# 14 PCA

## Principal Component Analysis (PCA)

- It is a Feature Extraction algorithm, not a feature selection, convert n features to k, where k < n
- **Principal Component Analysis (PCA)** is one of the most popular techniques for dimensionality reduction
- PCA transforms the original features into a new set of uncorrelated features called **principal components**
- These components are ordered by the amount of variance they capture from the data, with the first few components capturing most of the variability

### Working of PCA

1. **Standardize the Data**
    - Ensure that the data is centered around the origin (mean = 0) and has unit variance

    $$X_{\text{std}} = \frac{X-\mu}{\sigma}$$

    - where $u$ is the mean of the data, and $\sigma$) is the standard deviation
2. **Compute the Covariance Matrix**
    - The covariance matrix is a square matrix that shows the covariance (a measure of how much two variables change together) between pairs of features in the data

    $$\mathbf{C} = \frac{1}{n-1}\mathbf{X}_{\text{std}}^T\mathbf{X}_{\text{std}}$$

    $$\mathbf{C} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T$$
3. **Compute the Eigenvalues and Eigenvectors**
    - **Eigenvalues** : Indicate the amount of variance captured by each eigenvector (principal component)
    - **Eigenvectors** : Represent the directions of the principal components in the feature space

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

    - where ( \lambda ) is the eigenvalue and ( \mathbf{v} ) is the corresponding eigenvector
4. **Sort Eigenvectors by Eigenvalues**
    - The eigenvectors are sorted in decreasing order of their corresponding eigenvalues

- The top eigenvectors (with the largest eigenvalues) are chosen to form the new feature space

5. **Transform the Data**
   - The original data is projected onto the selected eigenvectors (principal components), creating a new dataset with reduced dimensions

   $$\mathbf{X}_{\text{pca}} = \mathbf{X}_{\text{std}}\mathbf{V}$$

   - where ( \mathbf{V} ) is the matrix of the top ( k ) eigenvectors

6. **Variance Explained by Each Principal Component**

$$\text{Variance(Component } i) = \frac{\lambda_i}{\sum_{j=1}^{k} \lambda_j}$$

## Objective of PCA

- **Maximizing Variance**
- **Minimizing Distances**

## Intuition Behind PCA

- PCA seeks to find new axes (principal components) that maximize the variance in the data
- The first principal component captures the most variance, and each subsequent component captures the maximum remaining variance while being orthogonal to the previous components

## Advantages of PCA

- **Reduces Complexity** : By focusing on the most significant components, PCA simplifies the dataset
- **De-correlates Features** : The resulting principal components are uncorrelated, which can improve the performance of machine learning algorithms
- **Facilitates Visualization** : PCA can reduce data to 2D or 3D for easy visualization

## Limitations of PCA

- **Loss of Information** : Some data variance is inevitably lost when reducing dimensions
- **Assumption of Linearity** : PCA assumes that the principal components are linear combinations of the original features, which may not capture complex, nonlinear relationships
- **Interpretability** : The new components may not have a clear or interpretable meaning

# Applications of PCA

- **Data/Image Compression** : Reducing the dimensionality of image data while preserving key features
- **Noise Reduction** : Eliminating noise by discarding low-variance components
- **Feature Extraction** : Selecting the most important features from a large set for better model performance
- **Data Visualization** : reducing the dimension to 2D or 3D, in order to visualize the dataset
- **Speed Up Computation** : reduces load on memory