**How do we perform reducing the dimensionality of data using Principal Component Analysis (PCA)?**

Reducing the dimensionality of data using **Principal Component Analysis (PCA)** involves transforming a dataset with many features (variables) into a smaller set of new features called **principal components**. These components capture the most important patterns in the data while minimizing information loss.

Here’s a step-by-step explanation of how PCA works:

**🔹 1. Standardize the Data**

PCA is affected by the scale of the data, so the first step is to **normalize** each feature (e.g., mean = 0, standard deviation = 1).

Example:  
If one feature is income (in thousands) and another is age (in years), standardization ensures both are on the same scale.

**🔹 2. Compute the Covariance Matrix**

The covariance matrix measures how variables in the data relate to one another—whether they increase or decrease together.

* For 2 features, the covariance matrix is a 2×2 matrix.
* It tells us about the direction and strength of relationships between features.

**🔹 3. Calculate the Eigenvectors and Eigenvalues**

* **Eigenvectors** indicate the **directions** of the new feature space (principal components).
* **Eigenvalues** represent the **magnitude** or importance of each eigenvector (how much variance it explains).

The first principal component has the highest variance, the second has the next highest, and so on.

**🔹 4. Select the Top Principal Components**

Choose the top *k* principal components that together explain a significant portion of the total variance (e.g., 95%).

This step reduces the dimensionality while preserving the most important information.

**🔹 5. Transform the Original Data**

Project the original dataset onto the selected principal components to get a new dataset with reduced dimensions.

This transformed data now has fewer features but still retains most of the structure of the original data.

**✅ Example Use Case:**

Imagine you have a dataset with 50 features. After applying PCA, you find that just 10 principal components explain 95% of the variance. You can now work with these 10 features instead of the original 50, making the data easier to visualize and process.