Principal Component Analysis: [PCA]

Dimensionality Reduction

Why we are Use PCA?

Some reasons are Used to perform PCA.

① Curse of Dimensionality:

Suppose we have 500 features in the dataset.

we want to train ML models with these features.

Suppose,

3 features → M_1 6 features → M_2 15 features → M_3 50 features → M_4

Acc1 > Acc2 ↑↑ > Acc3 ↑↑ < Ac4 ↓↓   So, model is Over feeded.

100 feature → M_5 500 features → M_6

< Acc5 ↓↓ < Acc6 ↓↓

Note that, The increases in features means Some features are not important at all. So, the accuracy will decreases.

The features would be,

→ House Size

→ No of bedrooms.

→ No of bathrooms. like that.

② <u>Model Performance Degrade:</u>

Suppose Consider a human being, The person thinks about not only the prices of the room appartment and also thinks about the facilities which comes along the room.

feature 1

$\boxed{\text{Loc A}}$ $\boxed{\text{Price}}$

2 Bhk ← 450k - 500k

3 Bhk ← 500k - 600k

Near Beach ← ↑↑↑

near to Celebrity house ← ↑↑↑

Grocery shop ← Its is not more important but machine also train this.

School ← ↑↑↑

↗ This says curse of dimensionality

Sometime, The domain expert also Confused to give accurate rate. This will shows the many features are feeded into the machines. So, the model is Overfeeded.

How to remove and prevent this Curse of dimensionality?

Two different ways are,

→ Feature Selection.

→ ~~Fe~~ Dimensionality Reduction (PCA).

Feature Selection	Dimensionality Reduction (PCA)
⇓	⇓
Take out and Select the important features.	Feature Extraction
	⇓
	It says we will derive the new feature from the set of features.

The feature extraction would be as,

<u>original</u>

$f_1 \quad f_2 \quad f_3 \quad O/P$

It must be lesser dimension ⇓ Feature Extraction

$D_1 \quad D_2 \quad O/P$

This is all about the feature engineering on PCA.

We will see about the feature selection and feature Extraction.

Feature Selection Vs Feature Extraction:

Why Dimensionality Reduction?

→ Prevent → Curse of dimensionality.

→ Improve the performance of model.

→ Visualise the data → Understand the data.

$\boxed{3d}$ $\boxed{2d}$ visualization.

$\boxed{100d}$

↓

$\boxed{3d}$ (or) $\boxed{2d}$ then visualize

Feature Selection:

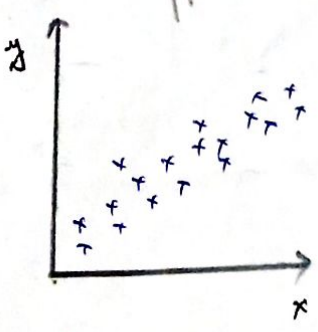
Feature Selection is the technique to select the most important feature from set of features.

Consider,

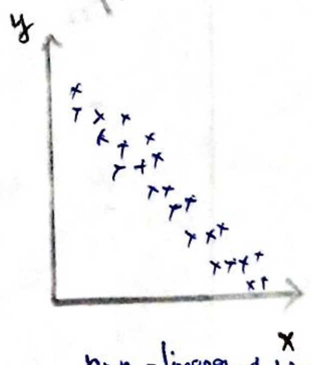
I/p	O/p
x	y
-	-
-	-
-	-
-	-

$\boxed{\begin{matrix} x\uparrow\ y\uparrow \\ x\downarrow\ y\downarrow \end{matrix}}$ $\boxed{\begin{matrix} x\downarrow\ y\uparrow \\ x\uparrow\ y\downarrow \end{matrix}}$

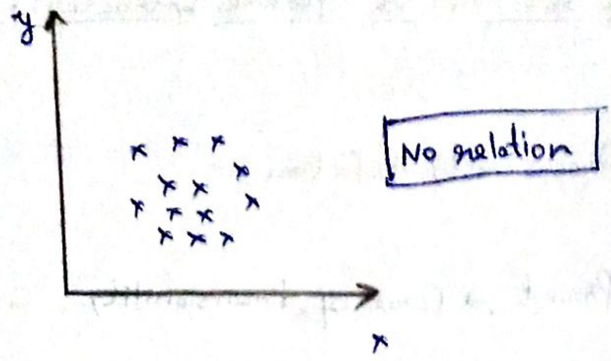
⇑ ⇑



Linear relation



non-linear relation



$$Cov(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x}) * (y_i - \bar{y})}{N-1} = +ve \text{ (or) } 0 \text{ (or) } -ve$$

$$\text{Pearson Correlation} = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y} = \boxed{-1 \text{ to } 1}$$

More towards +1 the feature x and y are more relationship.

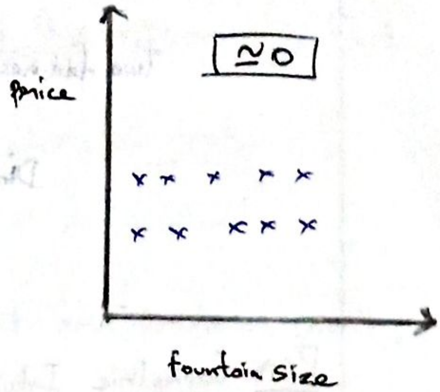
More towards -1 the feature x and y are no relationship.

Is equal to 0. The two features are no-correlated.

<u>Example,</u>

Dataset housing

House size	fountain size	Price
-	-	-
-	-	-



⇓ This is linear and more Correlated.

⇓ This is not an linear and Correlated.

Therefore, The house size is very important feature. Select that house size and neglect the fountain size for an analysis.

<u>Feature Extraction:</u>

Feature Extraction is the technique to extract the new feature from set of features.

Consider that,

Room Size	No of Rooms	Price
-	-	-
-	-	-
-	-	-

⇓ Transformation applied to extract new feature

House Size	Price
-	-
-	-

Two features → One feature

Dimensionality Reduction

PCA Geometric Intuition:

Consider the housing dataset,

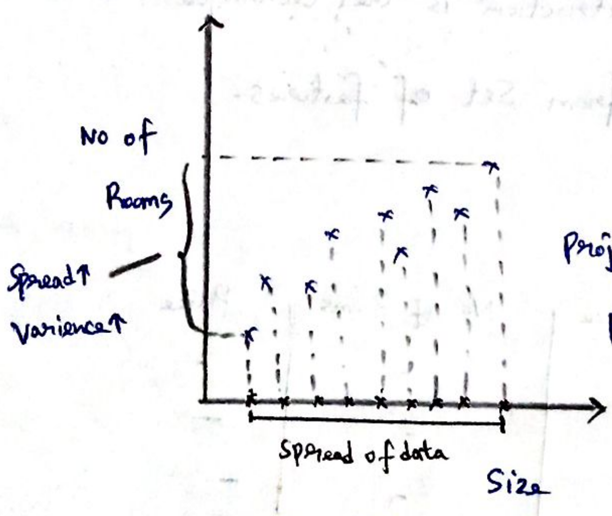
Size House	No of Room	Price (o/p)
-	-	-
-	-	-

PCA,

2 dimensions → 1 dimensions

Obiviously, we know the size house and no of Room are the linear relationship.

So, plot the points,



Projecting this points on 1D plane

Spread ↑

Varience ↑

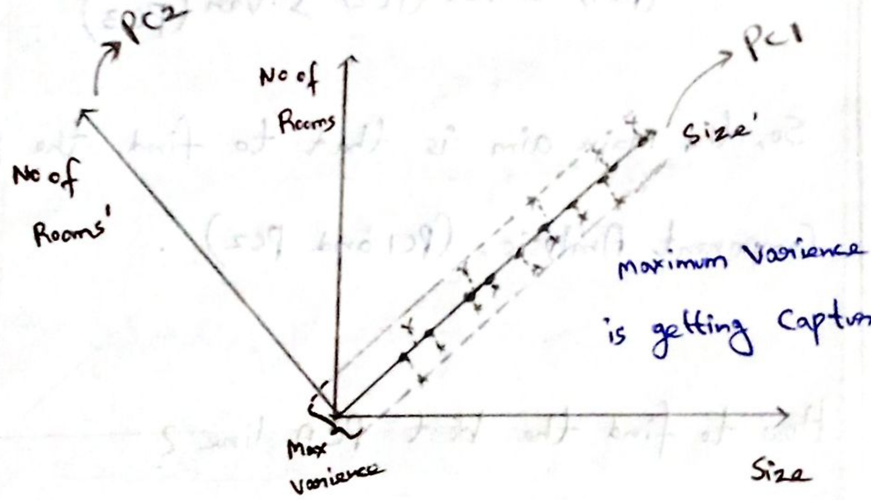
So, the main disadvantages of this approach is that the No of rooms information will be lost.

So, the model unable to predict and perform well.

→ Loss of information (No of Rooms)

Therefore, the different approach was used in PCA.

That different approach is nothing but the some transformation is applied to capture maximum varience of data.



How the axis line Created?

It is Created with the help of some transformation is nothing but Eigen decomposition on Matrix.

So, in this approach we tried 2D → 1D.

Therefore much information is not lost.

2 Dimensions

PC1, PC2

Var (PC1) > Var (PC2)

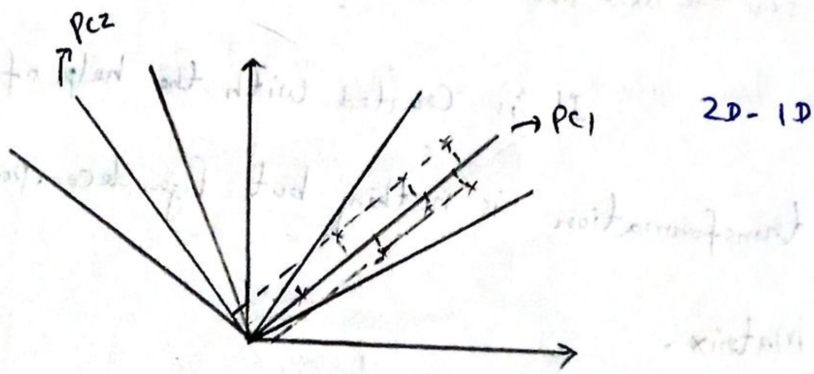
3 dimensions

PC1, PC2, PC3

Var (PC1) > Var (PC2) > Var (PC3).

So, the main aim is that to find the Prinicipal Component Analysis. (PC1 and PC2).

How to find the best PCA line?



2 Best PCA.

The finding of PCA is very ~~soor~~ simple to check the which principal component lines captures more varience of data points.

That ~~data~~ PCA lines are selected.

To get the best principal component which captures maximum varience.

Suppose we have,

$$3D \rightarrow 1D$$

$$\Downarrow$$

we know, $\underbrace{PC_1, PC_2, PC_3}_{\to 1D}$

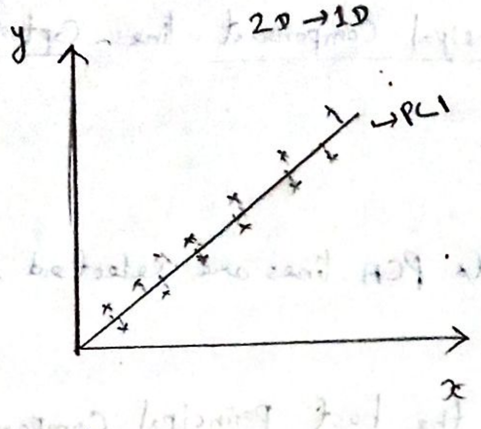
$$Var(PC_1) > Var(PC_2) > Var(PC_3)$$

Suppose, $3D \rightarrow 2D$,

$$\underline{PC_1, PC_2}, PC_3$$

↳ This two taken and convert to 2D.

<u>Maths Intuition behind PCA algorithm.</u>

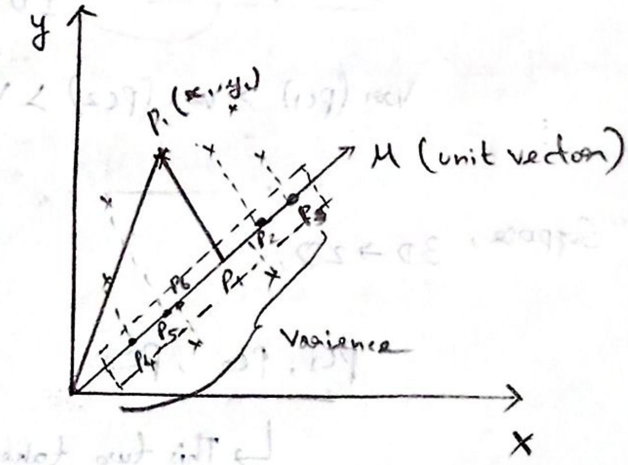


The two important things are,

1. projection.

2. Cost function related to Varience.

Consider that,



Suppose I want to project $P_1 (x_1, y_1)$ on unit vector,

$$\text{Proj}_{P_1} \mu = \frac{P_1 \cdot \mu}{\|u\|} \qquad \text{where, } \|u\| = 1$$

Therefore, $\boxed{Proj_{P_1}\mu = P_1 \cdot \mu}$ => Give Scalar value.

$\Downarrow$

P_1'

Scalar refers to only Magnitude.

So, In our case we want to find the varience. (i.e) varience which comes under the distance. Therefore we say magnitude. (Scalar).

So, Computing every points we need to project. It looks like,

$\boxed{P_0', P_1', P_2', P_3', P_4', \ldots P_n'}$ → Scalar value
$\Downarrow$
Varience (distance) spread.

The projecting point on the Unit vector we say P_i'.

Let's take,

$\boxed{P_0', P_1', P_2', P_3', P_4', \ldots P_n'}$

We use different notation as,

$x_0', x_1', x_2', x_3', x_4' \ldots x_n'$

Goal:

Find the best unit vector which captures maximum Varience.

$\bar{x}$ → Sample Mean.

So, Max Varience $= \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$

$\Downarrow$

So, this is the cost function of PCA.

Then, we Cannot obiviously try to find which unit Vector line is Capturing maximum varience.

Inorder to find that, we Something Used Called as Eigen Values and Eigen Vectors.

Eigen Vectors and Eigen Values:

To Calculate this,

- Firstly, Covarience matrix between features need to be find.
- Then the eigen Vectors and eigen Values will found (or) find out from this Covarience Matrix.
- Eigen Vector → Eigen Value → It shows the Magnitude of eigen Vector. Using this to Capture the Maximum Varience.

The Second Point shows Some mathematical equation to find eigen Value and eigen vector. Something are the,

$\boxed{Av = \lambda v}$ → It is nothing but linear Transformation of matrix.

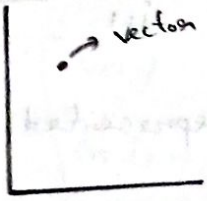
Using these, to find out the maximum varience of the datapoints Captured.

Eigen vectors and Eigen Values: [Linear Transformation]

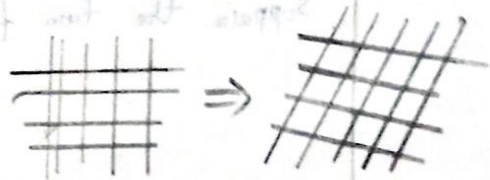
[Eigen decomposition of Covarience Matrix]

⇓

Eigen Vector and Eigen Values.



The linear transformation would be,



See the difference in some website.

[ ] * [V] = λ * V

↓

Eigen value

So,

$$\boxed{A * V = \lambda * V}$$

⇓ So this give

Eigen Vector

↓ which gives

Maximum magnitude

↳ from this we get principal components and get the maximum varience Captured.

The graphical visualization of this available in,

https://shad.io/Matvis/

Steps to Calculate Eigen Value and Vectors:

1. Firstly, Covarience of features.

Consider,

$$\boxed{x \quad y} \quad z \rightarrow \text{dependent feature}$$

$$Cov(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

Suppose the two feature, represented as 2x2,

	x	y
x	Var(x)	Cov(x,y)
y	Cov(y,x)	Var(y)

why var(x), because Cov(x,x),

$$Cov(x,x) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(x_i - \bar{x})}{N-1}$$

$$\cancel{= \sum_{i=1}^{n} \frac{x_i^2 - x_i\bar{x} + x_i^2 - x_i\bar{x} + x_i^2}{N-1}}$$

$$= \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{N-1} \Rightarrow Var(x)$$

That's why we put Cov(x, x) and Cov(y, y) are mentioned

at <u>Var(x)</u> and <u>Var(y)</u>.

Suppose, the same as 3x3 matrix

	X	Y	Z
X	Var(x)	Cov(x,y)	Cov(x,z)
Y	Cov(y,x)	Var(y)	Cov(y, z)
Z	Cov(z,x)	Cov(z,y)	Var(z)

Then Consider,

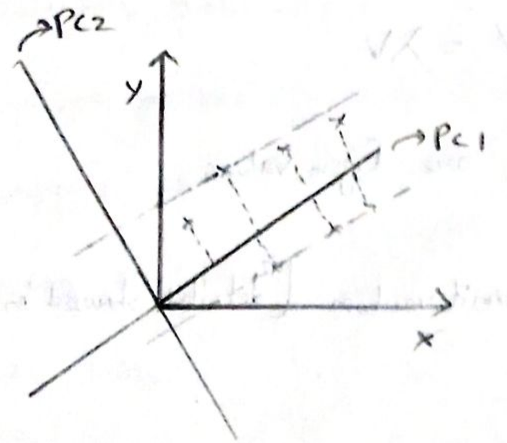
$$A = \begin{bmatrix} Var(x) & Cov(x,y) \\ Cov(y,x) & Var(y) \end{bmatrix}$$

$$\boxed{A.V = \lambda.V}$$

Where λ denotes λ_1, λ_2 i.e (f_1 and f_2)

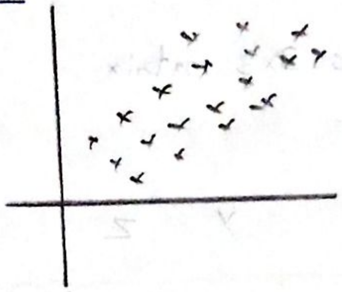
$$\underset{\downarrow}{\lambda_1} \to pc1 \quad \underset{\downarrow}{\lambda_2} \to pc2$$

Then the $\boxed{\lambda_1, \lambda_2}$ are the Eigen Values.



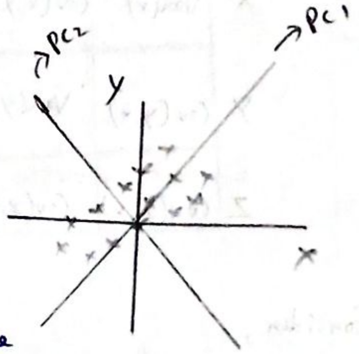
Overall steps for Calculations:

1] Change dimension



$2D \rightarrow 1D$

2) Standardize the data.



Once, standardization is applied. It comes to the center.

3] Covariance matrix x and y

$$A = \begin{array}{c|cc} & x & y \\ \hline x & Var(x) & Cov(x,y) \\ y & Cov(y,x) & Var(y) \end{array}$$

4] Find out Eigen vectors and Value:

$$AV = \lambda V.$$

λ_1, λ_2 are Eigen values.

$V \rightarrow$ Unit vector [detailed showed in graph in website]

$\boxed{\lambda_1, \lambda_2}$ $\lambda_1 - PC_1$

$\lambda_2 - PC_2$

λ_1 denotes the Magnitude of eigen vector.

$\Downarrow$ In this way

Capture maximum Varience.

Suppose,

$\boxed{\text{Change } 3D \rightarrow 2D}$

λ_1 λ_2 λ_3

$\downarrow$ $\downarrow$ $\downarrow$

PC_1 PC_2 PC_3

$\boxed{1D}$ $\boxed{1D}$ $\Rightarrow 2D$

$\boxed{\text{Change } 3D \rightarrow 1D}$

λ_1 λ_2 λ_3

$\boxed{1D}$

$\boxed{\text{Change } 2D \rightarrow 1D}$

λ_1, λ_2

$\downarrow$ $\downarrow$

PC_1 PC_2

$\boxed{1D}$

Likewise, we Change the dimension on PCA.