Principal Component Analysis (PCA)

- PCA can be abbreviated as Principal Component Analysis
- PCA comes under the Unsupervised Machine Learning category
- Reducing the number of variables in a data collection while retaining as much information as feasible is the main goal of PCA. PCA can be mainly used for Dimensionality Reduction and also for important feature selection.
- Correlated features to Independent features

Why Do We Need PCA?

- When a computer is trained on a big, well-organized dataset, machine learning often excels. One of the techniques used to handle the curse of dimensionality in machine learning is principal component analysis (PCA). Typically, having a sufficient amount of data enables us to create a more accurate prediction model since we have more data to use to train the computer. But working with a huge data collection has its own drawbacks. The curse of dimensionality is the ultimate trap.
- The title of an unreleased Harry Potter novel does not refer to what happens when your data has too many characteristics and perhaps not enough data points; rather, it refers to the curse of dimensionality. One can use dimensionality reduction to escape the dimensionality curse. Having 50 variables may be cut down to 40, 20, or even 10. The strongest effects of dimensionality reduction are found here.
- Overfitting issues will arise while working with high-dimensional data, and dimensionality reduction will be used to address them. Increasing interpretability and minimizing information loss. aids in locating important characteristics. Aids in the discovery of a linear combination of varied sequences.

When to use PCA?

- Whenever we need to know our features are independent of each other
- Whenever we need fewer features from higher features

Dimensionality Reduction Work in Real-Time Application

Assume there are 50 questions in all in the survey. The following three are among them: Please give the following a rating between 1 and 5:

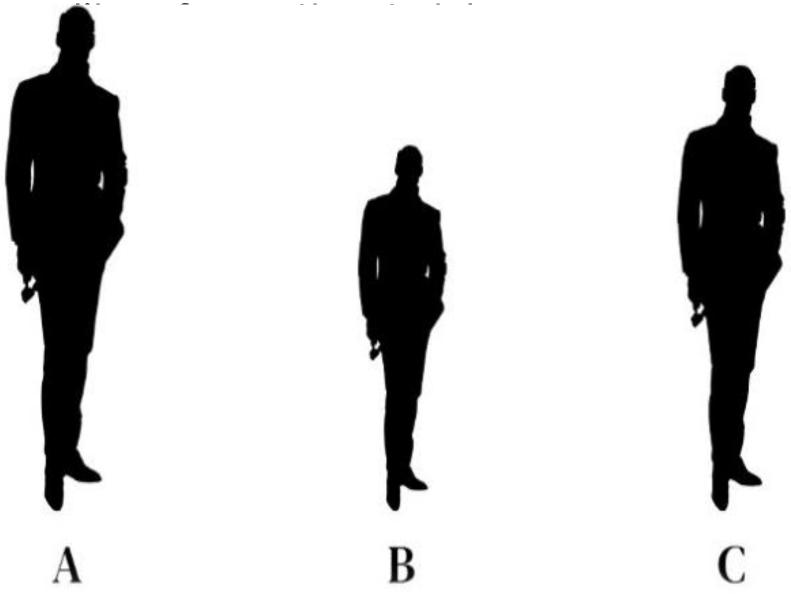
- I feel comfortable around people
- I easily make friends
- I like going out

These queries could appear different now. There is a catch, though. They aren't, generally speaking. They all gauge how extroverted you are. Therefore, combining them makes it logical, right? That's where linear algebra and dimensionality reduction methods come in! We want to lessen the complexity of the problem by minimizing the number of variables since we have much too many variables that aren't all that different. That is the main idea behind dimensionality reduction. And it just so happens that PCA is one of the most straightforward and popular techniques in this field. As a general guideline, maintain at least 70–80 percent of the explained variation.

Intuition behind PCA

- Let's assume we are playing a mind gar
need to find the tallest person.

Person	Height
A	145
B	160
C	185



Intuition behind PCA

- Now change the scenario, Can you guess who's who? It's tough when they are very similar in height.

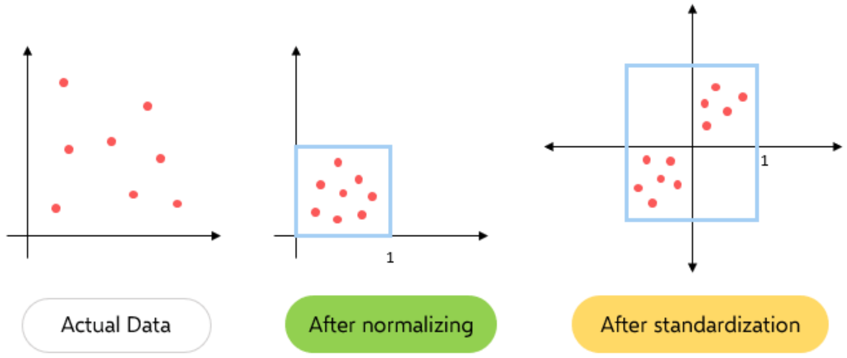
Person	Height
D	172
E	173
F	171

- Because of how much their heights vary, we previously had no issue telling a 185cm person from a 160cm and a 145cm person. Similar to this, our data contains more information when its variance is bigger. This explains why the terms PCA and maximum variance are frequently used together.

Basic Terminologies of PCA

Before getting into PCA, we need to understand some basic terminologies,

- **Variance** – for calculating the variation of data distributed across dimensionality of graph
- **Covariance** – calculating dependencies and relationship between features
- **Standardizing data** – Scaling our dataset within a specific range for unbiased output



Basic Terminologies of PCA

- **Covariance matrix** – Used for calculating interdependencies between the features or variables and also helps in reduce it to improve the performance
-

data value of X

mean value of X

data value of Y

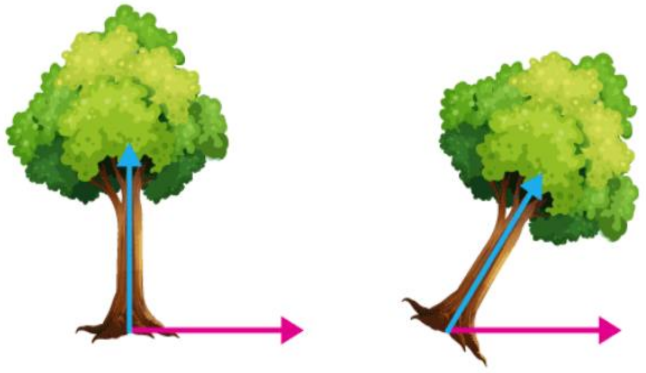
$$Cov(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

mean value of Y

Number of data values

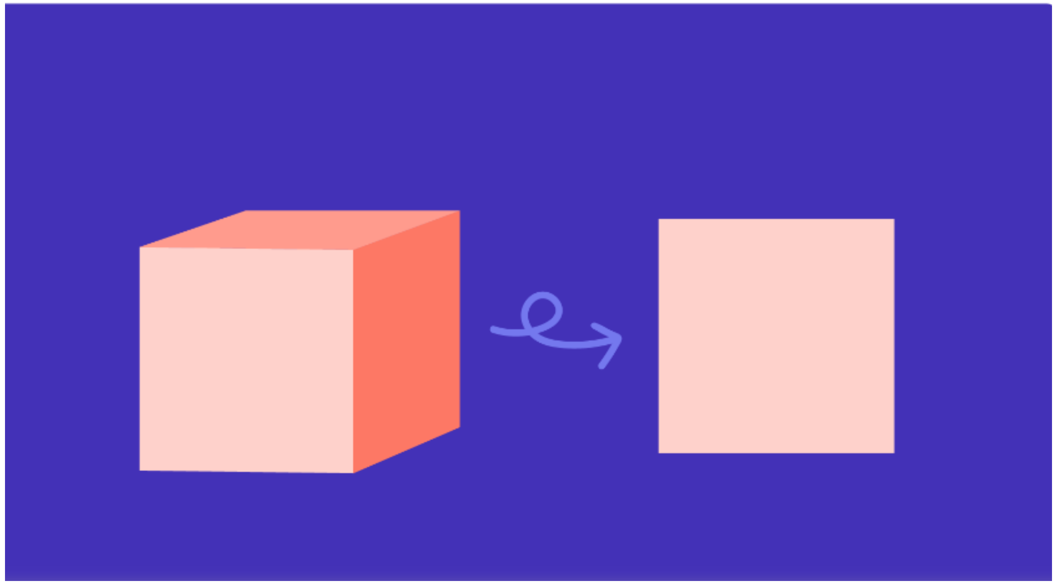
Basic Terminologies of PCA

- **Eigen Values and Eigen Vectors** – Eigenvectors' purpose is to find out the largest variance that exists in the dataset to calculate Principal Component. Eigenvalue means the magnitude of the Eigenvector. Eigenvalue indicates variance in a particular direction and whereas eigenvector is expanding or contracting X-Y (2D) graph without altering the direction.
- In this shear mapping, the blue arrow changes direction whereas the pink arrow does not. The pink arrow in this instance is an eigenvector because of its constant orientation. The length of this arrow is also unaltered, and its eigenvalue is 1. Technically, PC is a straight line that captures the maximum variance (information) of the data. PC shows direction and magnitude. PC are perpendicular to each other.



Basic Terminologies of PCA

- **Dimensionality Reduction –** Transpose of original data and multiply it by transposing of the derived feature vector. Reducing the features without losing information.



How does PCA work?

The steps involved for PCA are as follows-

- Original Data
- Normalize the original data (mean =0, variance =1)
- Calculating covariance matrix
- Calculating Eigen values, Eigen vectors, and normalized Eigenvectors
- Calculating Principal Component (PC)
- Plot the graph for orthogonality between PCs

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)

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

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

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

$$= \frac{4^2+0+5^2+1^2}{3} = \frac{\cancel{8}16+0+25+1}{3}$$

$$= \frac{\cancel{2}42}{3} = \underline{14}.$$

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

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

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

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

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

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

$$= -11$$

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

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

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

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

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

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

Hence covariance matrix can be

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

Step4:-

calculate Eigen value, Eigen vector, Normalized Eigen vector.

Inorder calculate Eigen value,

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

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

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

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

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

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

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

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

'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, $\lambda = 30.38$

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

S: Covariance matrix

λ_1: 30.38

I: Identity matrix

U_1: Eigen vector of $\lambda 1$

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

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

Hence,

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

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

$$\begin{pmatrix} -16.38 & -11 \\ -11 & -7.38 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$-16.38\, u_1 - 11 u_2 = 0 \rightarrow (1)$$

$$-11\, u_1 - 7.38 u_2 = 0 \rightarrow (2)$$

So, from this if we need to calculate u_1 & u_2

$$(1) \times 7.38 \Rightarrow 120.88 u_1 - \cancel{81.18} u_2 = 0$$

$$(2) \times -11. \Rightarrow +121\, u_1 + \cancel{81.18}\, u_2 = 0$$

$$0.12 u_1 = 0$$

$$\boxed{u_1 = 0}$$

then, apply u_1 in (1), then

$$-16.38 \times 0 - 11 u_2 = 0$$

$$\boxed{u_2 = 0}$$

this can't be possible, hence

$$\begin{bmatrix} (14-\lambda_1) & (-11) \\ (-11) & (28-\lambda_1) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

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

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

from (a),

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

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

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

Assume $\underline{A=1}$,

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

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 $\lambda_1 \Rightarrow \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.38 \end{bmatrix}$

From this, we are going to calculate PCs

Then if we want to normalize the eigen vectors,

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

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

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 vectors,

$$n_2 = \begin{bmatrix} 11/\sqrt{11^2+7.38^2} \\ 7.38/\sqrt{11^2+7.38^2} \end{bmatrix} = \begin{bmatrix} 11/13.24 \\ 7.38/13.24 \end{bmatrix} = \begin{bmatrix} 0.8308 \\ 0.5574 \end{bmatrix}$$

We are going to calculate the normalized eigenvector

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

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

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

$$= (-2.23 - 2.0755)$$

$$= -4.305$$

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

$$= 0 + 3.7359 = 3.7359$$

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

$$= 2.787 + 2.905 = 5.692$$

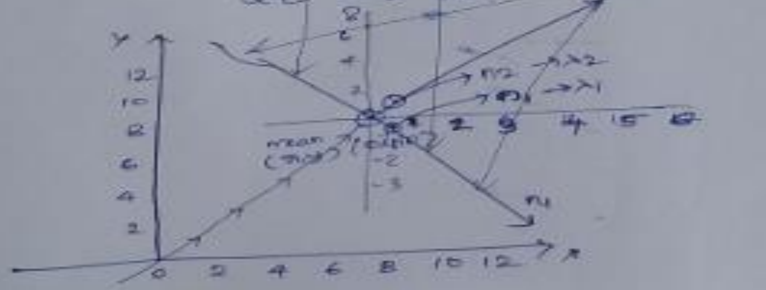
$$P_{14} = n_1^T \begin{bmatrix} 7-8 \\ 14-8.5 \end{bmatrix} = (0.5575 \; -0.8302) \begin{pmatrix} -1 \\ 5.5 \end{pmatrix}$$

$$= -0.5575 - 4.5661$$

$$= -5.123$$

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

2D $\xrightarrow{PCA}$ 1D

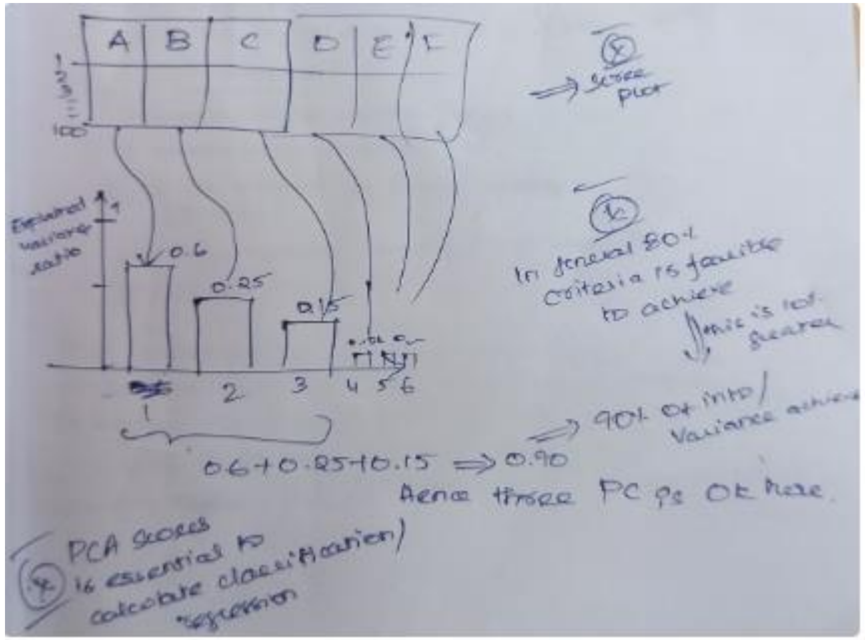


$n_1 \perp n_2$

Hence PCA is calculated and visually we can see how PC are orthogonal to each other.

How Many PCAs are Needed for Any Data?

- PCA has maximum variance (information), which will be good to select.
- Eigenvalues are used to find out which PCA has a maximum variance.



Advantage for Principal Component Analysis

- Used for Dimensionality Reduction
- PCA will assist you in eliminating all related features, sometimes referred to as multi-collinearity.
- The time required to train your model is now substantially shorter because to PCA's reduction in the number of features.
- PCA aids in overcoming overfitting by eliminating the extraneous features from your dataset.

Disadvantage for Principal Component Analysis

- Useful for quantitative data but not effective with qualitative data.
- Interpretation of PC is difficult from original data

Application for Principal Component Analysis

- Computer Vision
- Bio-informatics application
- For compressed images or resizing of the image
- Discovering patterns from high-dimensional data
- Reduction of dimensions
- Multidimensional Data – Visualization