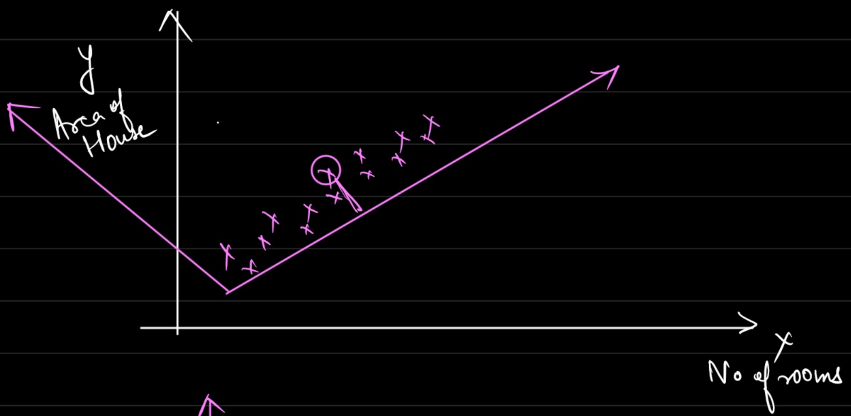
* Mathematical Explanation of PCA



$$\text{Proj of } P_1 \text{ on } u = \frac{P \cdot u}{\|u\|}$$

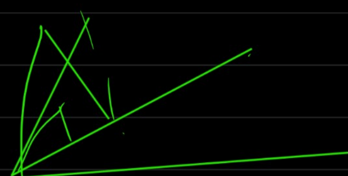
$$\text{Proj of } P_1 \text{ on } u = P \cdot u$$

$\Downarrow$

Scaler value

$\Downarrow$

Projection.



$d_0, d_1, d_2, d_3 \ldots d_n$

$\Downarrow$

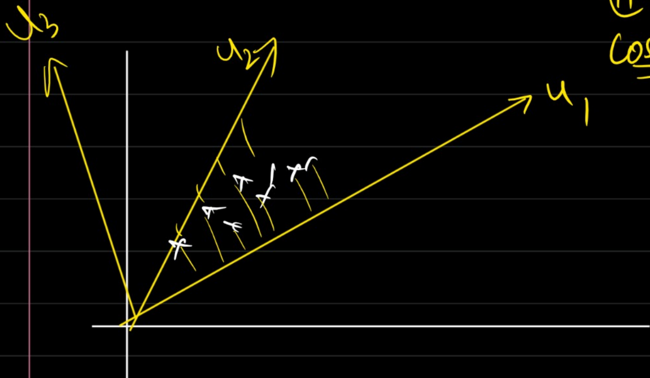
You want that unit vector where variance / spread is maximum.

$$\text{Var} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} \Rightarrow$$

Aim is to find that unit vector which captures the maximum variance after projection.

$\Downarrow$

Cost fn



spread $\uparrow$

$$\text{Var} = \frac{\sum (x_i - \bar{x})^2}{n}$$

$\Downarrow$

$$\text{Co-variance} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y - \bar{y})}{n}$$

* Covariance matrix.

$$\begin{array}{c|cc} & x_1 & x_2 \\ x_1 & cov(x_1,x_1) & cov(x_1,x_2) \\ x_2 & cov(x_2,x_1) & cov(x_2,x_2) \end{array}$$

$cov(x_1,x_1) = Var(x_1)$

$$\begin{bmatrix} var(x_1) & cov(x_1,x_2) \\ cov(x_2,x_1) & var(x_2) \end{bmatrix} \rightarrow \text{covariance matrix}$$

$$\begin{array}{c|ccc} & x_1 & x_2 & x_3 \\ \rightarrow x_1 & var(x_1) & cov(x_1,x_2) & cov(x_1,x_3) \\ x_2 & cov(x_1,x_2) & var(x_2) & cov(x_2,x_3) \\ x_3 & cov(x_3,x_1) & cov(x_3,x_2) & Var\ x_3 \end{array}$$

* if you decompose a Covariance matrix of features, then you will get eigen value and Eigen vector and Eigen vector with highest magnitude eigen value captures the maximum variance/spread.
(2nd highest magnitude – 2nd highest variance
3rd " " – 3rd " "
and so on)

Linear transformation of a matrix

$$A \cdot \vec{V} = \lambda \cdot \vec{V}$$

$\vec{V}$ – Eigenvector, λ – Eigen value

A – a matrix

{ when we use covariance matrix as A $\rightarrow$ Eigen decomposition of covariance matrix)
$\Downarrow$
Eigen vector, Eigen value.

Linear transformation

↳ A matrix transformation that brings changes in the coordinate system

$$\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \rightarrow \text{Matrix} \rightarrow \text{Linear transformation}$$

* → Many vectors were changing both magnitude and direction
* Few vectors changed only magnitude and not the direction.

⇓

→ these vectors are called Eigen vectors.

→ Eigen values are change in magnitude for Eigen vectors.

Eigen Vector

$(1, 0) \rightarrow (3, 0)$

3 time changes → Eigen value

* The vectors which only changes magnitude and not direction will be equals to dimension of matrix / no of features.

2x2 ⟶ 2 vectors

3 ⟶ 3 vector

No of features = No of Principal component

* Higest eigen Value → PC1

$$A \cdot \vec{V} = \lambda \cdot \vec{V}$$

(direction Same = V
Stretch / shrink = λ)

* Steps to Calculate Eigen Value & Eigen Vectors (to find PC's)

① Standardise the date (make the data mean centred.)

⇓

because it has been Observed PCA performs

better on mean - Centred data.

② Covariance matrix

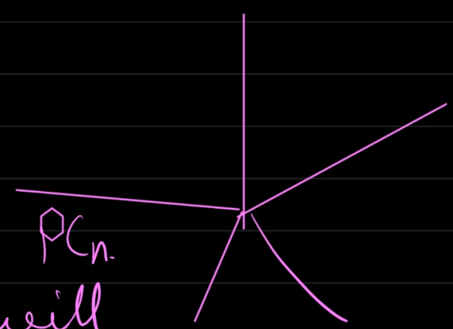
③ Eigen decomposition of Cov matrix.

$$Av = \hat{\lambda} v$$

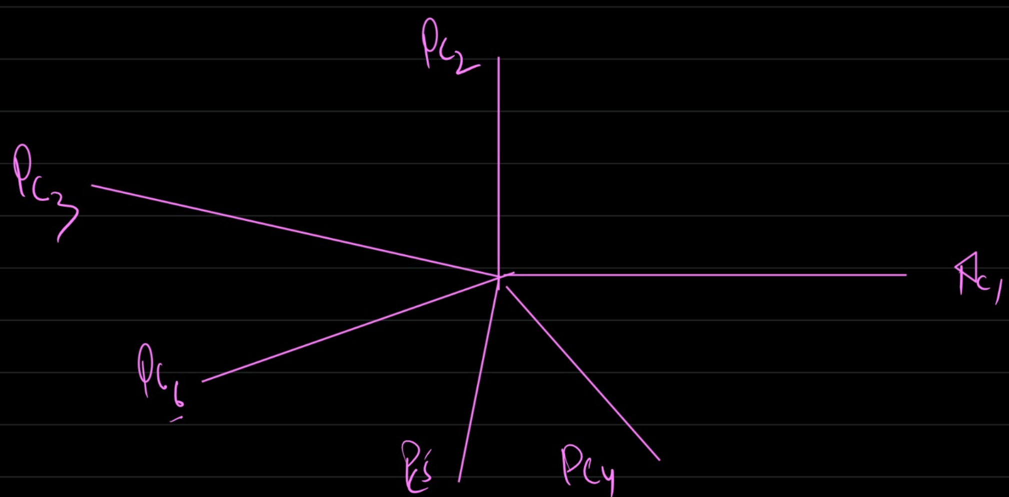
$\hat{\lambda} \rightarrow$ eigen value

$v \rightarrow$ Eigen Vector.

1000 f ⟶ 1000 PC's
↓

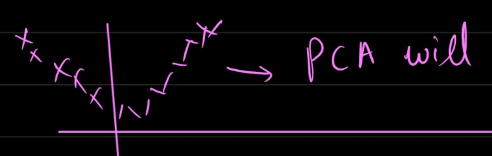


* $PC_1 > PC_2 > PC_3 \ldots PC_n$.
* Max of Variance/spread will be captured by first few Principal components
* PC's will be $\perp^r$ ~~to~~ each other.



* PCA fails?

⟶ Variance spread across all axis is same.



⟶ PCA will loose this pattern.