

**Evaluation of the effectiveness of  
state-of-the-art dimensionality reduction  
techniques, such as PaCMAP, UMAP,  
TriMAP, IVHD and t-SNE, on synthetic  
or sample datasets.**

Kamil Kwarciak, Daniel Kuc, Filip Reka

# 1 Introduction

In recent years, the advancement of technology and the accumulation of massive amounts of data have become an integral part of many scientific and industrial fields. As data becomes increasingly complex, we encounter challenges related to its analysis and interpretation. One such challenge is the processing and visualization of high-dimensional data. In such cases, dimensionality reduction techniques are incredibly useful. They allow us to transform data from a high-dimensional space to a lower-dimensional space while preserving as much information as possible. Dimensionality reduction techniques can help solve various problems such as noise reduction, elimination of correlations between variables, and visualization of data in a low-dimensional space.

## **TODO TUTAJ BEDZIE WIZUALIZACJA**

Within this project, we will focus on several advanced dimensionality reduction techniques that are commonly used.

- **PaCMAP** (Pairwise Controlled Manifold Approximation) is a technique that relies on utilizing neighborhood information and relationships between data points.
- **UMAP** (Uniform Manifold Approximation and Projection) is a method that combines the advantages of different techniques such as t-SNE and LargeVis to generate low-dimensional representations of data.
- **TriMAP** is a method that integrates topological and geometric information to create three-dimensional visualizations of data.
- **t-SNE** (t-Distributed Stochastic Neighbor Embedding) is a dimensionality reduction technique that transforms data into a low-dimensional space while preserving local relationships between points.
- **IVHD** (Interactive Visualization of High-Dimensional Data) is a modern and fast method that is a simplified implementation of MDS (Multidimensional Scaling). IVHD is based on the nearest neighbor graph.

We will provide a detailed description of the selected dimensionality reduction techniques and the methodology for evaluating their effectiveness. We will then conduct experiments on synthetic and sample datasets, analyzing the results and visualizations. Based on the collected data, we will be able to draw conclusions regarding the effectiveness of each technique and their potential applications.

## 2 Datasets description

The main goal of the project is to analyze datasets that do not exhibit linear separability. Consequently, the popular dimensionality reduction method, PCA, becomes insufficient. It is worth mentioning in this context the modification of this method known as kernel-PCA, which is used for non-linearly separable datasets. However, we will not focus on it in this article.

The proposed datasets are a creative reflection of imagination. They consist of generalized combinations of spaces with any number of dimensions, such as:

- hyperspheres
- hypercubes
- hypertoruses

### 2.1 First Dataset

...

### 2.2 Second Dataset

...

Based on the above datasets, we will conduct a study to examine the impact of changing the number of samples and data dimensionality on the effectiveness of the analyzed dimensionality reduction techniques.

### 3 Used methods

## 4 Datasets visualization

## 5 Conclusions