Statistics and Data Analysis
Unit 05 – Lecture 03 Notes

Tofik Ali

February 14, 2026

Topic

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

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

Detailed Notes

These notes are designed to be read alongside the slides. They expand each slide bullet into plain-language explanations, small worked examples, and common pitfalls. When a formula appears, emphasize (1) what each symbol means, (2) the assumptions needed to use it, and (3) how to interpret the final number in the problem context.

PCA Intuition

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

Explained Variance

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

Exercises (with Solutions)

Exercise 1: Scaling

Why scale features before PCA?

Solution

- To prevent large-unit features dominating variance.

Exercise 2: Components

Are PCA components original features?

Solution

- No; they are linear combinations.

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.

Exit Question

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

Demo (Python)

Run from the lecture folder:

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

Output files:

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

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.