Principal Component Analysis (PCA) –Geometric Intuition

➢ What is PCA?

- **Principal Component Analysis (PCA)** is a **dimensionality reduction** technique.
- It transforms the data to a **new coordinate system** where:
  - The **first axis (PC1)** captures the **maximum variance**.
  - The **second axis (PC2)** captures the **next highest variance**, and so on.
- These new axes are called **principal components**, and they are **orthogonal (perpendicular)** to each other.

➢ Benefits of Using PCA

- **Reduces dimensionality** while retaining most of the **important information**.
- Improves **visualization** of high-dimensional data in 2D/3D.
- Speeds up **training time** and reduces **overfitting** in ML models.
- Removes **redundancy** by combining correlated features.

➢ Geometric Intuition

- PCA finds the **direction (axis)** in feature space where the **data varies the most**.
- It **projects** the data onto this direction, thereby reducing dimensions while preserving variance.
- **Example:**
  - ❖ Imagine data scattered in a cloud—PCA finds the **longest stretch (maximum variance)** and aligns the first principal component along that stretch.
- Think of PCA as **rotating the coordinate axes** to better align with the **spread** of the data.

➢ What is Variance and Why Is It Important?

- **Variance** measures how much the data is **spread out**.
- PCA **prioritizes directions with high variance**, assuming that high variance = more **informative** data.
- Low variance directions are considered **less important** (e.g., noise), and can often be dropped without major information loss.

➢ Key Takeaway:

PCA uses a **geometric transformation** to rotate and project your data onto fewer, more **informative dimensions**, making it easier to process and visualize — without significantly sacrificing important structure or relationships in the data.
