


Explaining PCA (Principal Component Analysis)

What is PCA?

- PCA = Principal Component Analysis
- A way to simplify complex data
- Finds patterns and reduces dimensions
- Keeps the important information, drops redundancy

PCA in Simple Terms

-  Think of PCA as:
- Finding the best camera angle to capture the scene
- Packing for a trip: keeping only what matters
- Compressing a song or photo without losing the essence

PCA Methodology (Simplified)

- 1. Collect the data
- 2. Standardize features (make them comparable)
- 3. Build a correlation matrix (find relationships)
- 4. Find new directions (principal components)
- 5. Rank components using eigenvalues
- 6. Keep top components → reduce dimensions

Correlation Matrix

- Shows how features are related
- Detects redundancy in data
- Example: Height & Weight often move together
- Analogy: A friendship map (who always moves together)

Eigenvalues & Eigenvectors

- Eigenvectors = New directions (principal components)
 - → Like rotating the camera to the best angle
- Eigenvalues = Importance of each direction
 - → Brightness of the spotlight (big value = more info)
- • PCA keeps the directions with the biggest eigenvalues

PCA in Action

- 1. Correlation matrix → find related features
- 2. Eigenvectors → new smart directions
- 3. Eigenvalues → rank importance
- 4. Keep top directions → reduced dimensions

- Example: Fruits dataset
- New components: Sweetness Factor, Size Factor

Why PCA is Useful

- - Simplifies complexity
- - Removes redundancy
- - Speeds up analysis
- - Reveals hidden patterns
- - Makes visualization easier