38 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{pmatrix} -16.38 & -11 \\ -11 & -7.38 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$-16.38v_1 - 11v_2 = 0 \rightarrow (1)$

$-11v_1 - 7.38v_2 = 0 \rightarrow (2)$

so, from this if we need to calculate v_1 & v_2

$(1) \times 7.38 \Rightarrow 120.88v_1 - \cancel{81.18}v_2 = 0$

$(2) \times -11. \Rightarrow +121v_1 + \cancel{81.18}v_2 = 0$

$0.12v_1 = 0$

$\boxed{v_1 = 0}$

then, apply v_1 in (1), then

$-16.38 \times 0 - 11v_2 = 0$

$\boxed{v_2 = 0}$

this can't be possible, hence

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

$(14-\lambda_1)v_1 - 11v_2 = 0 \longrightarrow (a)$

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

from (a),

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

$(14-\lambda_1)v_1 = 11v_2.$

$$\frac{v_1}{11} = \frac{v_2}{14-\lambda_1} = A \text{ (Assigning)}$$

Assume $A=1$,

$$\frac{v_1}{11} = \frac{v_2}{14-\lambda_1} = A = 1.$$

Hence, $\frac{v_1}{11} = 1 \Rightarrow v_1 = 11.$

$\frac{v_2}{14-\lambda_1} = 1 \Rightarrow v_2 = 14-\lambda_1$

$= 14-30.38$

$= -16.38.$

Hence Eigen vector for $\lambda_1 \Rightarrow \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.38 \end{bmatrix}$

Then ~~so~~ if we want to normalize the eigen vector,

$$n_1 = \begin{bmatrix} 11/\sqrt{11^2+16.38^2} \\ -16.38/\sqrt{11^2+16.38^2} \end{bmatrix}$$ / dividing by the length.

$$= \begin{bmatrix} \frac{11}{19.73} \\ -16.38/19.73 \end{bmatrix} = \begin{bmatrix} 0.5575 \\ -0.8302 \end{bmatrix}$$

Now, calculate eigen vector for $\lambda_2 = 6.62$

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

$$\begin{bmatrix} (14-\lambda_2) & -11 \\ -11 & (23-\lambda_2) \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$(14-\lambda_2)U_1 - 11U_2 = 0 \rightarrow (c)$$

$$-11U_1 - (23-\lambda_2)U_2 = 0 \rightarrow (d)$$

from (c),

$$(14-\lambda_2)U_1 - 11U_2 = 0$$

$$(14-\lambda_2)U_1 = 11U_2$$

$$\frac{U_1}{11} = \frac{U_2}{14-\lambda_2} = B \text{ (Assume)}$$

Assume $B=1$,

$$\frac{U_1}{11} = \frac{U_2}{14-\lambda_2} = B = 1$$

Hence, $\frac{U_1}{11} = 1 \Rightarrow U_1 = 11$

$$\frac{U_2}{14-\lambda_2} = 1 \Rightarrow U_2 = 14-\lambda_2$$

$$= 14 - 6.62$$

$$= 7.38$$

Hence, Eigen Vector for $\lambda_2 = \begin{bmatrix} 11 \\ 7.38 \end{bmatrix}$

If we want to normalize eigen vector,

$$n_2 = \begin{bmatrix} 11/\sqrt{11^2+7.38^2} \\ 7.38/\sqrt{11^2+7.38^2} \end{bmatrix} = \begin{bmatrix} 11/13.24 \\ 7.38/13.24 \end{bmatrix}$$

$$= \begin{bmatrix} 0.8308 \\ 0.5574 \end{bmatrix}$$

Step 5: New dataset,

Feature	sample1	sample2	sample3	sample4
a	4	8	13	7
b	11	4	5	14

1st PC	P_{11}	P_{12}	P_{13}	P_{14}
	sample1	sample2	sample3	sample4

$$P_{11} = n_1^T \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5575 & -0.8302 \end{bmatrix} \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$$

$$= (-2.23 - 2.0755)$$

$$= -4.305$$

$$P_{12} = n_1^T \begin{bmatrix} 8-8 \\ 4-8.5 \end{bmatrix} \Rightarrow (0.5575 \; -0.8302) \begin{pmatrix} 0 \\ -4.5 \end{pmatrix}$$

$$= 0 + 3.7359 = 3.7359$$

$$P_{13} = n_1^T \begin{bmatrix} 13-8 \\ 5-8.5 \end{bmatrix} \Rightarrow (0.5575 - 0.8302) \begin{pmatrix} 5 \\ -3.5 \end{pmatrix}$$

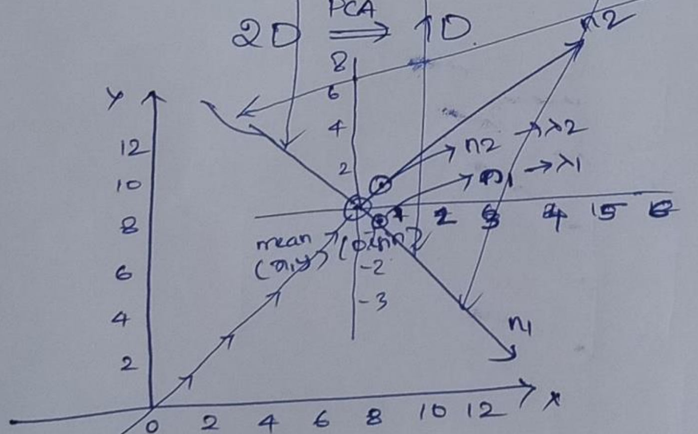
$$= 2.787 + 2.905 = \underline{5.692}$$

$$P_{14} = n_1^T \begin{bmatrix} 7-8 \\ 14-8.5 \end{bmatrix} = (0.5575 - 0.8302) \begin{pmatrix} -1 \\ 5.5 \end{pmatrix}$$

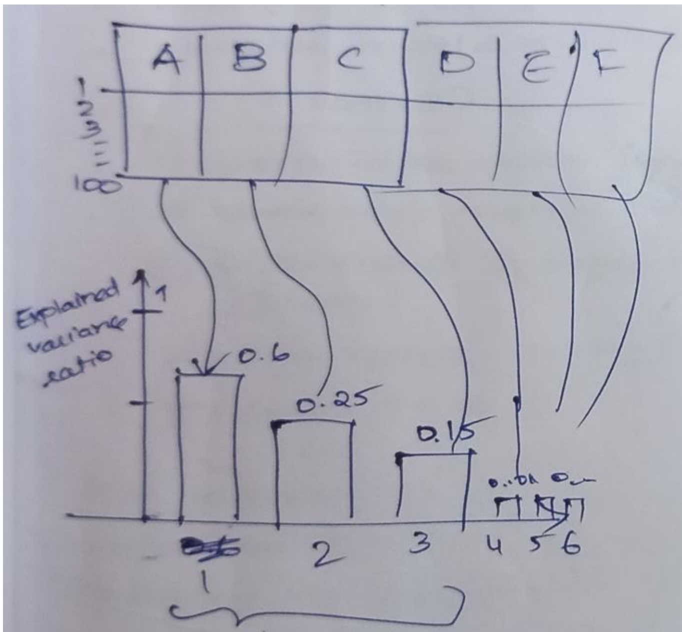
$$= -0.5575 - 4.5661$$

$$= -5.123$$

$$\boxed{PC1 = -4.305 \quad 3.7359 \quad 5.692 \quad -5.123}$$



$n_1 \perp^r n_2$



X

⇒ score plot

X

In general 80% criteria is feasible to achieve

⇓ this is 10% greater

⇒ 90% of info/ Variance achieved

$0.6+0.25+0.15 \Rightarrow 0.90$

Hence three PC is OK here.

X

PCA scores is essential to calculate classification/ regression