Statistics and Data Analysis

Unit 05 – Lecture 05: Kernel PCA and t-SNE

Tofik Ali

School of Computer Science, UPES Dehradun

February 17, 2026

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

Quick Links

Overview Kernel PCA t-SNE Exercises Demo Summary

Agenda

Learning Outcomes

- Explain why nonlinear methods are sometimes needed

Learning Outcomes

- Explain why nonlinear methods are sometimes needed
- Describe kernel PCA idea (high level)

Learning Outcomes

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

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

Kernel PCA: Key Points

- Implicitly map to higher-dimensional space via kernel

Kernel PCA: Key Points

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

Kernel PCA: Key Points

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

t-SNE: Key Points

- Mainly for 2D/3D visualization

t-SNE: Key Points

- Mainly for 2D/3D visualization
- Preserves local neighborhoods

t-SNE: Key Points

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

Exercise 1: Use case

Name one warning when interpreting t-SNE plots.

Solution 1

- Global distances and cluster sizes can be misleading.

Exercise 2: Randomness

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

Solution 2

- Set seed and check stability.

Exercise 3: Kernel PCA benefit

Why kernel PCA can help on circular data?

Solution 3

- It can capture nonlinear manifold structure.

Mini Demo (Python)

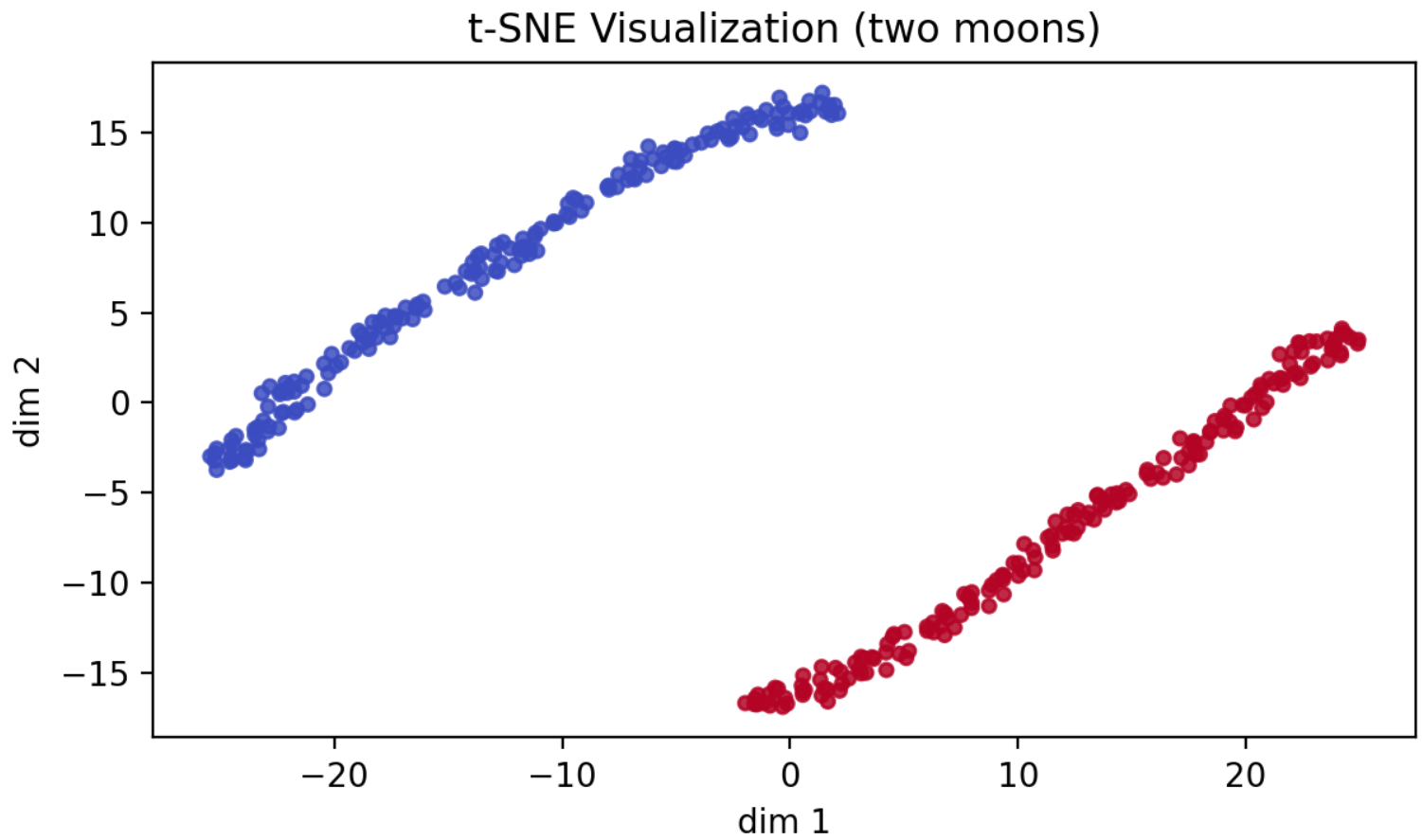
Run from the lecture folder:

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