Dimensionality Reduction:-

The number of input features present in the given dataset is known as dimensionality and the process to reduce these features is called dimensionality reduction.

A dataset contains a huge number of input features which makes the predictive modeling task more complicated. Because it is very difficult to visualize or make predictions with high number of features, for such cases dimensionality reduction techniques are required to use.

'It is a way of converting the higher dimensions dataset into lesser dimensions dataset ensuring it provides similar information'.

Principal Component analysis (PCA).

→ dimensionality reduction technique used in industry.

Original dataset $M \times N$ $\quad k < N$

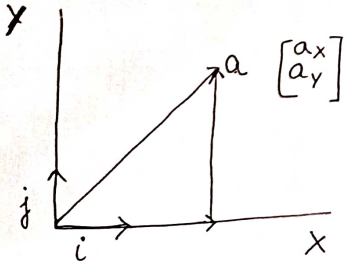
Transformed dataset $M \times k$

→ PCA is a statistical procedure to convert observations of possibly correlated variables into new principal components such that the new principal components are

a) New PC's are uncorrelated with each other.
b) Linear combinations of the original variables.
c) Capture maximum information

Fundamental Building block of PCA : Basis

Basis is a unit in which we express the vectors of a matrix.



$a = a_x i + a_y j$

i : 1 unit movement in +ve X $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$

j : 1 unit movement in +ve Y $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$a_x i = \begin{bmatrix} a_x \\ 0 \end{bmatrix}$

$a_y i = \begin{bmatrix} 0 \\ a_y \end{bmatrix}$

Basis vector $\{i\}$

Basis vector $\{j\}$

* Basis vector can be used to get to any point in space

Linear Combination -

ex

Height	Weight
165	55
155	71
165	88
160	105
160	94

Basis vectors = $\left\{ \begin{bmatrix} 1\,cm \\ 0\,kg \end{bmatrix} \begin{bmatrix} 0\,cm \\ 1\,kg \end{bmatrix} \right\}$

↓ represent movement along weight & height axis

$$\begin{bmatrix} 165 \\ 55 \end{bmatrix} = \underbrace{165 \begin{bmatrix} 1\,cm \\ 0\,kg \end{bmatrix}}_{\text{scaling}} + \underbrace{55 \begin{bmatrix} 0\,kg \\ 1\,kg \end{bmatrix}}_{\text{scaling}}$$

adding

Scaling and adding of certain vectors gives a linear combination of those vectors

Change of Basis

New Basis Representation (cm) = M × Old Basis Representation (ft)

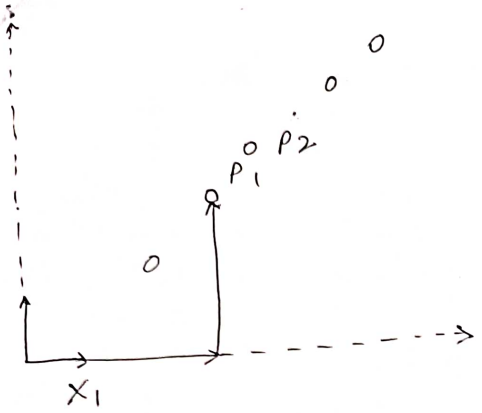
Or New Basis Representation (ft) = M^{-1} × Old Basis Representation (cm)

ex Basis: 1 ft 1 cm

Length: 1 ft 30.48 cm

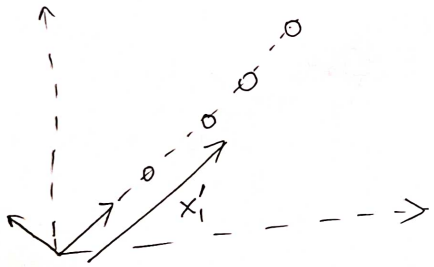
5.4 ft 164.6 cm → 5.4 (ft) = $\frac{30.48 \times 5.4}{NR}$ = M × O R

* Same information can be represented with different set of basis vector.



Point	X	Y
P_1	X_1	Y_1
P_2	X_2	Y_2
P_3	X_3	Y_3
P_4	X_4	Y_4
P_5	X_5	Y_5
P_6	X_6	Y_6

PCA can allow us to created new set of basis vectors.
Rotate ~~the axis~~ original set of basis vectors such that the original x axis is rotated counterclockwise & will coincide with line joining points



Point	X_{new}	Y_{new}
P_1	X_1'	0
P_2	X_2'	0
P_3	X_3'	0
P_4	X_4'	0
P_5	X_5'	0
P_6	X_6'	0

All representations are covered in x axis

* A 2-D dataset has been simplified to a 1-D dataset. All info is captured in new x axis.

How to find Basis

→ Variance (information)
→ The importance of a column can be checked by checking its variance values. If a column has more variance, then this column will contain more info

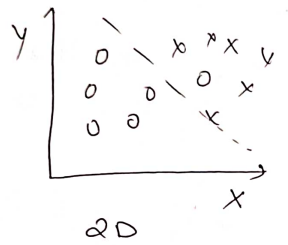
Height	Weight	Age
165	55	22
155	71	22
165	88	22
166	105	22
160	94	22
Variance 14.0	311.44	0

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$$

Linear Discriminant Analysis (LDA) –

→ dimensionality reduction technique
→ used for supervised classification problems.
→ project the features in higher dimension space into a lower dimension space.
→ whenever there is a requirement to separate two or more classes having multiple features efficiently, the LDA is considered the most common technique.
⇒ transforms 2D and 3D into 1D plane.

ex



2D

LDA creates new axis by using following criteria –

1) It maximizes the distance between mean of two classes
2) Minimizes variance within individual class

* New axis will increase the separation between the data points of the two classes and plot onto new axis