Statistics and Data Analysis
Unit 05 – Lecture 05 Notes

Tofik Ali

February 14, 2026

Topic

Nonlinear dimensionality reduction: kernel PCA and t-SNE (visualization).

Learning Outcomes

- Explain why nonlinear methods are sometimes needed
- Describe kernel PCA idea (high level)
- Describe t-SNE purpose (visualization) and pitfalls
- Choose PCA vs t-SNE appropriately

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.

Kernel PCA

- Implicitly map to higher-dimensional space via kernel
- Apply PCA in that space
- Captures nonlinear structure

t-SNE

- Mainly for 2D/3D visualization
- Preserves local neighborhoods
- Global distances can be misleading

Exercises (with Solutions)

Exercise 1: Use case

Name one warning when interpreting t-SNE plots.

Solution

- Global distances and cluster sizes can be misleading.

Exercise 2: Randomness

What should you do if t-SNE changes across runs?

Solution

- Set seed and check stability.

Exercise 3: Kernel PCA benefit

Why kernel PCA can help on circular data?

Solution

- It can capture nonlinear manifold structure.

Exit Question

Why should we avoid using t-SNE coordinates directly as model features (usually)?

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.