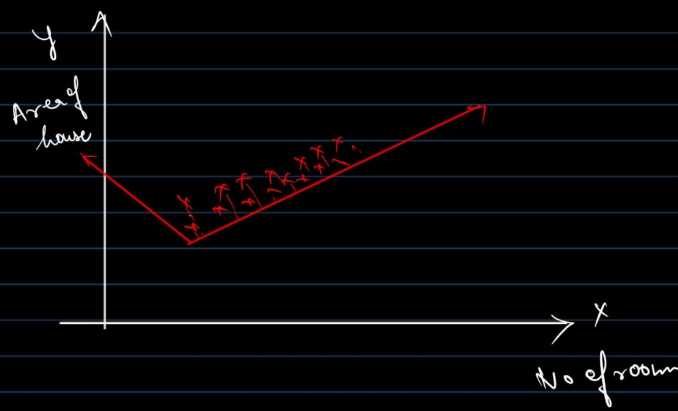
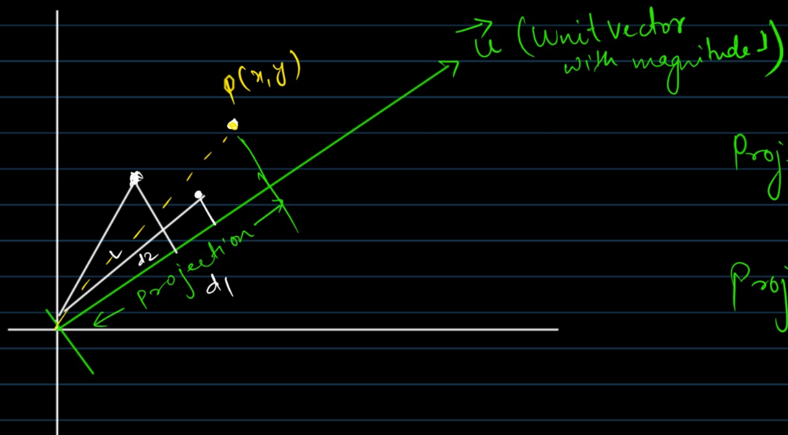
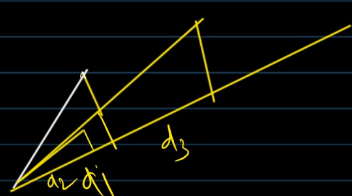
* Mathematical Explanation of PCA





Projection of P_1 on u = $\frac{P_1 \cdot u}{\|u\|}$

Projection of P_1 on u = $P_1 \cdot u$
⇓
Scaler value (only magnitude not direction)
⇓
Projection.

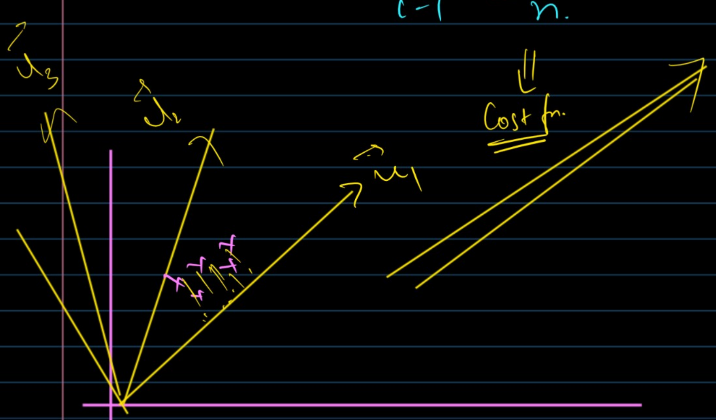


$\underline{d_1, d_2, d_3 \ldots d_n}$
⇓
you want the unit vector where the variance | spread is maximum.

→ $d_1, d_2 \ldots d_n$ are projections of different data points.

$$Var = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$$
⇓
Cost fn.

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



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

→ you will have multiple features.

$$\underline{\text{Covariance}} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

* Using Covariance ⇒ Covariance matrix will be created. It will help you to know the Unit vector which gives maximum spread. (Eigen value & Eigen vector)

Covariance matrix

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

$Cov(x_1,x_1) = Var(x_1)$

$$\begin{bmatrix} \underline{Var(x_1)} & Cov(x_1,x_2) \\ Cov(x_2,x_1) & \underline{Var(x_2)} \end{bmatrix}$$

Covariance matrix or Variance - covariance matrix.

$$\begin{array}{c|ccc} & x_1 & x_2 & x_3 \\ \hline x & Var(x_1) & Cov(x_1,x_2) & Cov(x_1,x_3) \\ x_2 & Cov(x_1,x_2) & Var(x_2) & Cov(x_2,x_3) \\ x_3 & Cov(x_3,x_1) & Cov(x_3\,x_2) & Var(x_3) \end{array}$$

* Theorem :- If you decompose a covariance matrix of features then you will get eigen value and Eigen vector.

→ Eigen vector with highest magnitude of eigen value captures the maximum variance / spread.

(2nd highest magnitude → 2nd highest variance
3rd " " → 3rd " "
and so on)

But what is decomposition ?

→ decomposition is some mathematical linear transformation

⇓

Linear transformation of a matrix

{ When we decompose Covariance matrix, we will get eigen value & Eigen vector }

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

where A is a matrix
$\vec{V}$ is a Eigen vector
λ is Eigen value (scalar)

* Linear transformation

→ It changes both magnitude & direction of vector.

→ A matrix transformation that brings changes in coordinate system.

$\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ → matrix → linear transformation

→ Many vectors were changing both magnitude & direction.

→ Few vectors changed only magnitude & Not direction.

⇓

these vectors are called as Eigen vector.

Eigen vector

$(1,0) \longrightarrow (3,3)$ → Eigen values are change in magnitude for eigen vectors

3 times changes ⇒ ∴ Eigen value is 3.

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

2×2 (2 features) ⟶ 2 vectors

3×3 (3 features) → 3 vectors.

* No of features = No of principal components

* Maximum var captured ⇒ highest Eigen value ⇒ PC_1

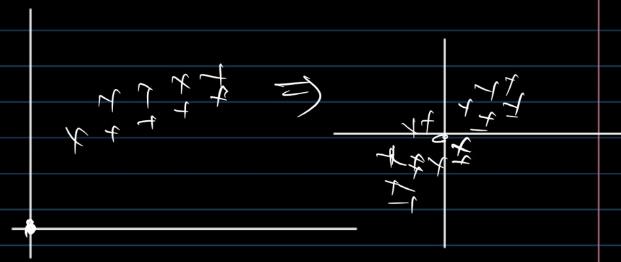
$$A \cdot \vec{V} = \lambda \cdot \vec{V}$$ (direction same = V, magnitude stretch/shrink = λ)

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

① Standardise the data (make the data mean centred)

⇓

It has been observed that PCA performs better on mean centred data.



② Calculate Covariance matrix

③ Eigen decomposition of Covariance matrix

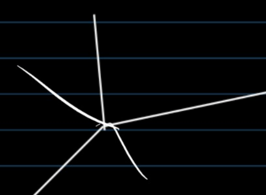
$$AV = \lambda V$$

$\lambda \rightarrow$ eigen value

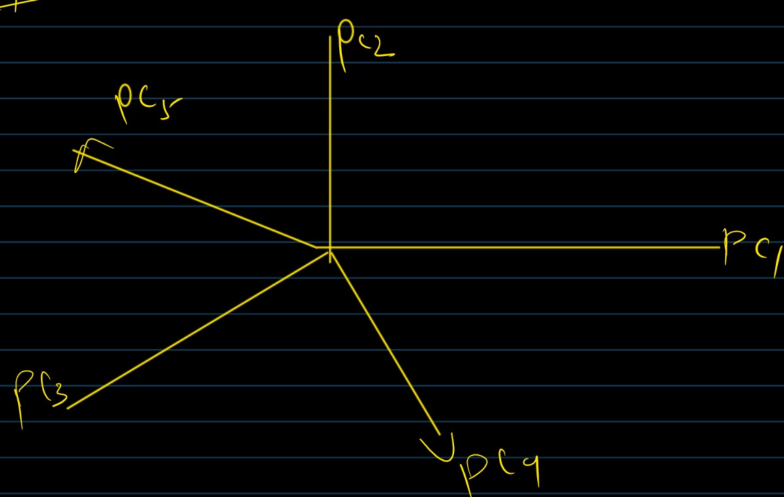
$V \rightarrow$ eigen vector

$$1000\ f \longrightarrow 1000\ PC's$$

* $PC_1 > PC_2 > PC_3 - - - PC_n$



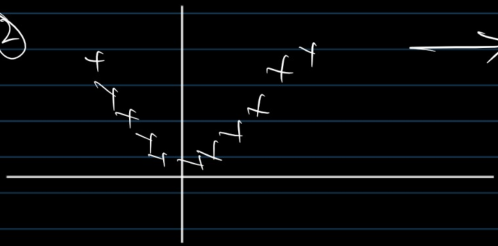
* Max of variance | spread will be captures by first few principal components.
* PCs will be $\perp^r$ to each other (as these are axis)
* PCA is an Unsupervised Algorithm.



* PCA fails.

①  $\longrightarrow$ Variance or Spread across all axis is same.

②



$\longrightarrow$ PCA will loose the pattern.