L05: PCA & t-SNE
Dimensionality Reduction for Visualization and Preprocessing

Methods and Algorithms – MSc Data Science

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

01_scree_plot/chart.pdf

02_principal_components/chart.pdf

03_reconstruction/chart.pdf

04_tsne_perplexity/chart.pdf

05_pca_vs_tsne/chart.pdf

06_cluster_preservation/chart.pdf

07_decision_flowchart/chart.pdf