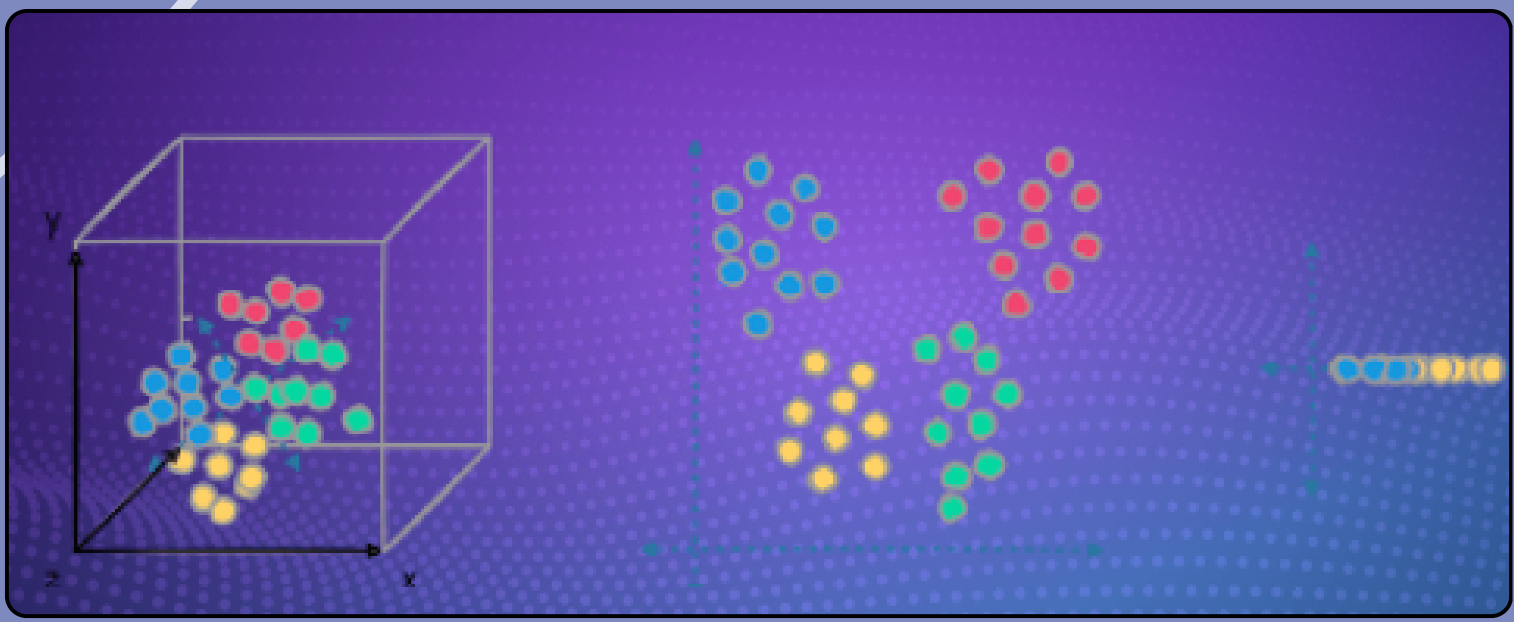
Dimensionality Reduction

What is it?
Dimensionality reduction is a technique that simplifies a dataset by reducing the number of features while preserving its essential characteristics. This is achieved through methods like:



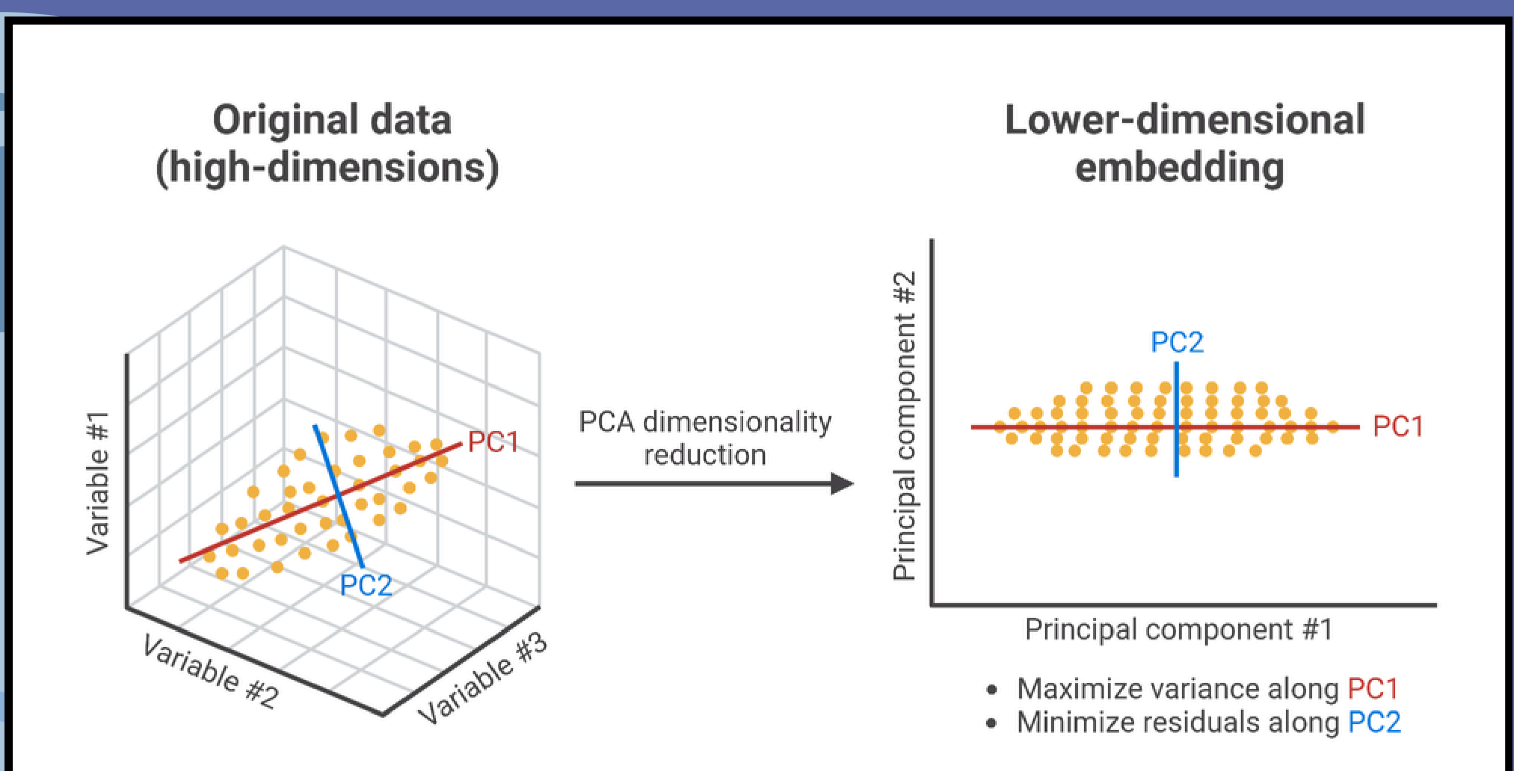
why is needed?

1. **Improved Model Performance**: Reduces the risk of overfitting, enhancing model generalization.
2. **Faster Computation**: Decreases processing time and resource requirements.
3. **Effective Visualization:** Makes it easier to explore and interpret complex datasets.
4. **Noise Reduction:** Eliminates irrelevant features, improving analysis quality.
5. **Feature Extraction:** Retains only the most informative features, aiding in insights.
6. **Data Compression**: Reduces data size for efficient storage and transmission.

Following are some of the most commanly used Dimensionality Reduction techniques

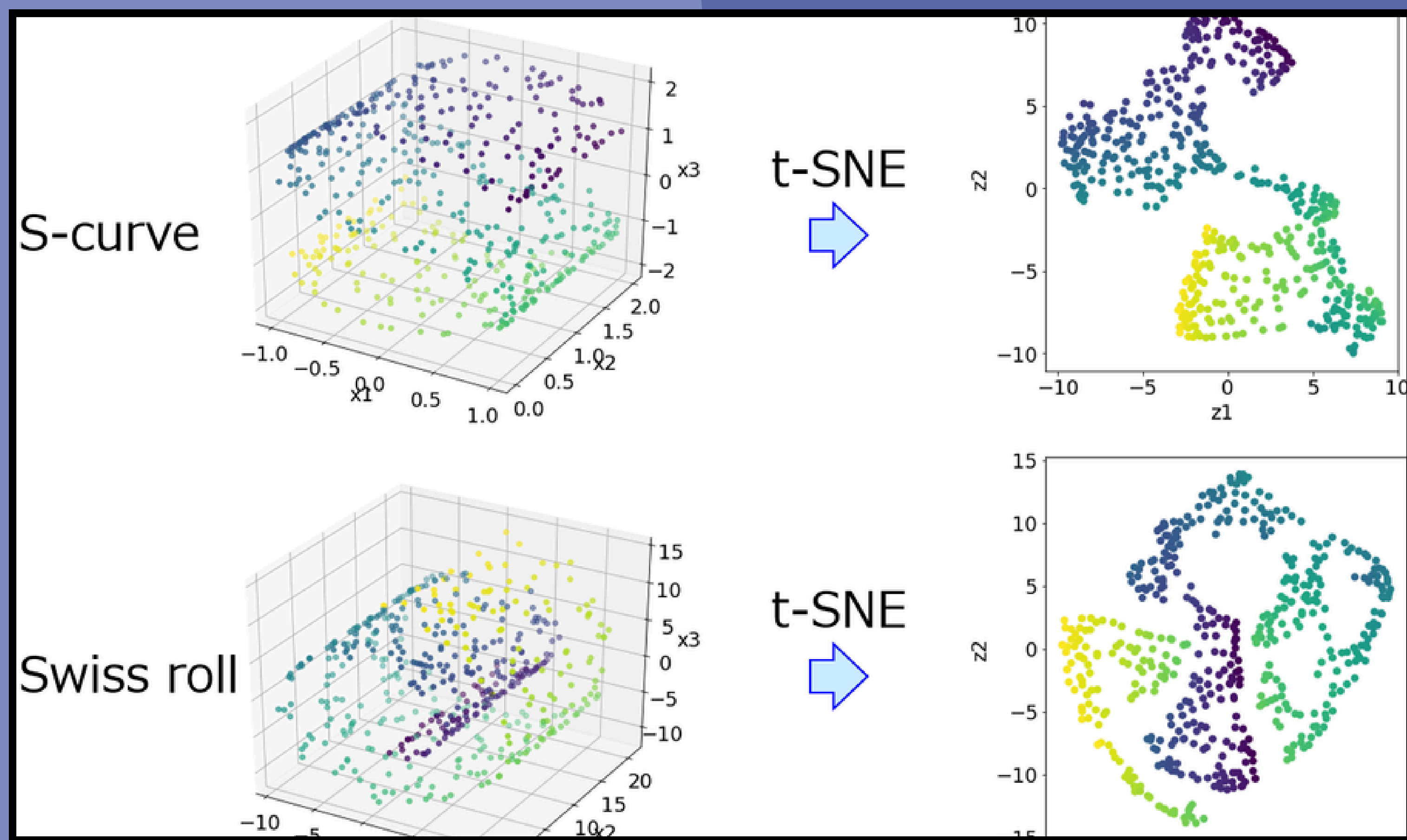
1. Principal Component Analysis (PCA)

- **Description**: Transforms data into a lower-dimensional space by finding the directions (principal components) that maximize variance.
- **Key Benefit:** Reduces data complexity while retaining most of the original variability.
- **Use Cases:** Data compression, noise reduction, and exploratory data analysis.



2. t-Distributed Stochastic Neighbor Embedding (t-SNE)

- **Description: Maps** high-dimensional data to a 2D or 3D space, preserving local structure and making clusters more distinct.
- **Key Benefit:** Excellent for visualizing complex data and identifying clusters.
- **Use Cases:** Visual exploration of data, pattern recognition in images or text data.



3. Linear Discriminant Analysis (LDA)

- **Description**: Projects data onto a lower-dimensional space to maximize class separability. Requires labeled data for training.
- **Key Benefit**: Enhances the separation between different classes, aiding classification tasks.
- **Use Cases**: Preprocessing step in supervised learning, face recognition, and text classification.

4. Autoencoders (Neural Networks)

- **Description:** Uses neural networks to learn an efficient, lower-dimensional representation of data through encoding and decoding processes.
- **Key Benefit:** Handles non-linear relationships and complex patterns in data.
- **Use Cases:** Image compression, anomaly detection, and feature extraction.

5. Singular Value Decomposition (SVD)

- **Description:** Decomposes a matrix into three simpler matrices, reducing dimensionality while preserving important features.
- **Key Benefit:** Effective in reducing the number of features in high-dimensional data.
- **Use Cases:** Natural language processing (NLP), recommender systems, and latent semantic analysis.

Deep Salunkhe 21102A0014 BE CMPN A