Principal Component Analysis

Renato R. Maaliw III, *DIT*
Professor 1, College of Engineering
Southern Luzon State University
Lucban, Quezon, Philippines

What is PCA?

PCA is a technique that reduces the number of variables in a dataset while preserving as much important information as possible.

Think of it as summarizing a long book into a short summary that **still captures** the main ideas.

Why do we need PCA?

Simplification: Dealing with too many variables can be confusing. PCA helps by reducing the number.

Why do we need PCA?

Visualization: It’s hard to visualize data with many dimensions. PCA allows us to plot data in 2D or 3D.

Why do we need PCA?

Noise Reduction: It filters out insignificant details, focusing on what's truly important.

How does PCA work?

1. Collecting Data

Suppose you have information about different fruits:

Apple: Red color, round shape, sweet taste.
Banana: Yellow color, long shape, sweet taste.
Lemon: Yellow color, oval shape, sour taste.

You have multiple characteristics (variables) for each fruit.

2. Standardizing the Data

Before comparing, we need to make sure each characteristic is on the same scale. It’s like adjusting the units so everything is measured similarly.

3. Finding Patterns

PCA looks for patterns in the data. It identifies characteristics that **vary** the most among the fruits.

In our example, taste might vary more than color.

4. Creating Principal Components

These patterns are turned into new variables called **principal components**. Each principal component is a combination of the original characteristics.

Principal Component 1 (PC1): Might represent overall sweetness.
Principal Component 2 (PC2): Could represent shape differences.

5. Reducing Dimensions

We can now describe each fruit using these principal components instead of the original characteristics.

Often, the **first few principal components** capture most of the important information.

An Everyday Analogy

Think about taking a photo of a 3D object.

The photo is 2D, but it **still shows** the **essential features** of the object.

PCA does something similar by projecting **high-dimensional** data onto **fewer dimensions.**

Thank you very much for listening.