L05: PCA & t-SNE

Dimensionality Reduction for Visualization and Preprocessing

Methods and Algorithms

Spring 2026

Outline

Learning Objectives

By the end of this lecture, you will be able to:

1. Apply PCA for dimensionality reduction and feature extraction
2. Interpret variance explained and choose number of components
3. Use t-SNE for visualization of high-dimensional data
4. Compare linear (PCA) vs non-linear (t-SNE) methods

Finance Application: Portfolio risk decomposition, asset clustering

---

From many features to meaningful low-dimensional representations

The Business Problem

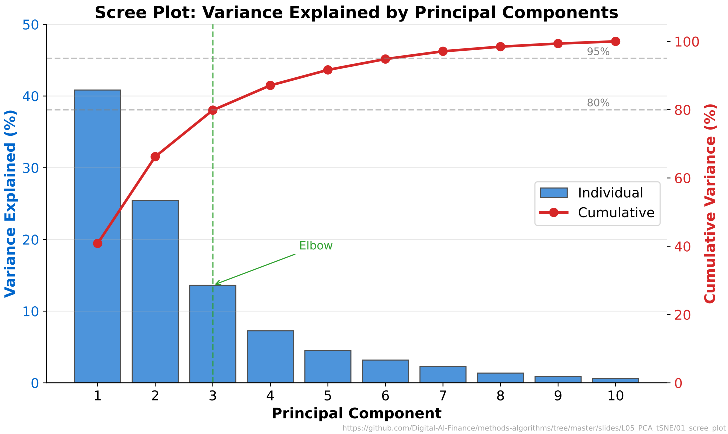
Curse of Dimensionality

- Portfolio with 100+ assets: hard to visualize relationships
- Customer data with dozens of features: redundant information
- High dimensions cause sparsity and computational issues

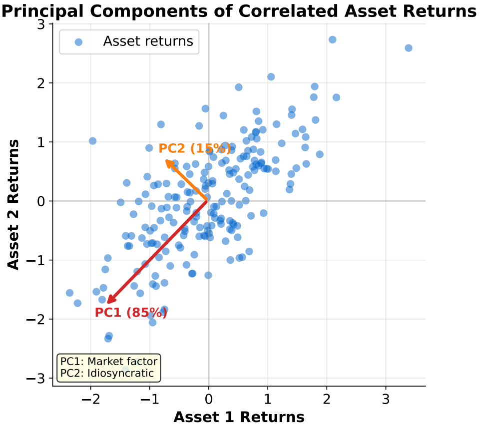
Solutions

- **PCA**: Linear projection preserving maximum variance
- **t-SNE**: Non-linear embedding preserving local structure

Reduce dimensions while preserving important information



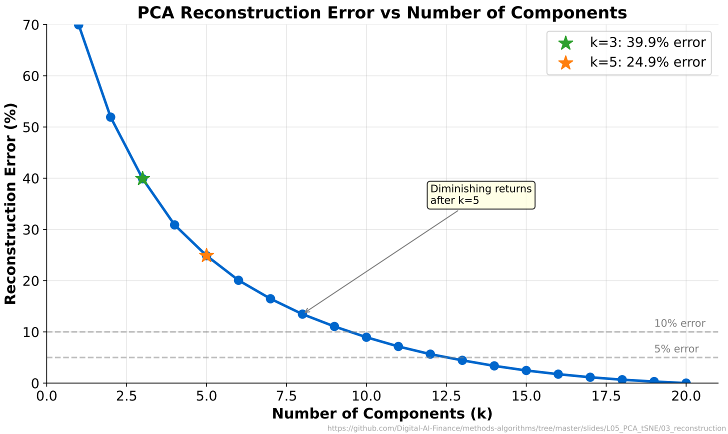
Choose k components capturing 80-95% of variance, or at the "elbow"



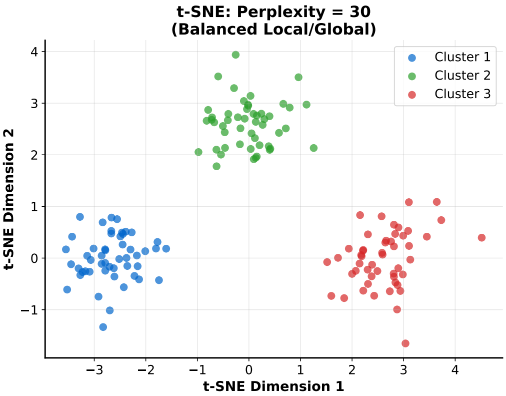
https://github.com/Digital-AI-Finance/methods-algorithms/tree/master/slides/L05_PCA_tSNE/02_principal_components

Principal components are orthogonal directions of maximum variance

Reconstruction Error



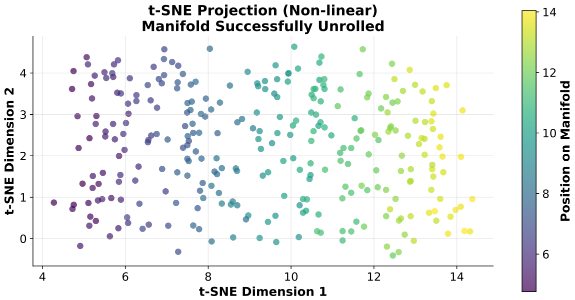
More components = lower error, but diminishing returns after elbow



https://github.com/Digital-AI-Finance/methods-algorithms/tree/master/slides/L05_PCA_tSNE/04b_tsne_perplexity_30

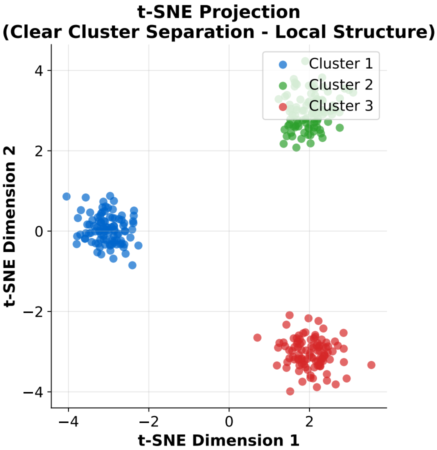
Perplexity controls local vs global structure preservation (try 5-50)

PCA vs t-SNE: Swiss Roll



https://github.com/Digital-AI-Finance/methods-algorithms/tree/master/slides/L05_PCA_tSNE/05b_tsne_swiss_roll

t-SNE unrolls non-linear manifolds that PCA cannot handle



https://github.com/Digital-AI-Finance/methods-algorithms/tree/master/slides/L05_PCA_tSNE/06c_tsne_cluster_projection

t-SNE better preserves cluster structure for visualization

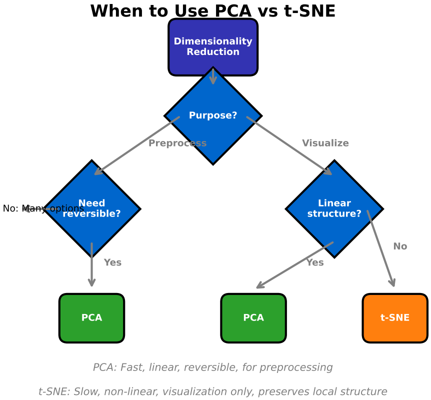
Hands-on Exercise

Open the Colab Notebook

- Exercise 1: Apply PCA to high-dimensional finance data
- Exercise 2: Visualize clusters with t-SNE
- Exercise 3: Compare PCA vs t-SNE for different datasets

Link: `https://colab.research.google.com/` [TBD]

Decision Framework



https://github.com/Digital-AI-Finance/methods-algorithms/tree/master/slides/L05_PCA_tSNE/07_decision_flowchart

PCA for preprocessing/speed, t-SNE for visualization only

References

- Jolliffe, I.T. (2002). *Principal Component Analysis*. Springer.
- van der Maaten, L. & Hinton, G. (2008). *Visualizing Data using t-SNE*. JMLR.
- James et al. (2021). *Introduction to Statistical Learning*. `https://www.statlearning.com/`