Assessing the Quality of Dimensionality Reduction Methods based on Fuzzy Simplicial Sets

Objective

Quantifying **locally** how the structures of a dataset are preserved under a dimensionality reduction. We use **evaluation measures** of the form

$$\text{eval}(\mathcal{X}, \mathcal{X}', i)$$

where

- $\mathcal{X}$ is the original dataset
- $\mathcal{X}'$ is the reduced dataset
- i is the index of the point where faithfulness is evaluated.

Methodology

We propose evaluation measures that aims to preserve **Fuzzy Simplicial Sets** (FSS).
Why are FSS a good choice?

- Theoretical and empirical validation to quantify neighbourhood relationships,
- robust to the choice of parameters
- computationally efficient to construct.

Evaluation measures

We present different evaluation measures:

- eval_{FSS}: By considering the local contribution to the global loss used in UMAP.
- eval_{KL}: By normalizing the FSS and computing the KL divergence.
- eval_{JSD}: Computing the JSD divergence between the normalized FSS.

A balanced measure

The measure eval_{JSD} shows some desirable properties:

- It measures **precision and recall**
- Its value is bounded between 0 and 1 which provides a **good interpretability**.

Applications

We demonstrate the usefulness of the evaluation measures to:

- **Reconstruct breaks in the structure**: Detecting regions that are close in the original dataset and not in the reduced one.
- **Detect intrinsic dimensionality**: The average local error increases when the embedding is made under the intrinsic dimensionality of the dataset.

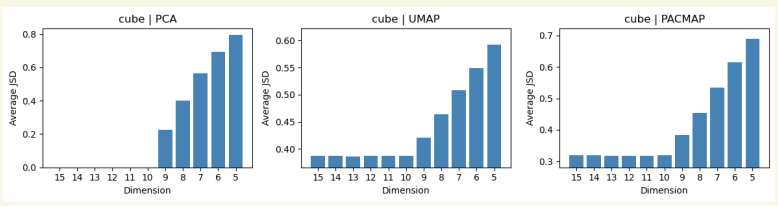

Figure: Average unfaithfulness values (eval_{JSD}) for PCA (left) UMAP (center) and PACMAP (right) applied to the embeddings of a 10-dimensional cube progressively to lower dimensions.

Benchmark datasets

We use 3 simple datasets to validate our evaluation measures

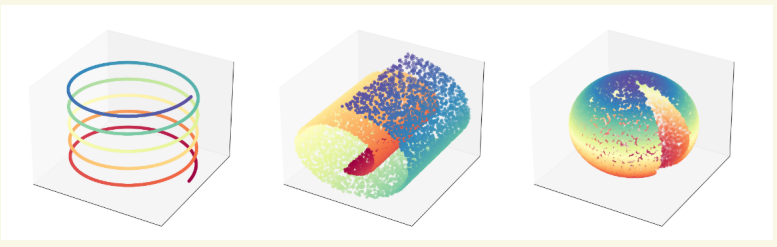

Figure: Benchmark datasets used.

Figure: Benchmark datasets embedded

We aim to detect the mistakes of the dimensionality reduction:

- The helix self-intersects in different locations.
- The Swissroll folds over itself, nearly breaks in some points, and develops artificial holes.
- The cut of the sphere is enhanced, and some artificial holes are created.

Results

The evaluation measures highlights areas where distortions occur.

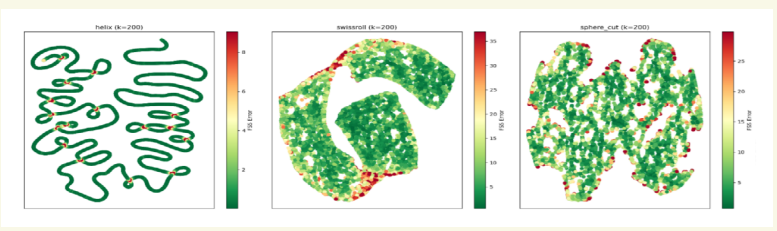

Figure: Clipped eval_{FSS} values

Victor Reyes Stephane Marchand-Maillet

Victor.Reyesmartin@unige.ch

University of Geneva - SISAP 2025