# PRINCIPAL COMPONENT ANALYSIS

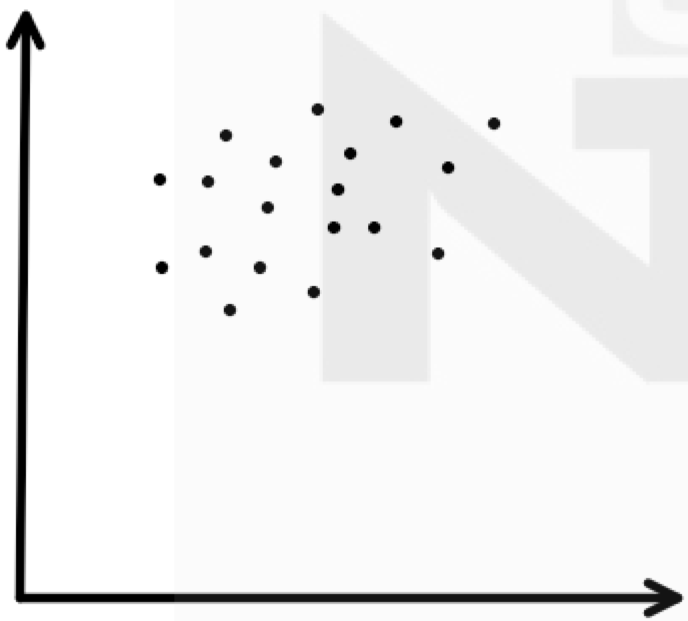
MUKESH KUMAR

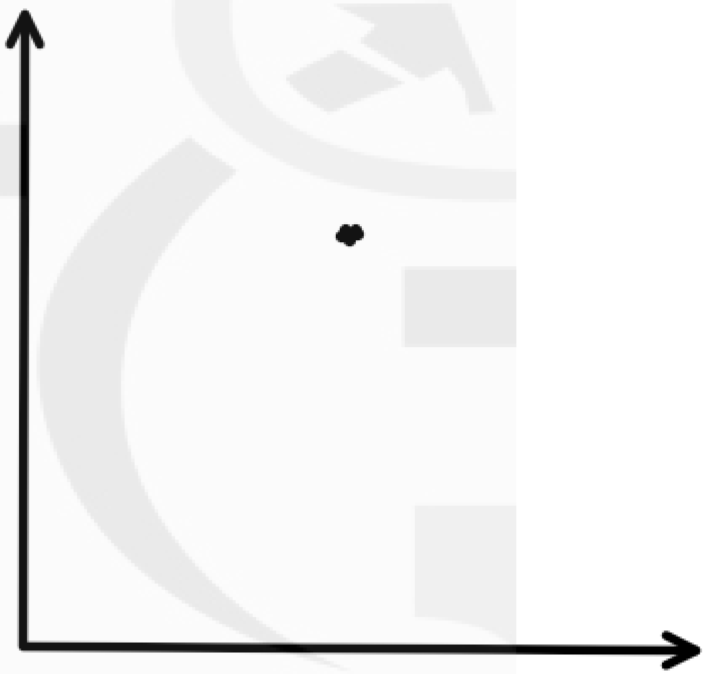
# Curse of Dimensionality

- In high dimensions, data becomes sparse.
- Distance-based models struggle, visualization becomes hard.
- We need **fewer, meaningful dimensions**.

# Variance = Information

# Variance



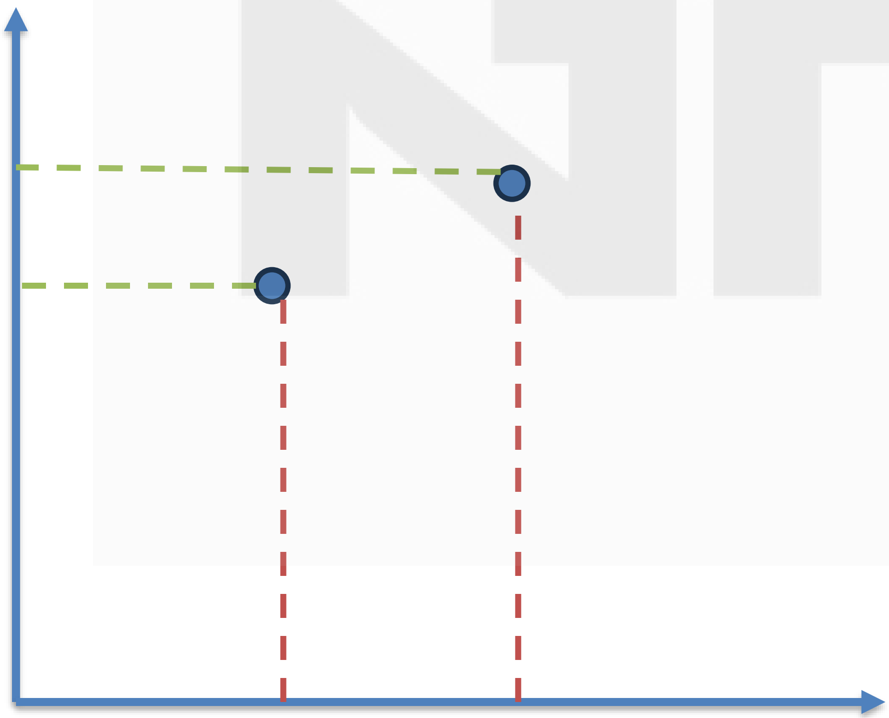


# Why Variance Matters

- Variance shows how much the data *spreads* or *changes*.
- Models **learn patterns from variance.**
- Retaining **maximum variance** means preserving the **most information.**

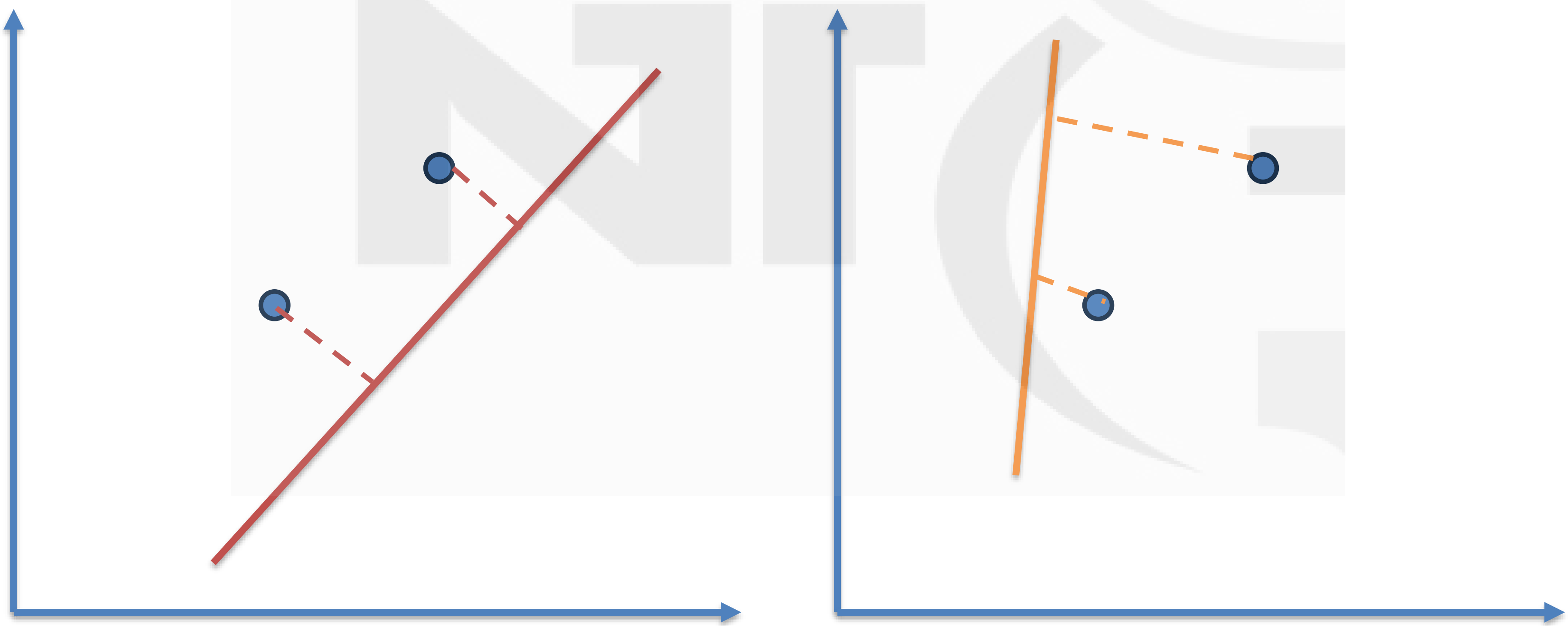
# Variance

Which direction/axis captures max variance



# Variance

When we project the points on lines, which line captures the max variance



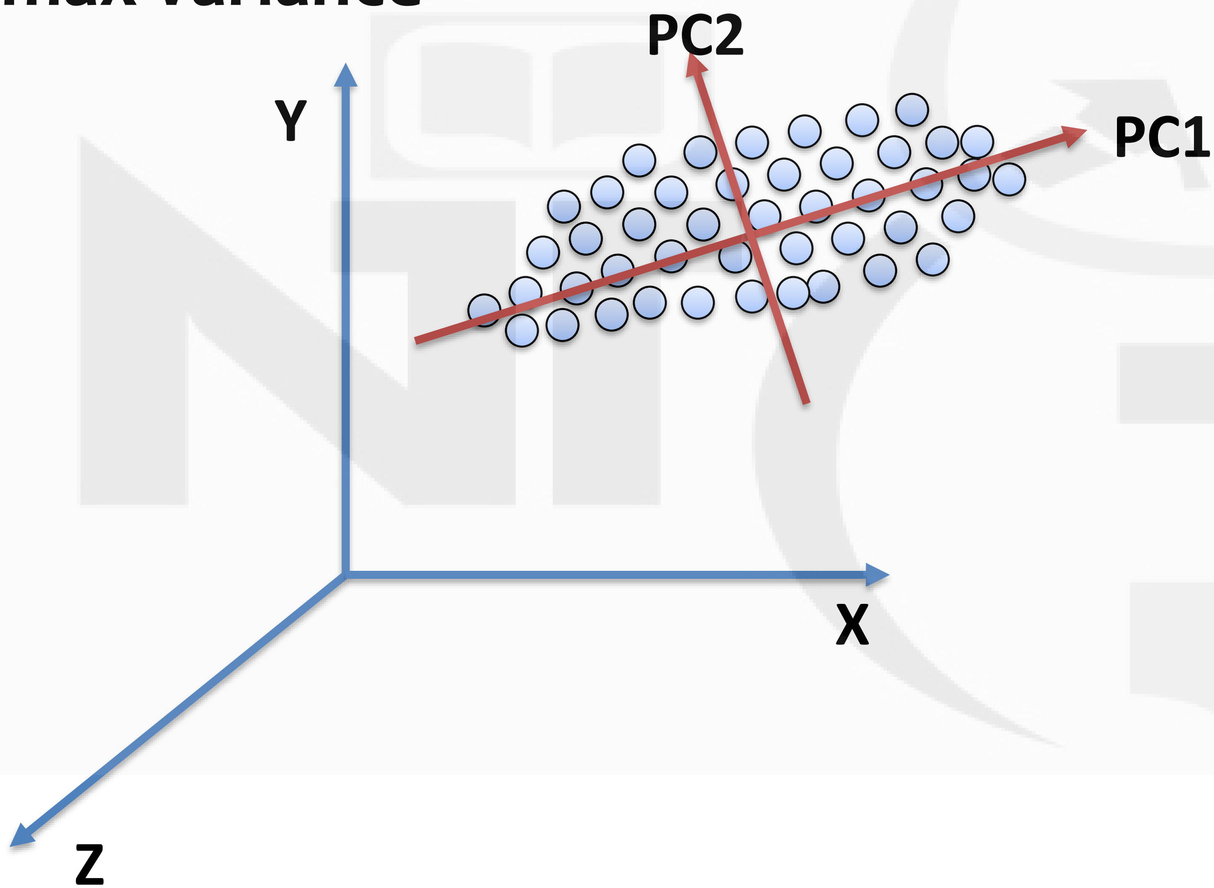
# What PCA Does

- PCA finds **new axes (Principal Components)**.
- These axes are linear combinations of original features.
- Each new axis captures **maximum remaining variance,** orthogonal to previous.
- Result: **Fewer dimensions**, but most of the **information retained**.

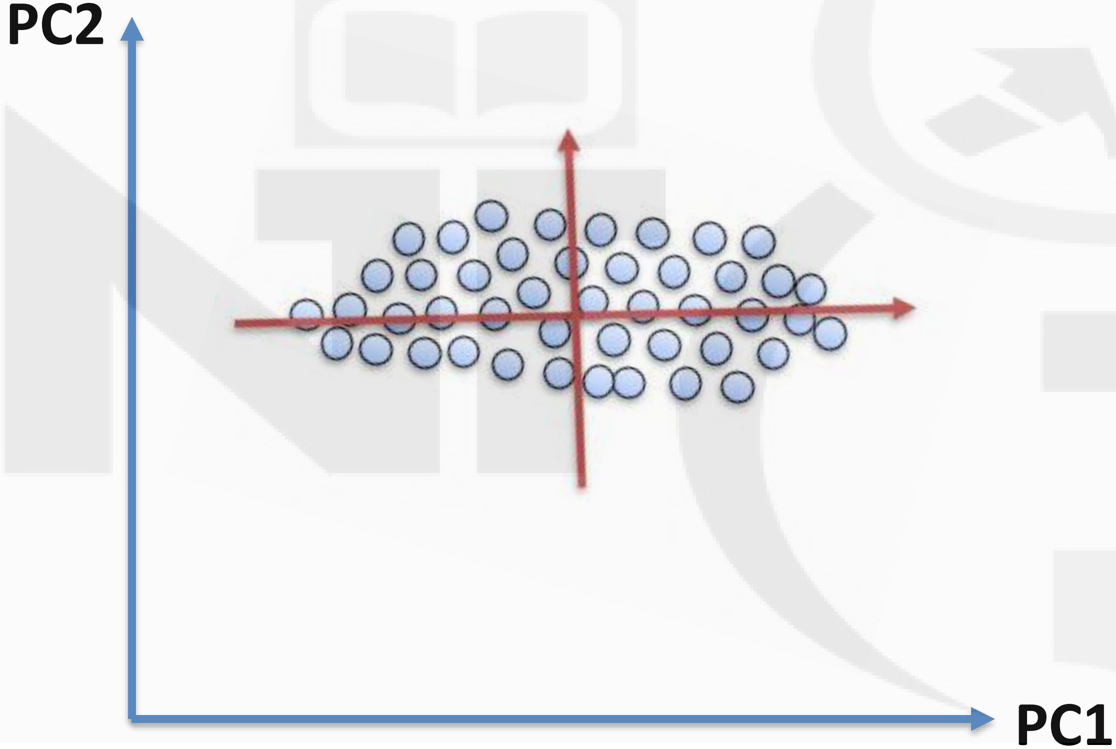
# How PCA Works

# Original 3D Axes

**Capture max variance**



# New Axes PC1 , PC2



- All PCs are orthogonal

# Principal components are linear combination of original features

- Features : $F1, F2, F3\ F4, F5, F6, F7, F8, F9\ldots Fn$
- Principal Components: (outcome of projection)
  – $PC1 = a1.F1 + a2.F2+a3.F3 \ldots.+a9.F9\ldots.+anFn$
  – $PC2 = b1.F1 + b2.F2+b3.F3 \ldots.+b9.F9 \ldots +bnFn$
  – $PC3 = c1.F1 + c2.F2+c3.F3 \ldots.+c9.F9+\ldots.+cnFn$

# Simple Definition

- PCA finds out most important features (Principal Components) which affect the target variable the most.

# What is Principal Component Analysis

- **Principal Component Analysis (PCA)** is a statistical technique used to transform a high-dimensional dataset into a lower-dimensional form while retaining as much of the original variability as possible.

- PCA achieves this by identifying the directions (principal components) along which the variance of the data is maximized.

- It simplifies the complexity in high-dimensional data while preserving trends and patterns.

# Principal Components

- Principal components are new axes formed by linear combinations of the original variables.

- The first principal component captures the maximum variance in the data.

- Subsequent principal components are orthogonal (uncorrelated) to the previous ones and capture the remaining variance.

# Dimensionality Reduction

- Reduces the number of variables under consideration by creating new uncorrelated variables (principal components).
- Helps in simplifying models and reducing computational costs.
- Makes it easier to visualize data in 2D or 3D space.

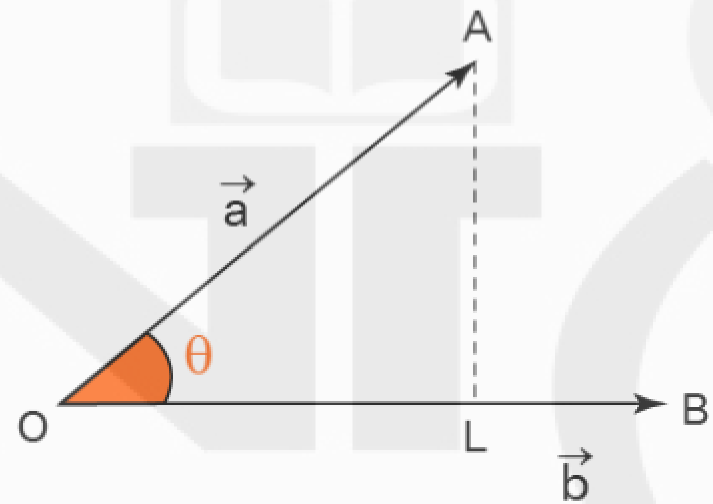
# Practical Use of PCA

- **Data Preprocessing:** PCA is often used before feeding the data into machine learning models, especially when dealing with high-dimensional data.
- **Visualization:** Reducing data to two or three dimensions can make it easier to visualize clusters or patterns in the data.
- **Noise Reduction:** By retaining only the principal components with the highest variance, PCA can help remove noise (less important variance).

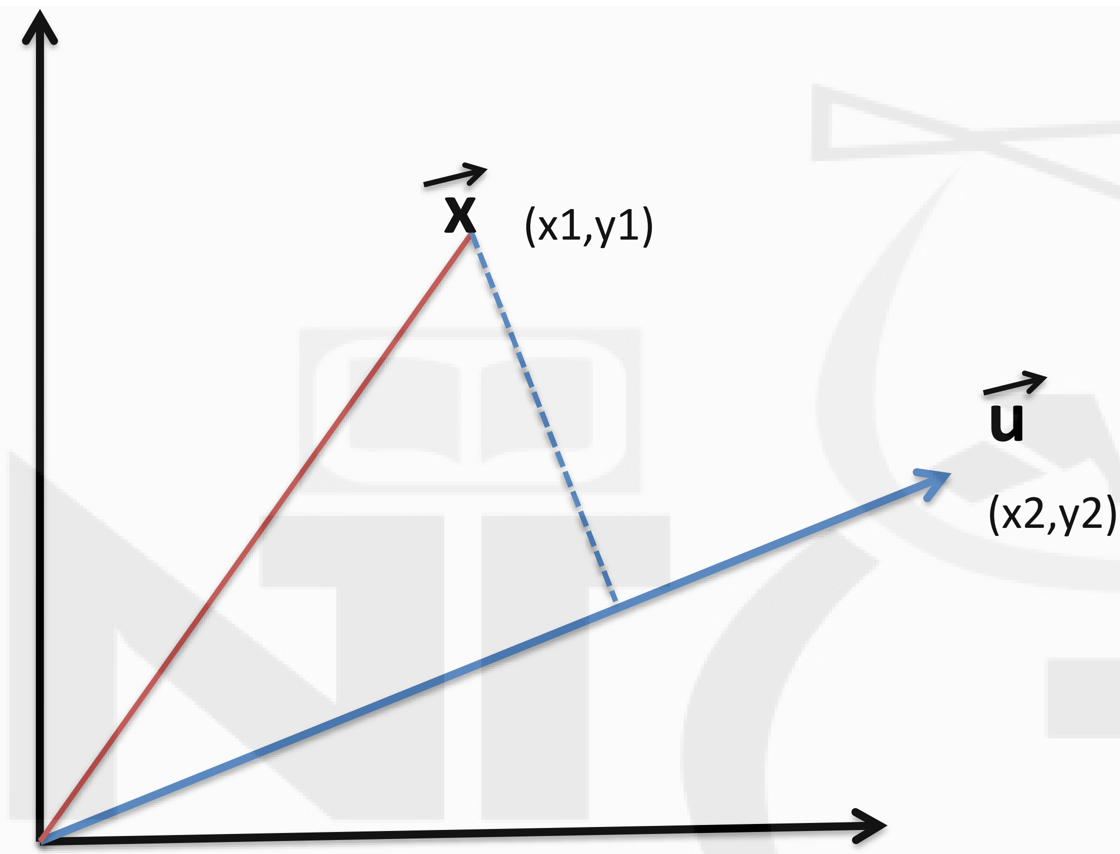
# PCA Demo

# PCA LOSS FUNCTION

# Projecting a Vector




$$OL = \frac{\vec{a}\,.\,\vec{b}}{|\vec{b}|}$$




$$x.u \,/\, |u| \;=\; \vec{x}.\vec{u} \;=\; u^T x$$

# Dot Product

The dot product of two vectors is a fundamental operation in linear algebra. To clarify the dot product between two vectors $\mathbf{a} = [x_1, y_1]$ and $\mathbf{b} = [x_2, y_2]$, we can break it down as follows:

## Definition

The dot product of two vectors $\mathbf{a}$ and $\mathbf{b}$ is a scalar quantity and is calculated as:

$$\mathbf{a} \cdot \mathbf{b} = x_1 x_2 + y_1 y_2$$

## Steps to Calculate the Dot Product

1. **Multiply corresponding components**: Multiply the first component of $\mathbf{a}$ by the first component of $\mathbf{b}$, and the second component of $\mathbf{a}$ by the second component of $\mathbf{b}$.
2. **Sum the products**: Add the products obtained in the first step.

## Example

Consider the vectors:

$$\mathbf{a} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

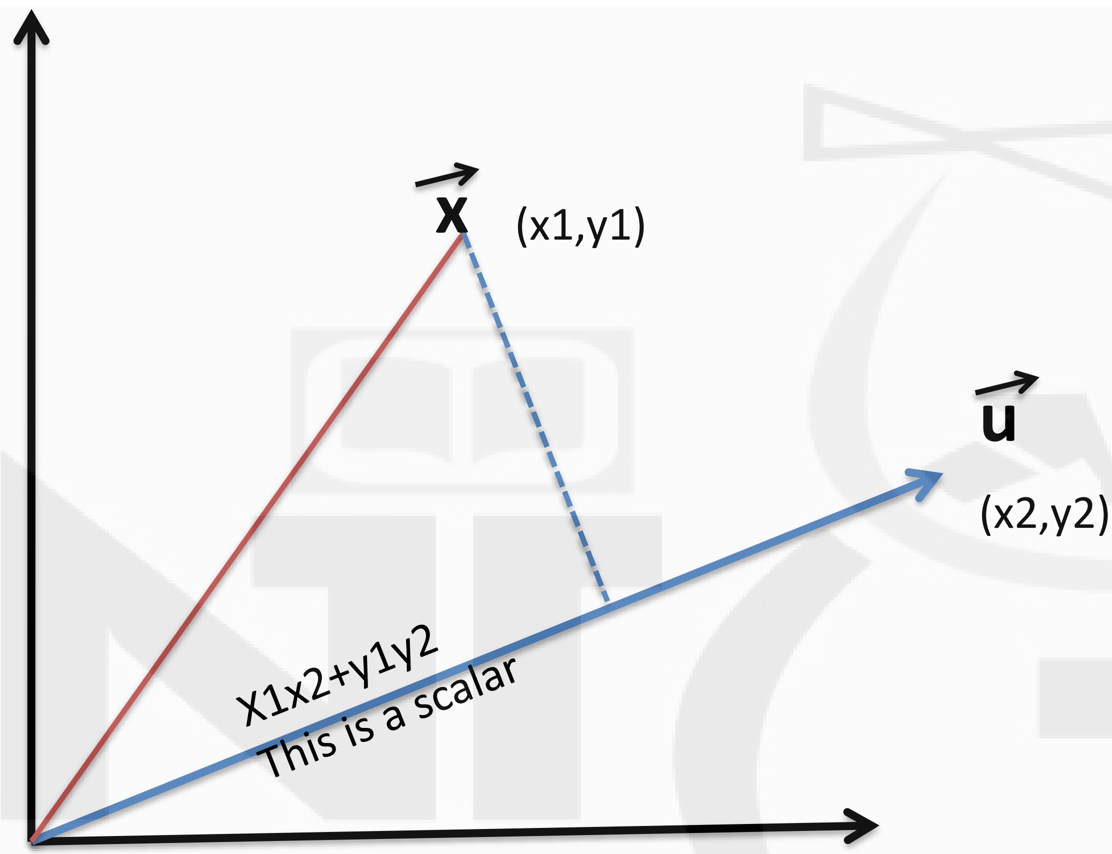
$$\mathbf{b} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

To compute the dot product using the transpose:

$$\mathbf{a}^T = \begin{bmatrix} 3 & 4 \end{bmatrix}$$

Then:

$$\mathbf{a}^T\mathbf{b} = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 3 \times 2 + 4 \times (-1) = 6 - 4 = 2$$




$$x.u \,/\, |u| \;=\; \vec{x}.\vec{u} \;=\; u^T x$$

# Projections for all the points

$u^T x_1, u^T x_2, u^T x_3, u^T x_4, u^T x_5, \ldots\ldots\ u^T x_n$

Variance will be $\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$$\mathbf{1/n \sum (u^T x_1 - u^T \bar{x})^2}$$

and the variance of the projected data is given by

$$\frac{1}{N}\sum_{n=1}^{N}\left\{\mathbf{u}_1^{\mathrm{T}}\mathbf{x}_n - \mathbf{u}_1^{\mathrm{T}}\overline{\mathbf{x}}\right\}^2 = \mathbf{u}_1^{\mathrm{T}}\mathbf{S}\mathbf{u}_1$$

where $\mathbf{S}$ is the data covariance matrix defined by

$$\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^{\mathrm{T}}.$$

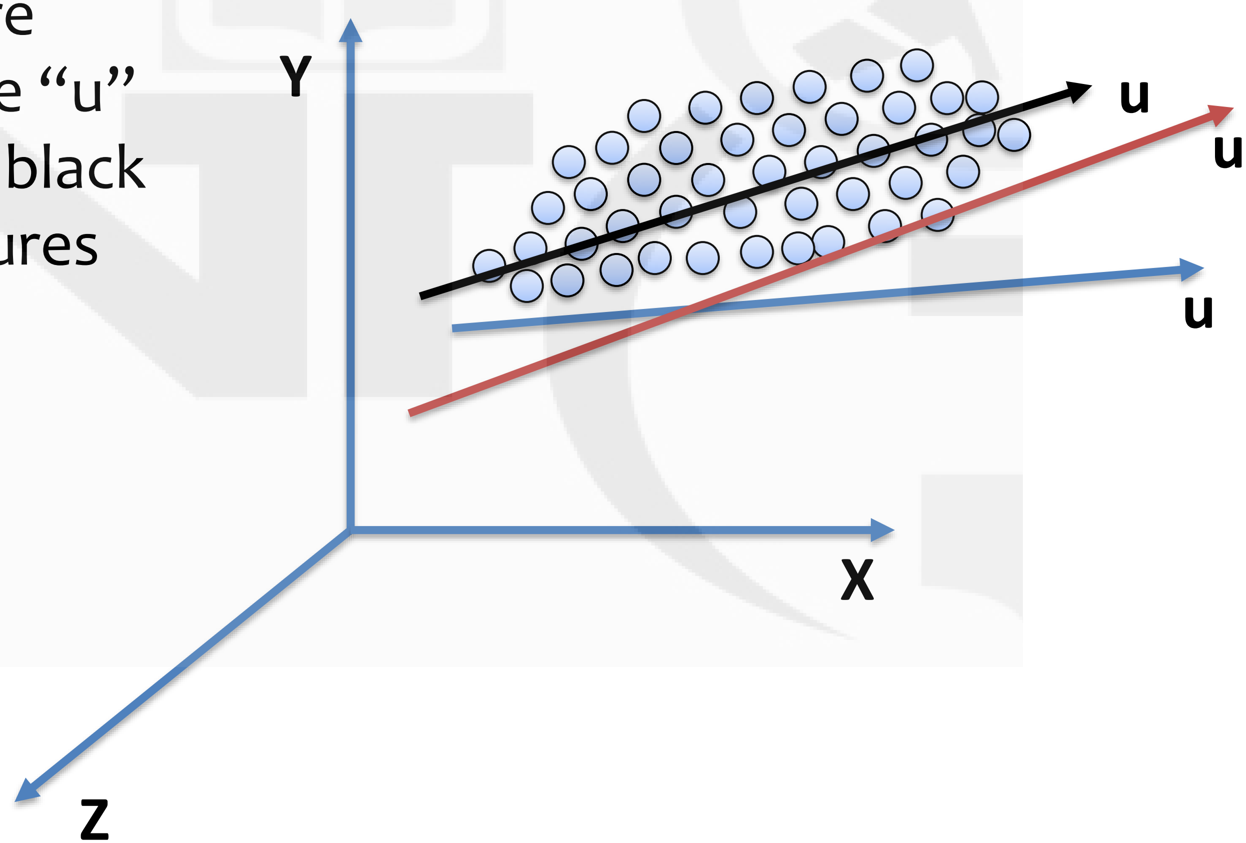
# Our Final Objective Function is

- This equation represents all the original points project on the new vector

- In the equation “u” is the new vector we are looking for such that it captures the max variance

$$\mathbf{u}^\top S \mathbf{u}$$

- It could be one of the few vectors shown here , but we are interested in the “u” represented by black line which captures max variance



# Final Objective Fun

- We need to find the "u" vector which is indeed a unit vector i.e. dot of transpose with itself is 1.

$$\max_{\mathbf{u}} \quad \mathbf{u}^\top S\mathbf{u} \quad \text{subject to} \quad \mathbf{u}^\top\mathbf{u} = 1$$

- It needs to be a unit vector otherwise the projected values may get scaled changing the true variance of data

# Apply Langrange Multiplier

$$\mathcal{L}(\mathbf{u}, \lambda) = \mathbf{u}^\top S\mathbf{u} - \lambda(\mathbf{u}^\top \mathbf{u} - 1)$$

- Equate the derivative of function wrt u to 0 to find min or max of the function

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}} = 2S\mathbf{u} - 2\lambda\mathbf{u} = 0 \quad \Rightarrow \quad S\mathbf{u} = \lambda\mathbf{u}$$

- https://en.wikipedia.org/wiki/Lagrange_multiplier

# Eigen Equation

- When we solved the objective function, we got an equation which is a famous eigen equation

- Which means Principal components are indeed the eigen vectors.

$$S\mathbf{u} = \lambda \mathbf{u}$$

# Connecting to Eigenvectors and Eigenvalues

From the eigenvalue equation ,we can immediately recognize the following:

- The vector **u** is an **eigenvector** of the covariance matrix **S** , because it satisfies the equation

- The scalar ( lambda ) is the corresponding **eigenvalue,** which in the context of PCA represents the variance captured by that eigenvector (or principal component).

# Eigen Definition

- **Eigenvector**: A vector that doesn't change its direction under a given transformation, only its scale changes.

- **Eigenvalue**: The factor by which the eigenvector is stretched or compressed during the transformation.

# Standard Eigen Equation

$$A\mathbf{v} = \lambda\mathbf{v}$$

- $\lambda$ tells you how much the eigenvector $\mathbf{v}$ is stretched or compressed.
  - If $\lambda > 1$, the eigenvector gets **stretched**.
  - If $0 < \lambda < 1$, the eigenvector gets **compressed**.
  - If $\lambda = 0$, the eigenvector is **collapsed** to zero.
  - If $\lambda < 0$, the eigenvector is flipped and stretched or compressed.

# Conclusion

Thus, we have derived that:

- The vector (u) that maximizes the variance in the data (the principal component direction) is an **eigenvector** of the covariance matrix ( S ).

- The associated eigenvalue ( lambda ) tells us the **variance** (or amount of information) captured along that direction.

# PCA in a nutshell

- PCA works by finding the directions of maximum variance in the data and projecting the data onto these directions.

- These directions are the eigenvectors of the covariance matrix of the data, and the amount of variance explained by each direction is the eigenvalue.

- To perform PCA, you must first standardize the data to have zero mean and unit variance

# PCA Performance issues

1. PCA effectiveness depends upon the scales of the attributes. If attributes have different scales, PCA will pick variable with highest variance rather than picking up attributes based on correlation

2. Changing scales of the variables can change the PCA

3. Interpreting PCA can become challenging due to presence of discrete data

4. Presence of skew in data with long thick tail can impact the effectiveness of the PCA (related to point 1)

5. PCA assumes linear relationship between attributes. It is ineffective when relationships are non linear

# Additional Info

# Covariance

- **Covariance** is a statistical measure that indicates the extent to which two variables change together.

- If the variables tend to show similar behavior (i.e., if one increases, the other tends to increase, and if one decreases, the other tends to decrease), the covariance will be positive.

- If one variable tends to increase while the other decreases, the covariance will be negative.

- Covariance is a key concept in statistics, particularly in fields like finance, where it is used to assess the relationship between the returns on different assets.

For a sample of $n$ observations, the sample covariance can be calculated as:

$$\mathrm{Cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

Where:

- $X_i$ and $Y_i$ are the $i$-th observations of variables $X$ and $Y$, respectively.
- $\bar{X}$ is the sample mean of $X$.
- $\bar{Y}$ is the sample mean of $Y$.

## Interpretation

- **Positive Covariance**: Indicates that the variables tend to move in the same direction. For example, if $\mathrm{Cov}(X, Y) > 0$, an increase in $X$ generally corresponds to an increase in $Y$.
- **Negative Covariance**: Indicates that the variables tend to move in opposite directions. For example, if $\mathrm{Cov}(X, Y) < 0$, an increase in $X$ generally corresponds to a decrease in $Y$.
- **Zero Covariance**: Indicates that there is no linear relationship between the variables.

## Example Calculation

Suppose we have two variables $X$ and $Y$ with the following values:

$$X = [2, 4, 6, 8]$$
$$Y = [1, 3, 2, 5]$$

1. Calculate the means:
   $\bar{X} = \frac{2+4+6+8}{4} = 5$
   $\bar{Y} = \frac{1+3+2+5}{4} = 2.75$

2. Calculate the deviations from the mean:
   $(X_i - \bar{X}) = [-3, -1, 1, 3]$
   $(Y_i - \bar{Y}) = [-1.75, 0.25, -0.75, 2.25]$

3. Calculate the product of the deviations and sum them:
   $\sum(X_i - \bar{X})(Y_i - \bar{Y}) = (-3)(-1.75) + (-1)(0.25) + (1)(-0.75) + (3)(2.25) = 5.25 - 0.25 - 0.75 + 6.75 = 11$

4. Divide by $n - 1$:
   $\text{Cov}(X, Y) = \frac{11}{3} \approx 3.67$

So, the covariance between $X$ and $Y$ is approximately 3.67, indicating a positive relationship.