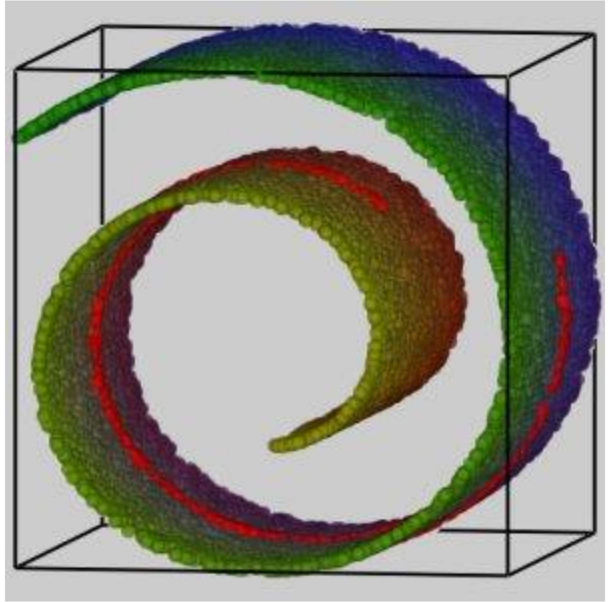
Fundamentals of AI

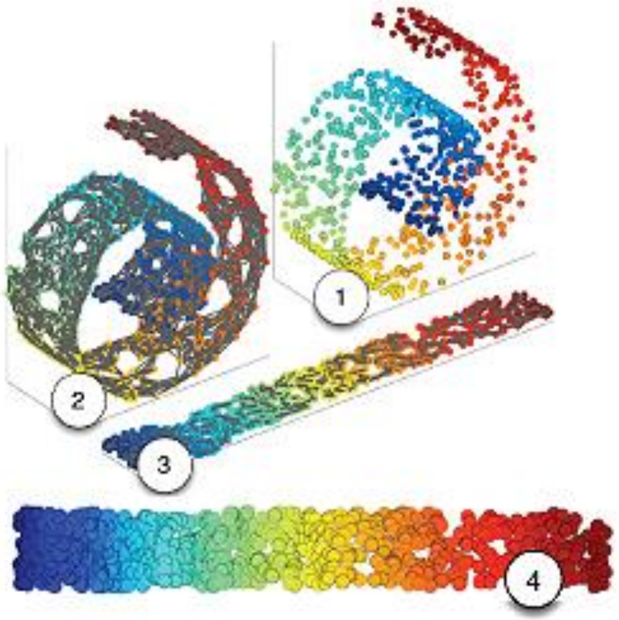
Manifold learning and non-linear dimred

Introduction to manifold learning

Notion of manifold

- *m*-dimensional manifold is a topological space with the property that each point has a neighborhood that is homeomorphic to the Euclidean space of dimension *m*
- m-dimensional manifold can be embedded into the R^p, where $p>m$
- Manifold can be characterized by geometry and topology
- Manifolds need not to be connected or closed
- Notion of geodesic distance





Manifold learning

- *Principal manifold*: an m-dimensional manifold which ‘passes through the middle of the data point cloud’
- PCA constructs *linear principal manifold*
- Any matrix factorization defines a *linear manifold* (ICA, NMF)
- Assumption of uniformly sampled manifold as a generative data model
- Sometimes no explicit manifolds are learnt, only projections onto hypothetical manifolds (projective methods)
- In other cases, we learn explicit manifold geometry or topology (injective methods)
- What is learnt is frequently not a manifold (strictly speaking)!

Manifold learning and dimensionality reduction

- Frequently used in the sense of *non-linear dimensionality reduction*
- Main role of manifold learning is for data visualization in 2D or 3D
- Frequently, manifold learning is applied after other ways of dimensionality reduction (e.g., after PCA): more stable results

Use of kNN or ε-graphs

- Many (but not all!) of the methods start with reconstructing local proximity relations
- kNN graph connects a data point with k nearest neighbours
- ε-graph connects a data point with any other data point in a ball of radius ε
- Common confusion: kNN graph or ε-graph is not a ‘manifold’

Meaning of ‘embedding’

- In mathematics, an embedding (or imbedding) is one instance of some mathematical structure contained within another instance
- ‘m-Manifold is embedded in R^p’
- ‘lower-dimensional embedding’ (e.g., Locally Linear Embedding (LLE)) = result of projection
- ‘embedded space’ = low-dimensional (or ***base*** *or* ***latent***) space
- Graph embeddment

What you shall learn in this Lecture

- Be able to distinguish **projective** and **injective** methods of manifold learning
- Have idea about **principal manifolds** (non-linear PCA) and their generalizations
- Understand **t-SNE** and **UMAP** as the most used projective methods, and know their parameters
- Understand the principles of **artificial neural network-based methods** for manifold learning, especially the principles of **generative data modeling**