Nonlinear Dimensionality Reduction: t-Stochastic Neighbor Embedding (t-SNE)

Prof. Endri Raco

Polytechnic University of Tirana

Erasmus+ Exchange Professor

Advanced Multivariate Analysis

Polytechnic University of Catalonia (UPC)

October 15, 2025

Welcome to Advanced Multivariate Analysis

Today's Journey

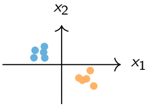
- **2-hour deep dive** into t-SNE
- **Mathematical foundations** to practical insights
- **Three key parts:**
  1. SNE - The original idea
  2. t-SNE - Solving the crowding problem
  3. Hyperparameters & interpretation

$$\boxed{p_{j|i} = \frac{e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma_i^2}}{\sum_{k \neq i} e^{-\|\mathbf{x}_i - \mathbf{x}_k\|^2/2\sigma_i^2}}}$$

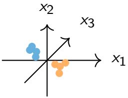
Core t-SNE Formula (We'll derive this today)

The Curse of Dimensionality

Our Intuition Works Here:



2D: Simple



3D: Manageable

But Not Here:

100D? 1000D?

Problems:

- Distance concentration
- Volume: $V_n(r) \propto r^n$
- Sparse data

"Geometric intuition fails"

Goals of Dimensionality Reduction

Goal 1: Visualization

$\mathbf{x} \in \mathbb{R}^{784}$
(MNIST)

$\downarrow$ compress

$\mathbf{y} \in \mathbb{R}^{2}$
(2D plot)

Key: "See" hidden structure

Goal 2: Feature Extraction

Noisy features

$\downarrow$ extract

Essential structure

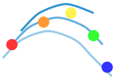
Benefits:

- Noise reduction
- Efficiency
- Better ML

Challenge: Preserve relationships while reducing dimensions

When Linear Methods Falter

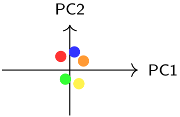
Swiss Roll Dataset



3D Manifold

True Structure:
2D manifold in 3D space

PCA Projection



PCA Result

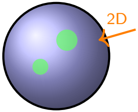
Problem:
Preserves variance,
destroys local structure

Need methods that preserve **local relationships**

The Manifold Hypothesis

"High-dimensional data often lies on or near a much lower-dimensional manifold"

Example: Earth's Surface



2D surface in 3D

Mathematical Form

Data: $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$
where $\mathbf{x}_i \in \mathbb{R}^D$

Assumption:
$\exists$ manifold $\mathcal{M}$ with dim $d \ll D$:

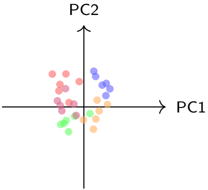
$$\mathbf{x}_i \approx f(\mathbf{z}_i) + \epsilon_i$$

- $\mathbf{z}_i \in \mathbb{R}^d$ (low-dim)
- $f : \mathbb{R}^d \to \mathbb{R}^D$
- ϵ_i (noise)

Goal: Uncover this hidden low-dimensional structure

Preserving Neighborhoods: t-SNE in Action

PCA on MNIST Digits

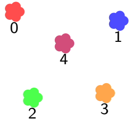


PCA: Mixed Clusters

Problems:

- Classes overlap significantly
- Linear projection limitations
- Poor cluster separation

t-SNE on MNIST Digits



t-SNE: Clear Separation

Advantages:

- Distinct clusters
- Preserves local structure
- Reveals true relationships

"t-SNE focuses on preserving **local neighborhood structure***"*
Similar points in high-D $\rightarrow$ nearby points in low-D

Today's Journey: From Theory to Mastery

Part 1: SNE
(Foundation)
Gaussian Kernels

KL Divergence
Symmetric SNE
Limitations

Part 2: t-SNE
(Innovation)
Crowding Problem

Student's t-dist.
Heavy Tails
Optimization

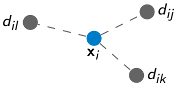
Part 3: Practice
(Application)
Perplexity & LR

Interpretation
Common Pitfalls
t-SNE vs UMAP

From Distances to Probabilities: The Core Idea

Central Insight: Convert distances between points into probabilities
that represent neighborhood relationships

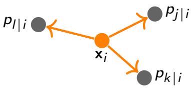
Traditional Approach



Euclidean Distances

Problem: How to weight different distances?

SNE/t-SNE Approach



Probabilities

Solution: Probabilities naturally normalize!

"What is the probability that point i would pick point j as its neighbor?"