Statistics and Data Analysis
Unit 05 – Lecture 03 Notes
Principal Component Analysis (PCA)

Tofik Ali

February 17, 2026

Topic

PCA: variance-maximizing projection; explained variance; scaling.

How to Use These Notes

These notes are written for students who are seeing the topic for the first time. They follow the slide order, but add the missing 'why', interpretation, and common mistakes. If you get stuck, look at the worked exercises and then run the Python demo.

Course repository (slides, demos, datasets): `https://github.com/tali7c/Statistics-and-Data-Analysis`

Time Plan (55 minutes)

- 0–10 min: Attendance + recap of previous lecture
- 10–35 min: Core concepts (this lecture's sections)
- 35–45 min: Exercises (solve 1–2 in class, rest as practice)
- 45–50 min: Mini demo + interpretation of output
- 50–55 min: Buffer / wrap-up (leave 5 minutes early)

Slide-by-slide Notes

Title Slide

State the lecture title clearly and connect it to what students already know. Tell students what they will be able to do by the end (not just what you will cover).

Quick Links / Agenda

Explain the structure of the lecture and where the exercises and demo appear.

- Overview

- PCA Intuition
- Explained Variance
- Exercises
- Demo
- Summary

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

Why these outcomes matter. **PCA** finds new axes (principal components) that capture maximum variance. It is a rotation of the feature space. Because PCA is variance-based, it is sensitive to scaling: standardize features first unless all features are already comparable. **Explained variance** tells how much of the total variance is captured by the first k principal components. Use it to choose the number of components: keep enough to capture most variance while still reducing dimensionality.

PCA Intuition: Key Points

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

Explanation. **PCA** finds new axes (principal components) that capture maximum variance. It is a rotation of the feature space. Because PCA is variance-based, it is sensitive to scaling: standardize features first unless all features are already comparable.

Explained Variance: Key Points

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

Explanation. **Explained variance** tells how much of the total variance is captured by the first k principal components. Use it to choose the number of components: keep enough to capture most variance while still reducing dimensionality.

Exercises (with Solutions)

Attempt the exercise first, then compare with the solution. Focus on interpretation, not only arithmetic.

Exercise 1: Scaling

Why scale features before PCA?

Solution

- To prevent large-unit features dominating variance.

Walkthrough. **PCA** finds new axes (principal components) that capture maximum variance. It is a rotation of the feature space. Because PCA is variance-based, it is sensitive to scaling: standardize features first unless all features are already comparable.

Exercise 2: Components

Are PCA components original features?

Solution

- No; they are linear combinations.

Walkthrough. **PCA** finds new axes (principal components) that capture maximum variance. It is a rotation of the feature space. Because PCA is variance-based, it is sensitive to scaling: standardize features first unless all features are already comparable.

Exercise 3: Choosing k

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

Solution

- Add next PC(s) until target reached.

Mini Demo (Python)

Run from the lecture folder:

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

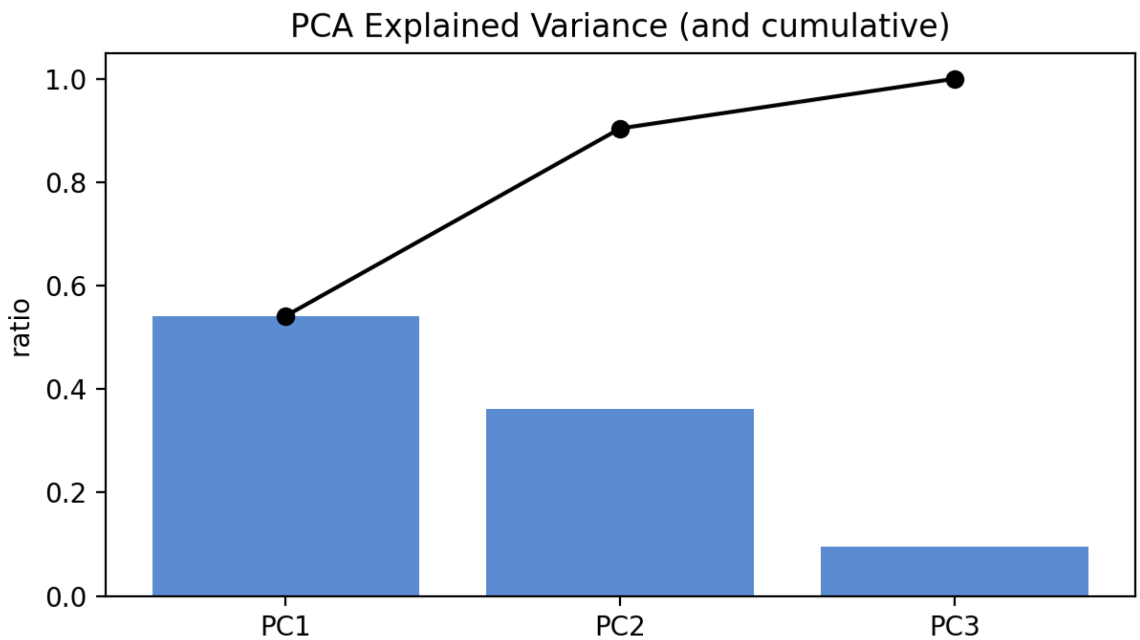
Output files:

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

What to show and say.

- Runs PCA on correlated features and plots the 2D projection.
- Reports explained variance ratios to discuss choosing k.
- Use it to show why scaling changes PCA results dramatically.

Demo Output (Example)



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?

Suggested answer (for revision). PCA can remove noise/redundancy; discarding low-variance directions can improve generalization if those directions are mostly noise.

References

- Montgomery, D. C., & Runger, G. C. *Applied Statistics and Probability for Engineers*, Wiley.
- Devore, J. L. *Probability and Statistics for Engineering and the Sciences*, Cengage.

- McKinney, W. *Python for Data Analysis*, O'Reilly.

Appendix: Slide Deck Content (Reference)

The material below is a reference copy of the slide deck content. Exercise solutions are explained in the main notes where applicable.

Title Slide

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

Quick Links

Overview PCA Intuition Explained Variance Exercises Demo Summary

Agenda

- Overview
- PCA Intuition
- Explained Variance
- Exercises
- Demo
- Summary

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
- Components are orthogonal
- PC1 captures most variance

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)

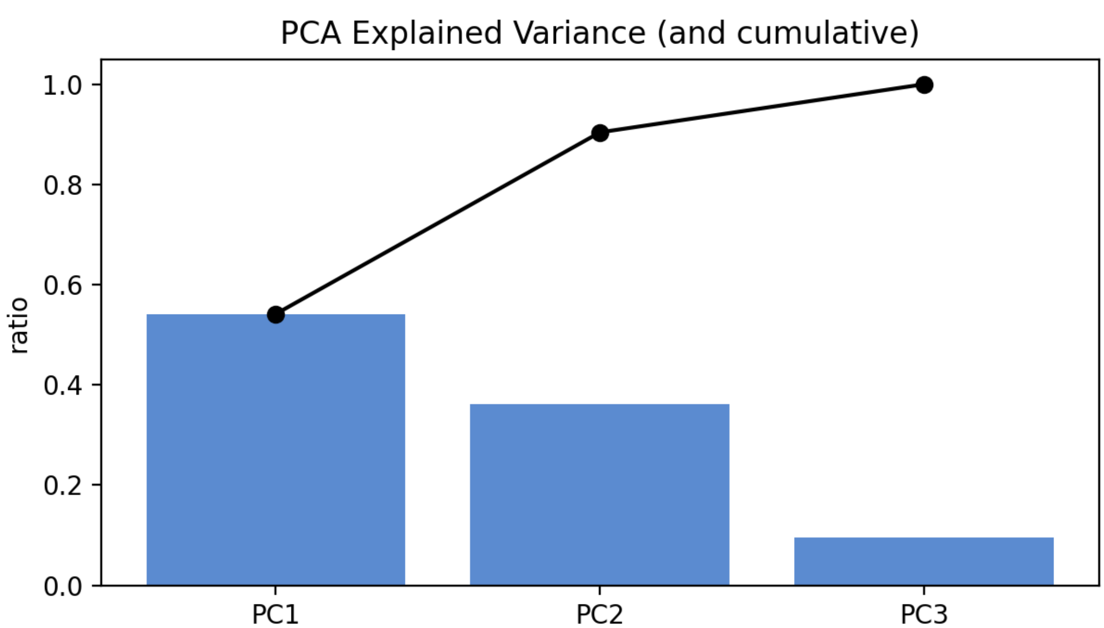
Run from the lecture folder:

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

Outputs:

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

Demo Output (Example)



Summary

- Key definitions and the main formula.
- 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?