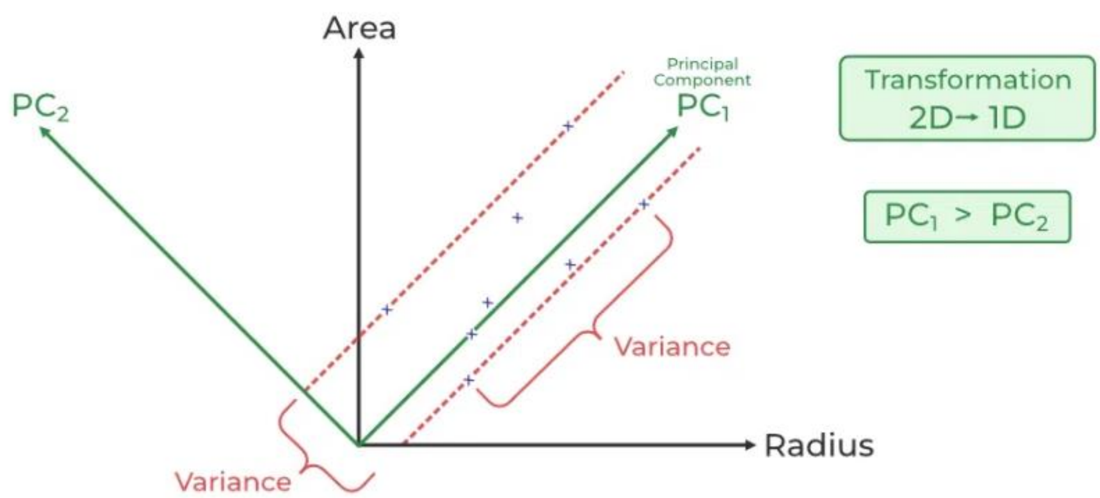
Principal Component Analysis (PCA) By:Loga Aswin

- PCA is an unsupervised learning method focusing on dimensionality reduction in machine learning.
- It uses statistical processes to transform correlated features into a set of linearly uncorrelated ones via orthogonal transformation.
- These transformed features are referred to as Principal Components (PCs).
- PCA aids exploratory data analysis and predictive modeling by unveiling strong patterns in datasets.
- Its goal is to identify a lower-dimensional surface for projecting high-dimensional data.
- PCA operates by prioritizing variance in each attribute, as higher variance indicates a more distinct split between classes, facilitating dimensionality reduction.



Steps for PCA:

- **Data Standardization:**

  Standardize the dataset to ensure consistent scales across features.
- **Covariance Matrix and Eigendecomposition**

  Compute the covariance matrix and perform eigendecomposition to find eigenvalues and eigenvectors.

- **Selection of Principal Components**

  Choose top eigenvalues and corresponding eigenvectors to form principal components.

- **Projection onto New Feature Space**

  Project the original data onto the principal components to create a lower-dimensional representation.

- **Variance Analysis**

  Analyze the variance explained by each principal component to understand data structure.

- **Utilization for Modeling or Analysis**

  Apply the reduced-dimensional data for modeling or analysis purposes to benefit from lower complexity and computational requirements.

Terminologies:

- Eigenvalues and Eigenvectors: Eigenvalues represent scaling factors, while eigenvectors signify directions of maximal variance in the data.

- Covariance Matrix: A matrix displaying the relationships and variances among different features in the dataset.

- Principal Components (PCs): Transformed variables derived from the original features, ordered by variance explained.

- Explained Variance: The amount of variance in the data accounted for by each principal component.

- Dimensionality Reduction: Process of reducing variables while retaining significant information.

- Projection: Transformation of data onto a new feature space defined by the principal components.

Why We Apply PCA:

- **Dimensionality Reduction:** It helps in reducing the number of features while retaining the essential information, reducing computational complexity and potential overfitting.
- **Multicollinearity Handling:** PCA addresses multicollinearity issues by transforming correlated features into a set of uncorrelated variables.
- **Data Visualization:** It aids in visualizing high-dimensional data by projecting it onto a lower-dimensional space while preserving significant variance.
- **Noise Reduction:** PCA can mitigate the impact of noise in data by focusing on the most significant variations.

Advantages of PCA:

- Dimensionality Reduction.
- Enhanced Model Performance.
- Data Interpretation

Disadvantages of PCA:

- Loss of Interpretability.
- Information Loss.
- Assumption of Linearity.

Applications of PCA:

- **Image and Signal Processing:** PCA is used in facial recognition, image compression, and denoising signals.

- **Bioinformatics:** Analyzing gene expression data and DNA microarrays.
- **Finance and Economics:** Reducing dimensions in financial data analysis, portfolio management, and risk assessment.
- **Medical Diagnostics:** Identifying patterns in medical imaging or patient data to assist in diagnoses.
- **Natural Language Processing:** Reducing dimensions in text analysis, document classification, and sentiment analysis.