PCA

Date :

Steps involved

1) Find The covariance matrix of the standardized data.

2) Find the eigen values and eigen vectors for the covariance matrix.

- Eigen vectors are those vectors which remain on its own span; only getting affected it its magnitude.
- Eigen vector with highest λ ~~of~~ amounts for max variance;
- These eigen vectors are called Principal Components.
- Eigen vectors are found when for eg; 2D space gets squished onto a line or dot (1D).
- At that instance, $|A - \lambda I| = 0$.

3) Now that we have got the PCs, we have to project the original transformed data onto our PCs.

- This is achieved by dot product of X with Eigen basis.

4) Our data set which initially had d dimensions will have only m dimensions.

5) Select ~~dimensions~~ PCs which together account for atleast 90% of the variation.

Teacher's Signature :

Data	Att1	Att2
A	1	0.9
B	2	2.2
C	3	2.6

Standardisation of the given data

- Done so that all the variables are on a comparable scale

$\mu = 0 ; \sigma = 1$

Att1	Att2
-1.22	-1.38
0	0.43
1.22	0.95

Take transpose (needed as np.cov accepts features as rows)

-1.22	0	1.22
-1.38	0.43	0.95

Find covariance matrix

	A1	A2
A1	1.5	1.42
A2	1.42	1.5

Find eigen vectors for above matrix

$$A \vec{v} = \lambda \vec{v}$$

$$\begin{pmatrix} 1.5 & 1.42 \\ 1.42 & 1.5 \end{pmatrix} \vec{r} = \lambda \vec{v}$$

$$A\vec{v} - \lambda \vec{v} = 0$$

$$(A - \lambda I)\vec{v} = 0 \quad -① $$

$$\Rightarrow \det (A - \lambda I) = 0$$

$$\det \left[\begin{pmatrix} 1.5 & 1.42 \\ 1.42 & 1.5 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right] = 0$$

$$\begin{vmatrix} 1.5 - \lambda & 1.42 \\ 1.42 & 1.5 - \lambda \end{vmatrix} = 0$$

$$(1.5 - \lambda)^2 - 2.0164 = 0$$

$$\lambda^2 + 2.25 - 3\lambda - 2.0164 = 0$$

$$\lambda^2 - 3\lambda + 0.2336 = 0$$

$$\lambda = 2.92 \text{ and } 0.08$$

Put $\lambda = 2.92$ in ①

$$\begin{pmatrix} -1.42 & 1.42 \\ 1.42 & -1.42 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$v_1 = v_2$$

If $v_1 = 1$; $v_2 = 1$

Eigen vector $e_1 \rightarrow \frac{1}{\sqrt{2}} \vec{i} + \frac{1}{\sqrt{2}} j$

$$(0.707, 0.707)$$

Put $\lambda = 0.08$ in ①

$$\begin{pmatrix} 1.42 & 1.42 \\ 1.42 & 1.42 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$v_1 = -v_2$$

If $v_1 = 1$; $v_2 = -1$

Eigen vector $e_2 \rightarrow \frac{1}{\sqrt{2}} \vec{i} - \frac{1}{\sqrt{2}} \vec{j}$

$$(0.707, -0.707)$$

Teachers Signature..............................

Date

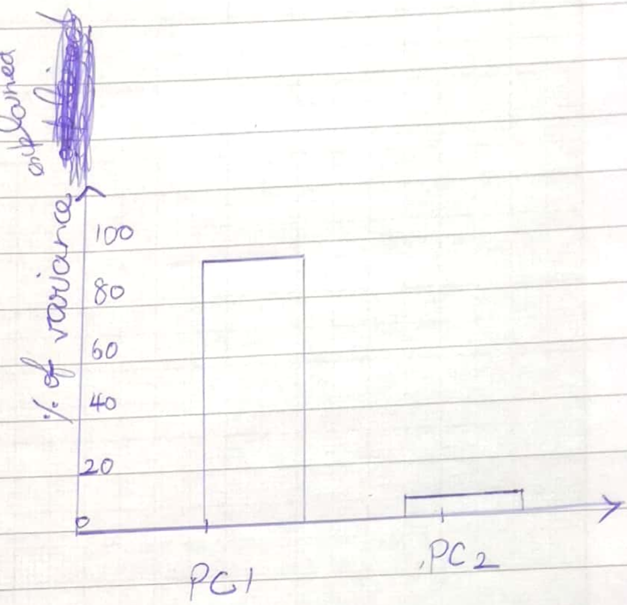
Find variance explained by each PC

$\lambda_1 = 2.92$

$$\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right) \times 100\ \% = 98\ \%$$

$\lambda_2 = 0.08$

$$\left(\frac{0.08}{2.92+0.08}\right) \times 100\ \% = 2\ \%$$



Projecting old points onto PC1

Since PC1 explains 98% of the variation, we can take PC1 alone.

Projecting a vector onto another vector is done by using dot product.

$$X\_std \cdot eig\_vect^{T}$$

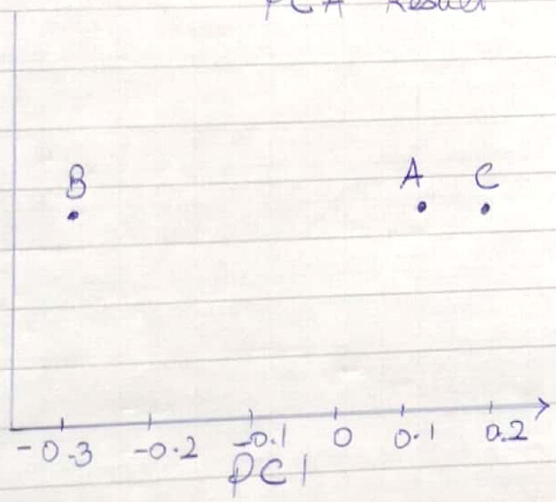
$$= \begin{pmatrix} -1.22 & -1.38 \\ 0 & 0.43 \\ 1.22 & 0.95 \end{pmatrix} \cdot \begin{pmatrix} 0.707 & -0.707 \end{pmatrix} = \begin{pmatrix} -1.22 \\ 0 \\ 1.22 \end{pmatrix} \times 0.707 + \begin{pmatrix} -1.38 \\ 0.43 \\ 0.95 \end{pmatrix} \times (-0.707)$$

$$= \begin{pmatrix} -1.22 \times 0.707 \\ 0 \times 0.707 \\ 1.22 \times 0.707 \end{pmatrix} + \begin{pmatrix} -1.38 \times -0.707 \\ 0.43 \times -0.707 \\ 0.95 \times -0.707 \end{pmatrix}$$

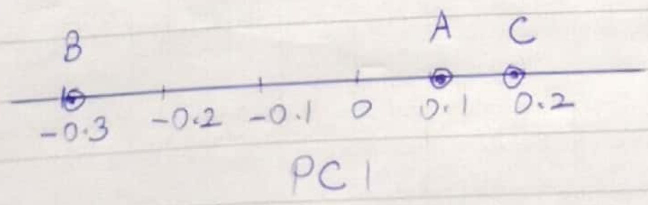
Teachers Signature..............................

$$= \begin{pmatrix} 0.1110 \\ -0.304 \\ 0.1930 \end{pmatrix}$$

PCA Result



PCA Result



We started with 2 dimensions. Now we have reduced it to a single dimension still explaining 98% of the variation in the data.