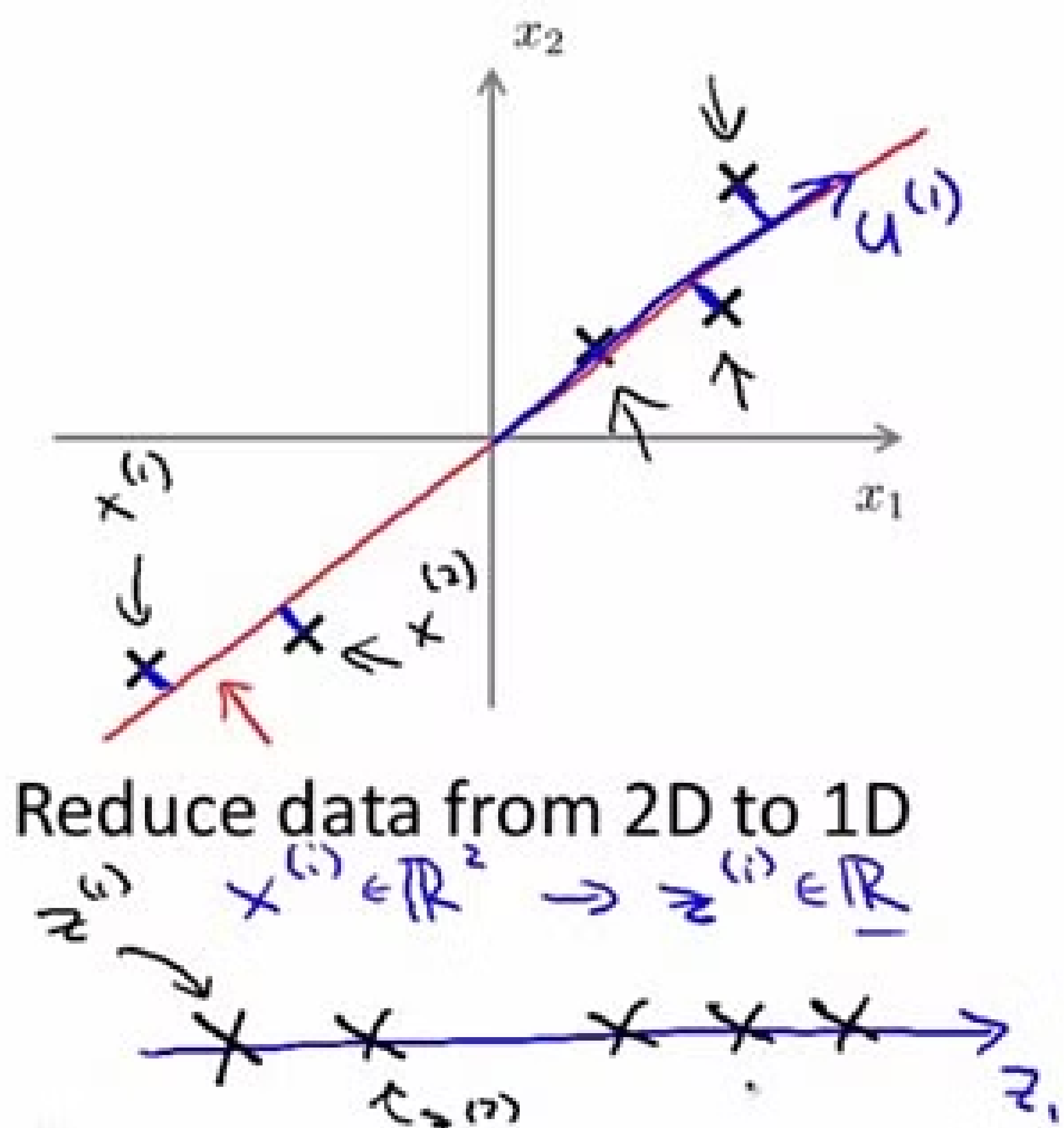
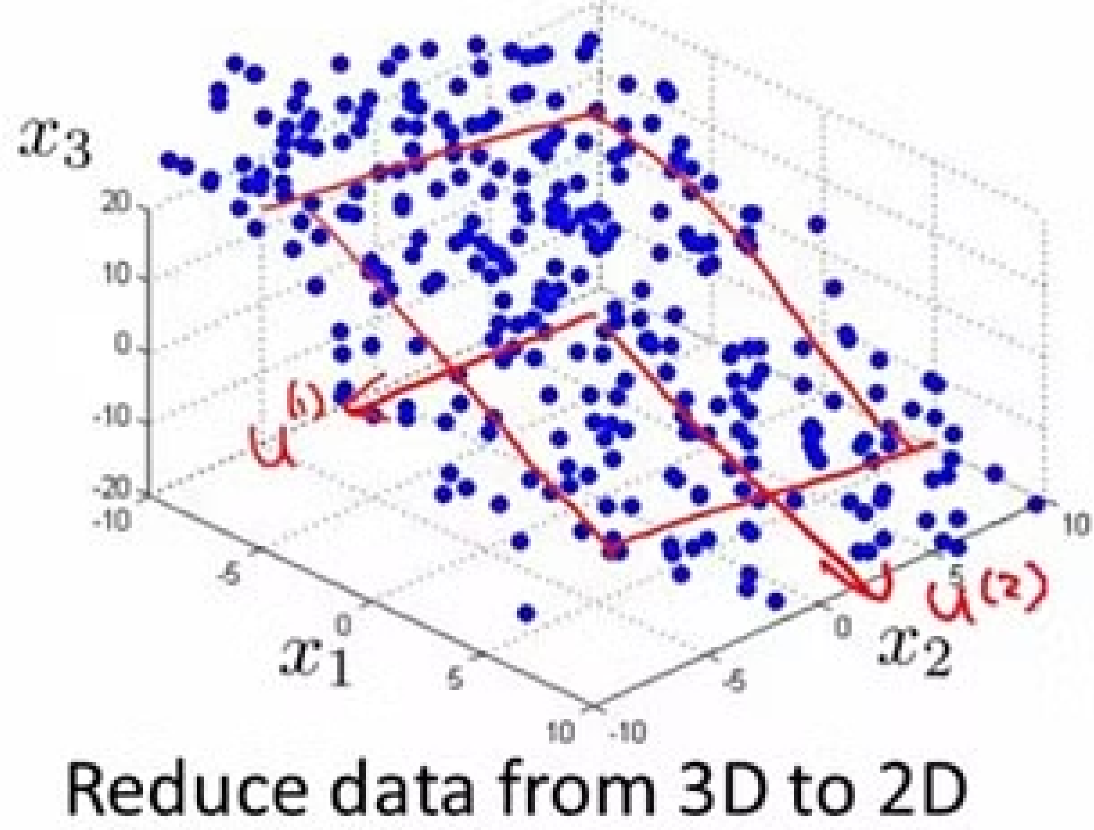
# Principal Component Analysis (PCA)





- PCA is a dimensionality reduction technique used in data analysis and machine learning.

- It helps uncover patterns and relationships in high-dimensional data.

- PCA transforms the data into a new set of variables called principal components.

# Step 1

## Standardize the data:

- If features are measured on different scales, standardize them to have zero mean and unit variance.
- Ensures each feature contributes equally to PCA.

# Step 2

## Compute the covariance matrix:

- **Calculate the covariance matrix of the standardized data.**
- **Represents the relationships between different features.**

# Step 3

**Compute the eigenvectors and eigenvalues:**

- Eigenvectors represent directions of maximum variance.
- Eigenvalues indicate the amount of variance captured by each eigenvector.

# Step 4

## Select the principal components:

- Sort eigenvalues in descending order.
- Choose the top-k eigenvectors (principal components) based on explained variance.

# Step 5

**Project the data onto the new feature space:**

- **Multiply the standardized data by the selected eigenvectors.**
- **Obtain a new lower-dimensional representation of the data.**

# Benefits of PCA:

- **Dimensionality reduction: PCA reduces the number of features while preserving most of the information.**
- **Data visualization: PCA can visualize high-dimensional data in a lower-dimensional space.**

# Benefits of PCA:

- **Feature extraction: PCA identifies the most important features that contribute to the variance in the data.**
- **Noise reduction: PCA can eliminate noise by removing low-variance components.**

# Applications of PCA:

- Exploratory data analysis
- Pattern recognition
- Image and signal processing
- Machine learning feature selection

# Thank You