

# Statistics and Data Analysis

## Unit 05 – Lecture 03: Principal Component Analysis (PCA)

Tofik Ali

School of Computer Science, UPES Dehradun

February 14, 2026

<https://github.com/tali7c/Statistics-and-Data-Analysis>

# Quick Links

Overview

PCA Intuition

Explained Variance

Exercises

Demo

Summary

# Agenda

- 1 Overview
- 2 PCA Intuition
- 3 Explained Variance
- 4 Exercises
- 5 Demo
- 6 Summary

# Learning Outcomes

- Explain PCA as a variance-maximizing linear projection

# Learning Outcomes

- Explain PCA as a variance-maximizing linear projection
- State why scaling is important before PCA

# Learning Outcomes

- Explain PCA as a variance-maximizing linear projection
- State why scaling is important before PCA
- Interpret explained variance ratio and scree plot

# Learning Outcomes

- Explain PCA as a variance-maximizing linear projection
- State why scaling is important before PCA
- Interpret explained variance ratio and scree plot
- Use PCA for visualization and noise reduction

# PCA Intuition: Key Points

- Find new axes (components) capturing maximum variance



# PCA Intuition: Key Points

- Find new axes (components) capturing maximum variance
- Components are orthogonal

# PCA Intuition: Key Points

- Find new axes (components) capturing maximum variance
- Components are orthogonal
- PC1 captures most variance

# Explained Variance: Key Points

- Explained variance ratio per component

# Explained Variance: Key Points

- Explained variance ratio per component
- Choose  $k$  via scree plot / cumulative variance target

# Explained Variance: Key Points

- Explained variance ratio per component
- Choose  $k$  via scree plot / cumulative variance target
- Validate downstream performance

# Exercise 1: Scaling

Why scale features before PCA?

# Solution 1

- To prevent large-unit features dominating variance.

## Exercise 2: Components

Are PCA components original features?



# Solution 2

- No; they are linear combinations.

## Exercise 3: Choosing $k$

If first 2 PCs explain 88% and you need 90%, what do you do?

## Solution 3

- Add next PC(s) until target reached.

# Mini Demo (Python)

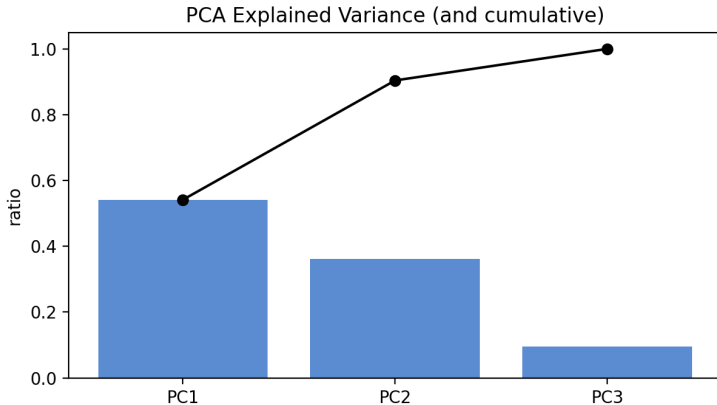
Run from the lecture folder:

```
python demo/demo.py
```

Outputs:

- images/demo.png
- data/results.txt

# Demo Output (Example)



# Summary

- Key definitions and the main formula.

# Summary

- Key definitions and the main formula.
- How to interpret results in context.

# Summary

- Key definitions and the main formula.
- How to interpret results in context.
- How the demo connects to the theory.



# Exit Question

Why might PCA improve a model even though it discards some variance?