PCA → Principle Component analysis

PCA tries to find the linear Combination of Variables which contain much information by looking at the Variance using Orthogonal linear projections.

PCA is a data Reduction technique that Transforms large no of Correlated Variables to a smaller Set of Uncorrelated Variables Called Principle Components. Applied majorly in Image processing

Advantages of PCA

→ Dimensionality Reduction
→ Avoidance of Multicollinearity
→ Variables are ordered in Terms of standard error
→ Overfitting Mitigation.

---

3 Main steps

Ex bike & bicycle.

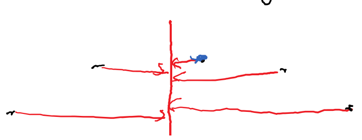
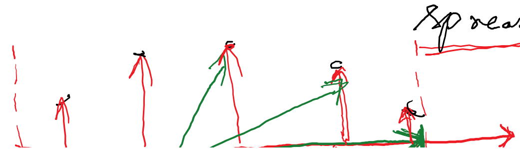
→ Feature Selection
→ Feature reduction/Extraction
→ Dimensionality reduction

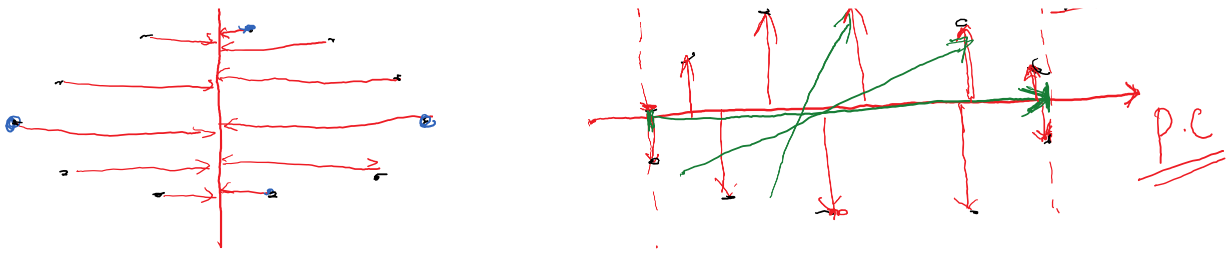
Concept → We measure data in terms of principle axis rather than regular X Y & Z axis.

→ What are principle Components? → They are underlying structures in the data.

→ They are the directions where the Variation is more & the data is more spread out.

Spread out = large variance → PC is the line in the direction of max spread out.



(PC) which is having most variance ⇒ The line in the direction of maximum spread

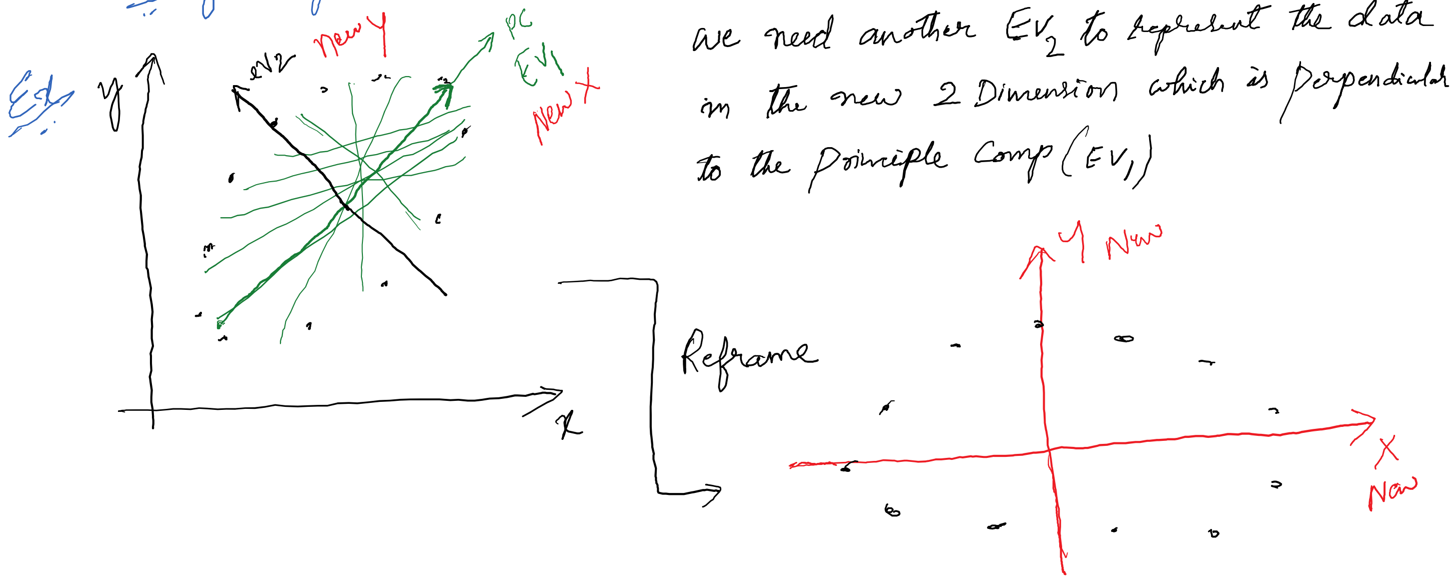
---

Instead of lines we have Eigen Vectors & Eigen values.
They represent how much variation is present & in what direction.

The Eigen vector with the maximum Eigen Value is the P.C.

No of Eigen Vectors = Dimension of the dataset ∴ 2D = 2 EV'S, 3D = 3 EV'S.

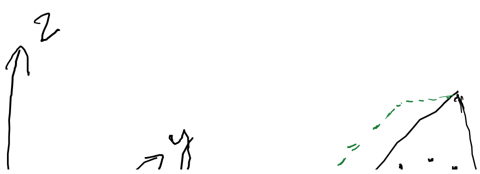
→ Eigen Vectors put the data in to new dimension, which is equal to the no of original dimension

Eg: 

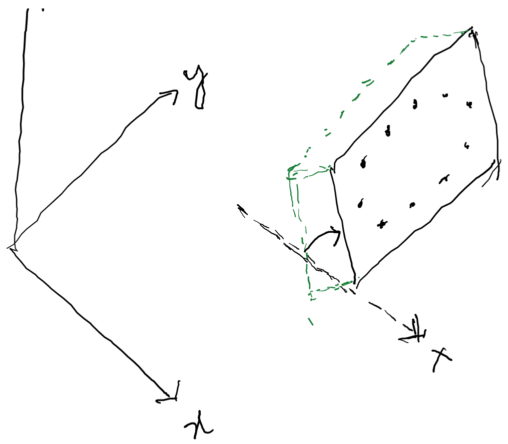
we need another EV_2 to represent the data in the new 2 Dimension which is perpendicular to the Principle Comp (EV_1)

There is no change in the data, we are looking at it from a different angle (How it helps in Analysis ?)

Dimensionality Reduction → Reduces data down in to basic Components by removing any unnecessary Variables

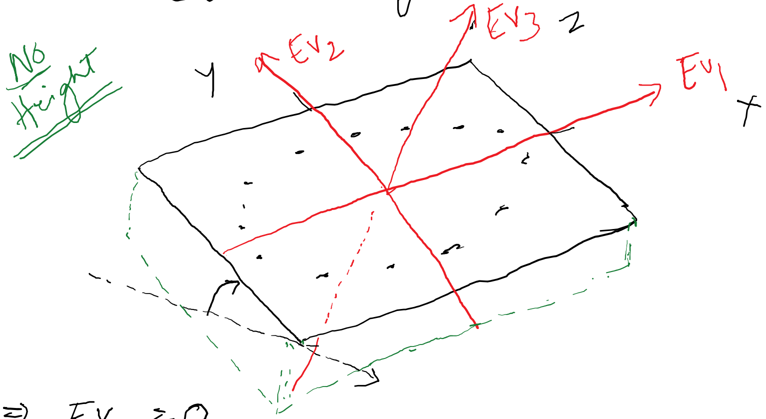


Oval points are on a plane having length & width but no Height.



length & width but no Height.

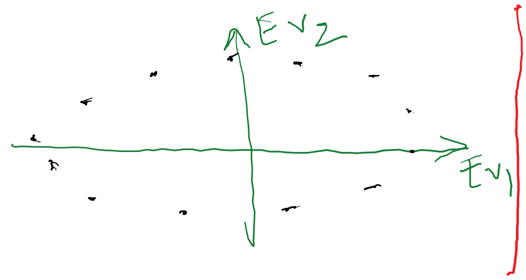
⇒ 3D ⇒ 3 eigen Vectors.



⇒ Since we don't have height ⇒ $EV_3 = 0$

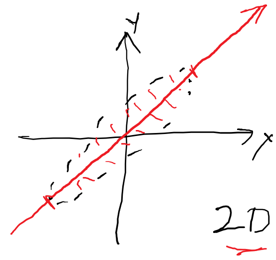
We Eliminated the EV that was of no use ⇒ 3D represented in 2D

⇒

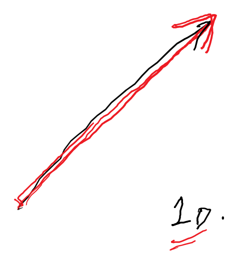


Z is removed

Ex



⇒



Imagin it had some height also ⇒ x, y, z = (10, 8, 0.1)

0.1 in z gives very less information than 10 & 8 (x & y)

So many dimensions having less or no information can be eliminated

---

Few terminologies

→ **Dimensionality** - No of Random Var / features / Columns

→ **Correlation** → How strongly 2 var are related to each other (+1 to -1)

→ **Orthogonal** → Uncorrelated Variables ⇒ Correlation = 0

→ **Eigen Vectors** → Non Zero Vector (v) of Square matrix (A)

v is Eigen Vector if Av is a Scalar Multiple of v

$$\Rightarrow \quad Av = \lambda v$$

v is eigen Vector & λ is eigen Value

V is eigen vector & λ is eigen value

→ Co-variance matrix → Matrix of co-variance btn 2 pairs.

The $(i, j)^{th}$ element is the co-var btn i^{th} & j^{th} variable

---

Principle of PCA → Linear combination of optimally weighed observed variables.

PCA on 2D → Normalize the data

→ Subtracting respective means from the no's in the respective columns.

⇒ Centered to $\mu = 0$.

→ Calculate Co-variance Matrix

$$\text{Covar mat} = \begin{array}{c|cc} & x_1 & x_2 \\ \hline x_1 & Var[x_1] & Cov[x_1, x_2] \\ x_2 & Cov[x_2, x_1] & Var[x_2] \end{array}$$

Where

$Var[x_1] = cov[x_1, x_1]$ & $Var[x_2] = Cov[x_2, x_2]$

→ Calculate the Eigen Values & Eigen Vectors. (Math)

$$\boxed{\text{determinant}(\lambda I - A) = 0}$$

I = Identity matrix

$$\boxed{(\lambda I - A)v = 0.}$$

→ Choose components from feature vectors → Take significant components

→ Forming a principle component.

⇒ New Data = Feature vector^T × Scaled data^T

→ Matrix containing the principle components.

Feature Vector → Matrix formed using the EV's we selected to keep

Feature vector → Matrix formed using the EV's we selected to keep

Scaled Data → Scaled version of original dataset.

* No of PC's can be decided by a "Scree plot" similar to Elbow in K-means.