28-03-2024

Dimensionality reduction

Dimensionality reduction (DR)

MNIST handwritten digit data (28x28=784)

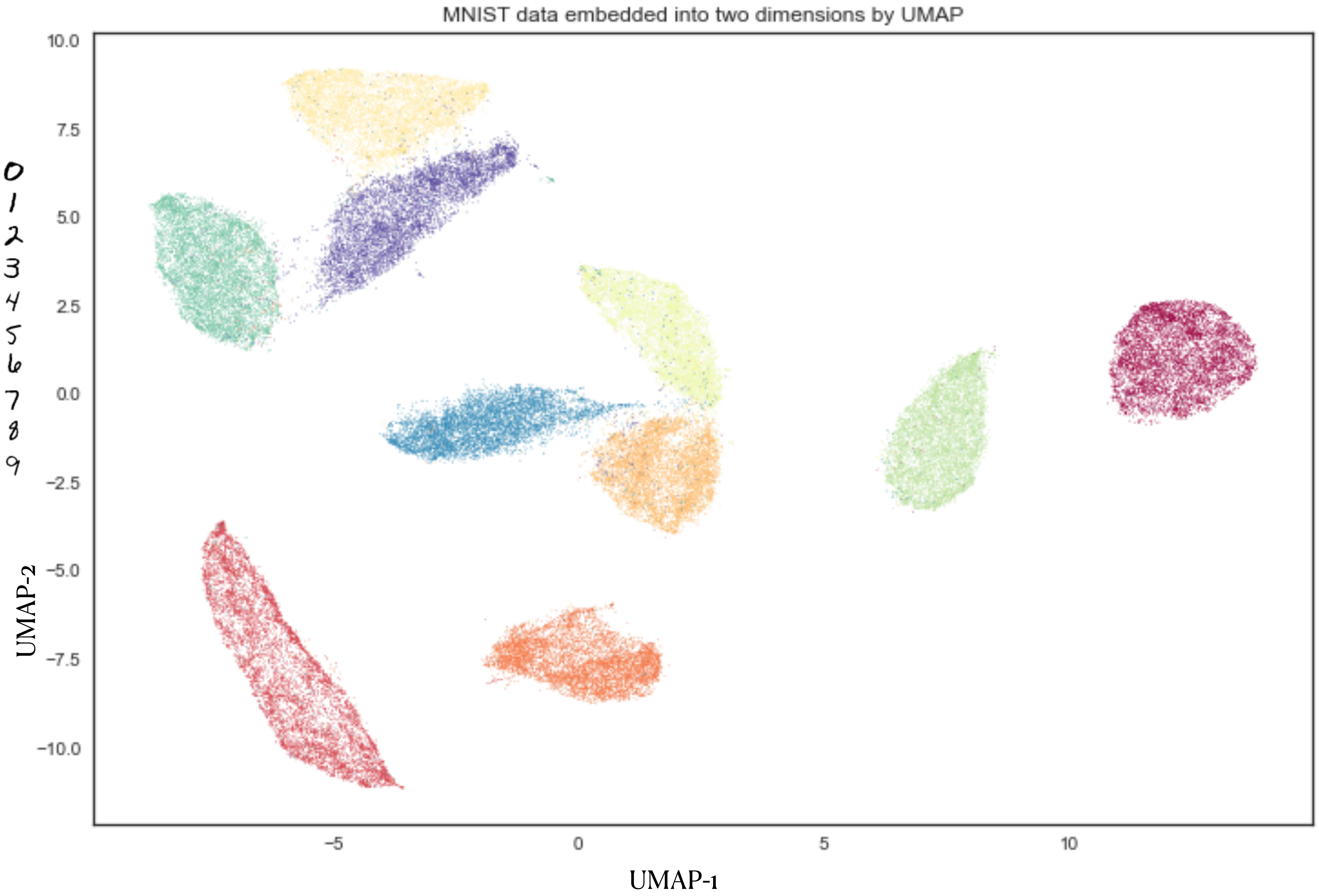
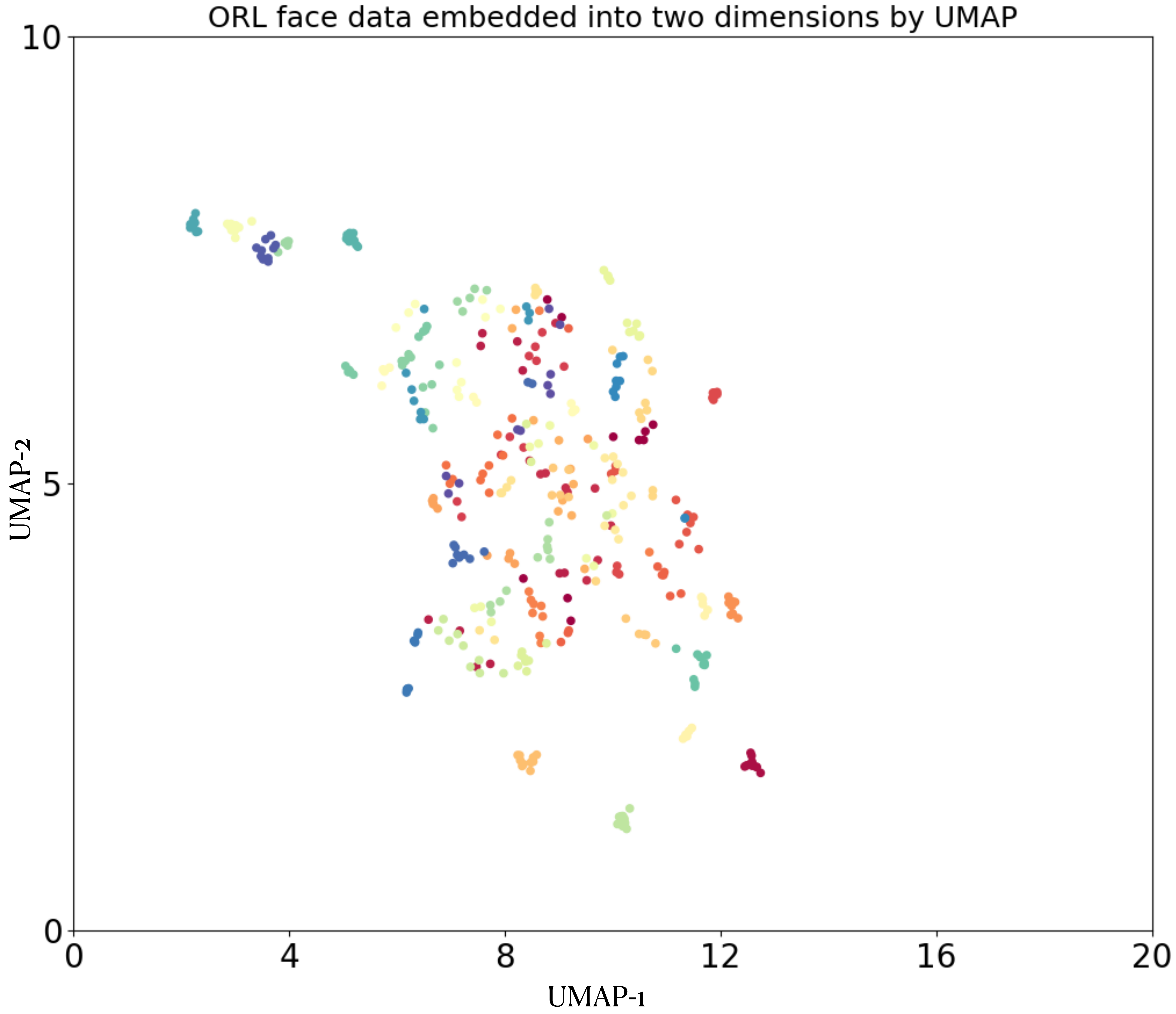


Image: Wikipedia

Dimensionality reduction (DR)

ORL face data (64x64=4096)

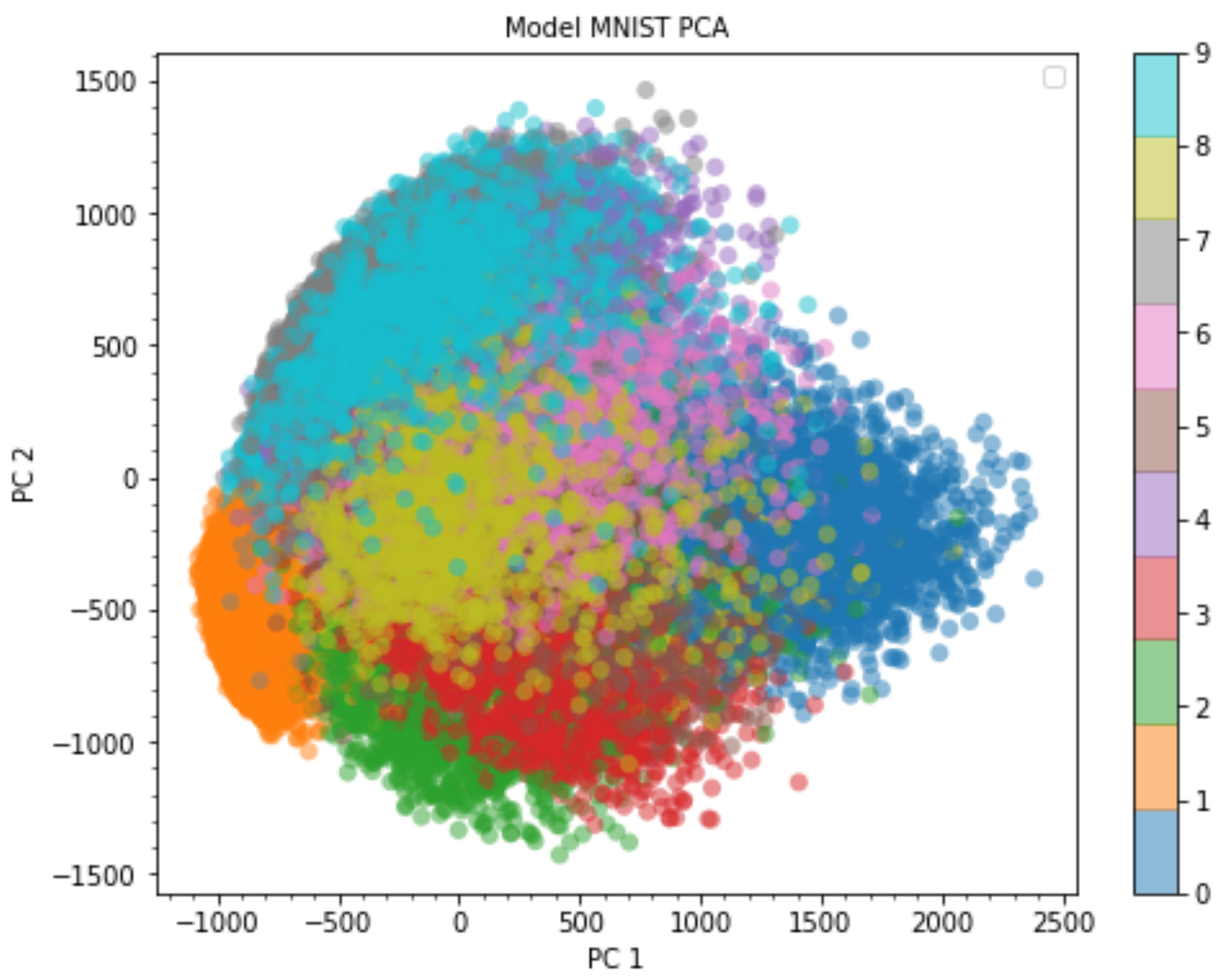


Dimensionality reduction (DR)

- This is a another class of ML algorithms
- Represent high dimensional data into a (very)-low dimensions
- Let $X_1, X_2, \ldots, X_n$ be given data points (observation), $X_i \in R^d$
- Task is to find a low dimensional representation of X_i say $Z_i \in R^l$ such that $l < d$
  - Find a projection mapping $P : R^d \to R^l$ such that $Z_i = P(X_i)$; $i = 1, \cdots, n$
- There are two types of dimensionality reductions
  - Linear DR
  - Non-linear DR

Linear dimensionality reduction (LDR)

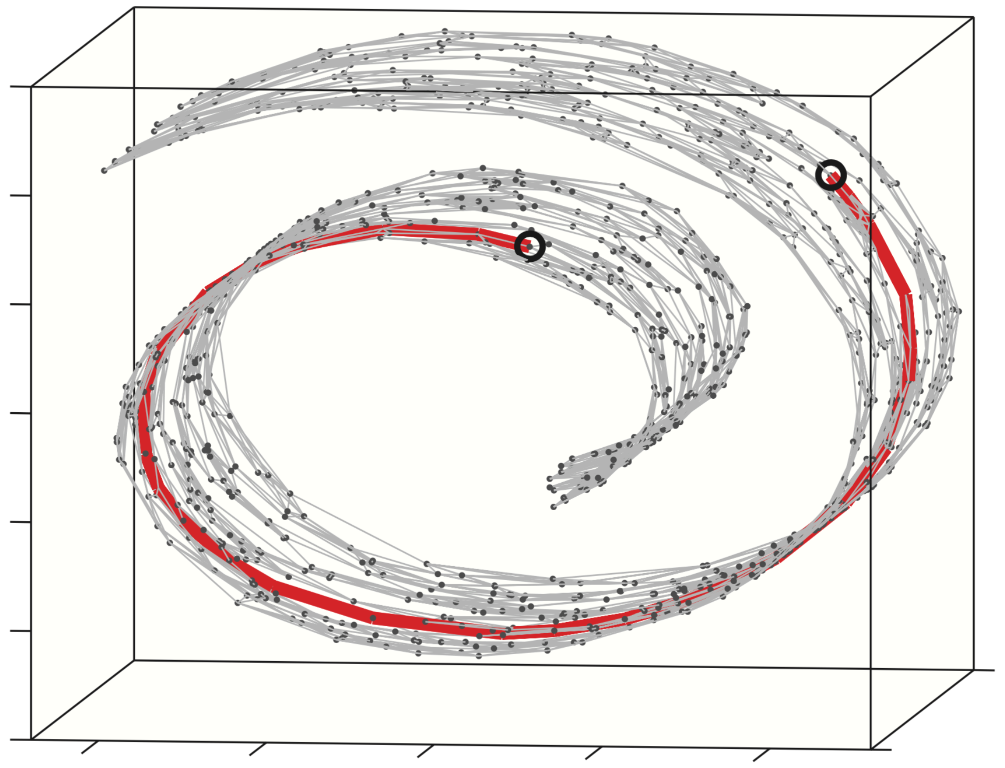
- Find a linear transformation $P \in R^{d \times l}$ with some objective $f_{X_i; i=1,\cdots,n}(:)$; where $l < d$
  - $Z_i = P^T X_i$
- Depending on the objective $f_{X_i; i=1,\cdots,n}(:)$, there are different types of LDR
  - PCA: Principle component analysis - Karl Pearson, On lines and planes of closest fit to systems of points in space,1901
  - MDS (metric): Multi-dimensional scaling - J B Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, 1964
  - …

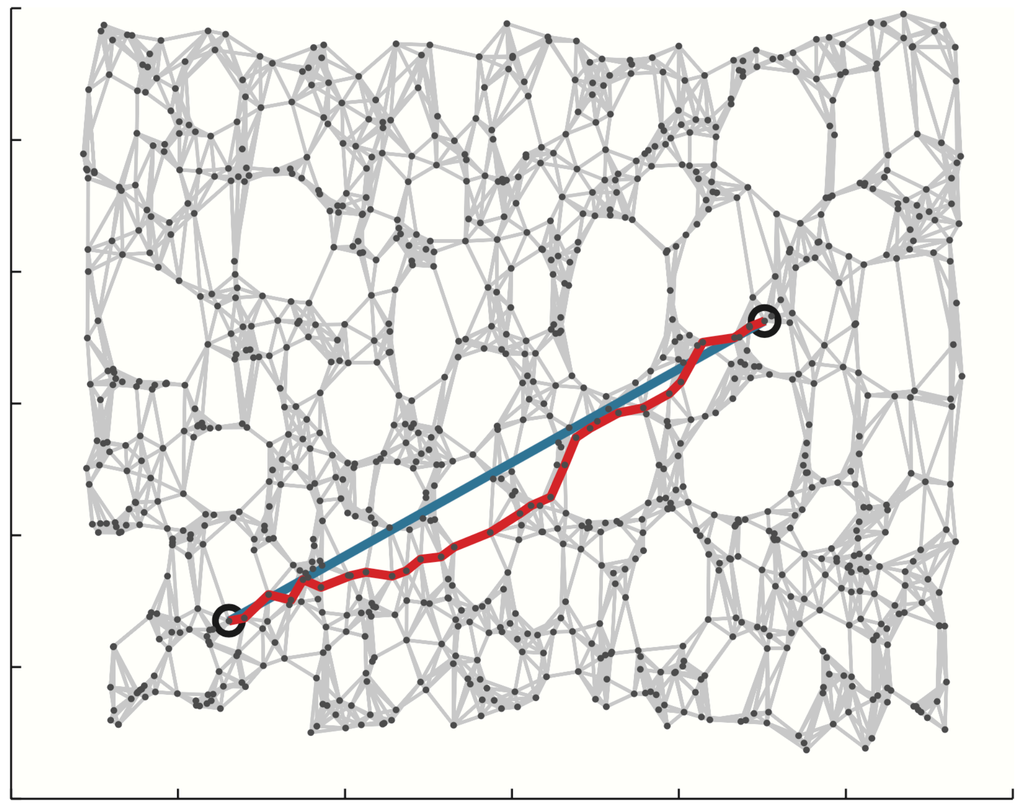


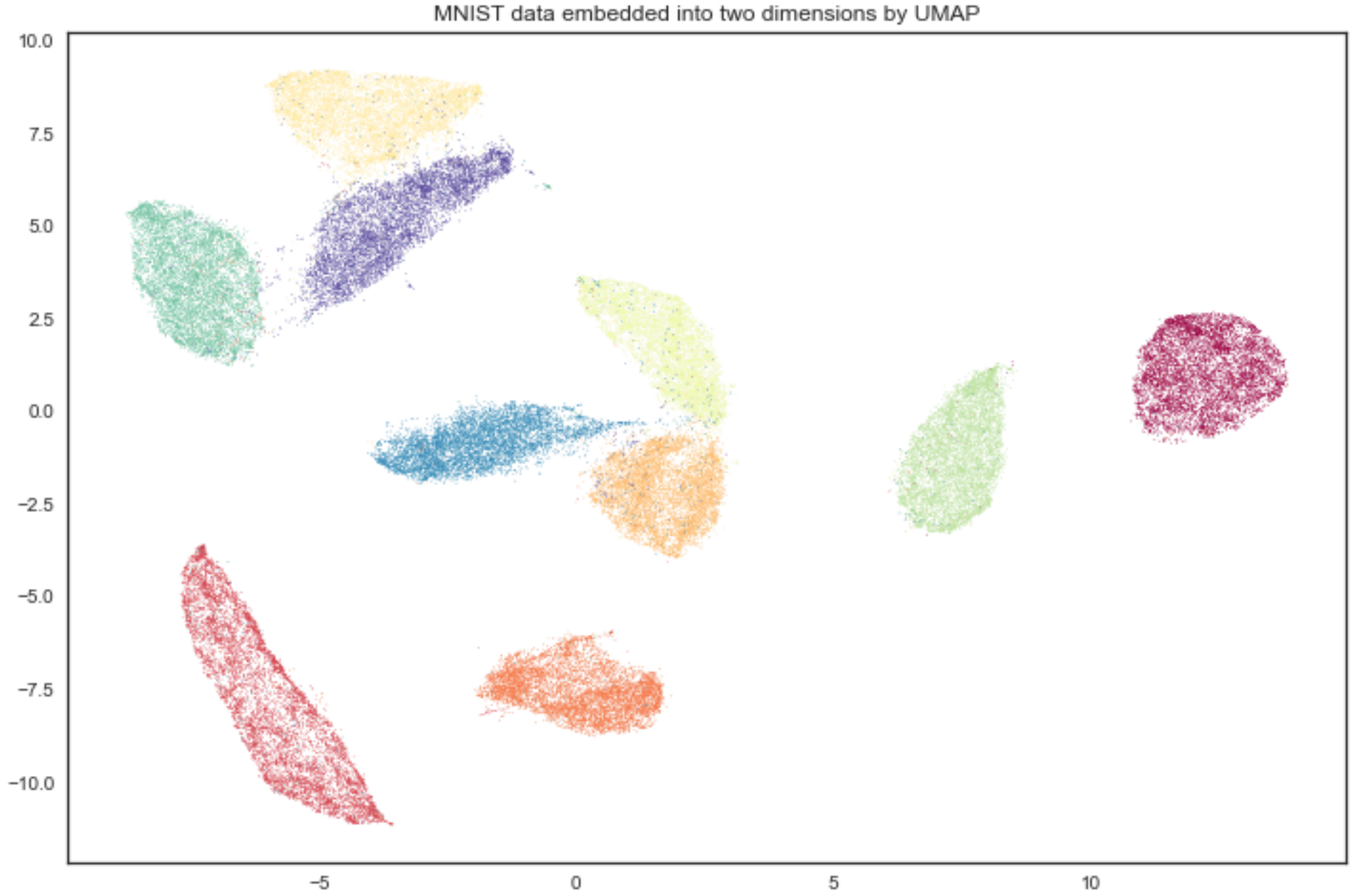
Images: https://zeta-learn.com/writeups/examples/samples.html

Non-linear dimensionality reduction (NLDR)

- Try to capture some spatial structure of the data when projected into lower dimension
  - Manifold learning
- Different types of NLDR
  - Isomap
  - Uniform Manifold Approximation and Projection (UMAP)
  - T-distributed stochastic embedding (t-SNE)



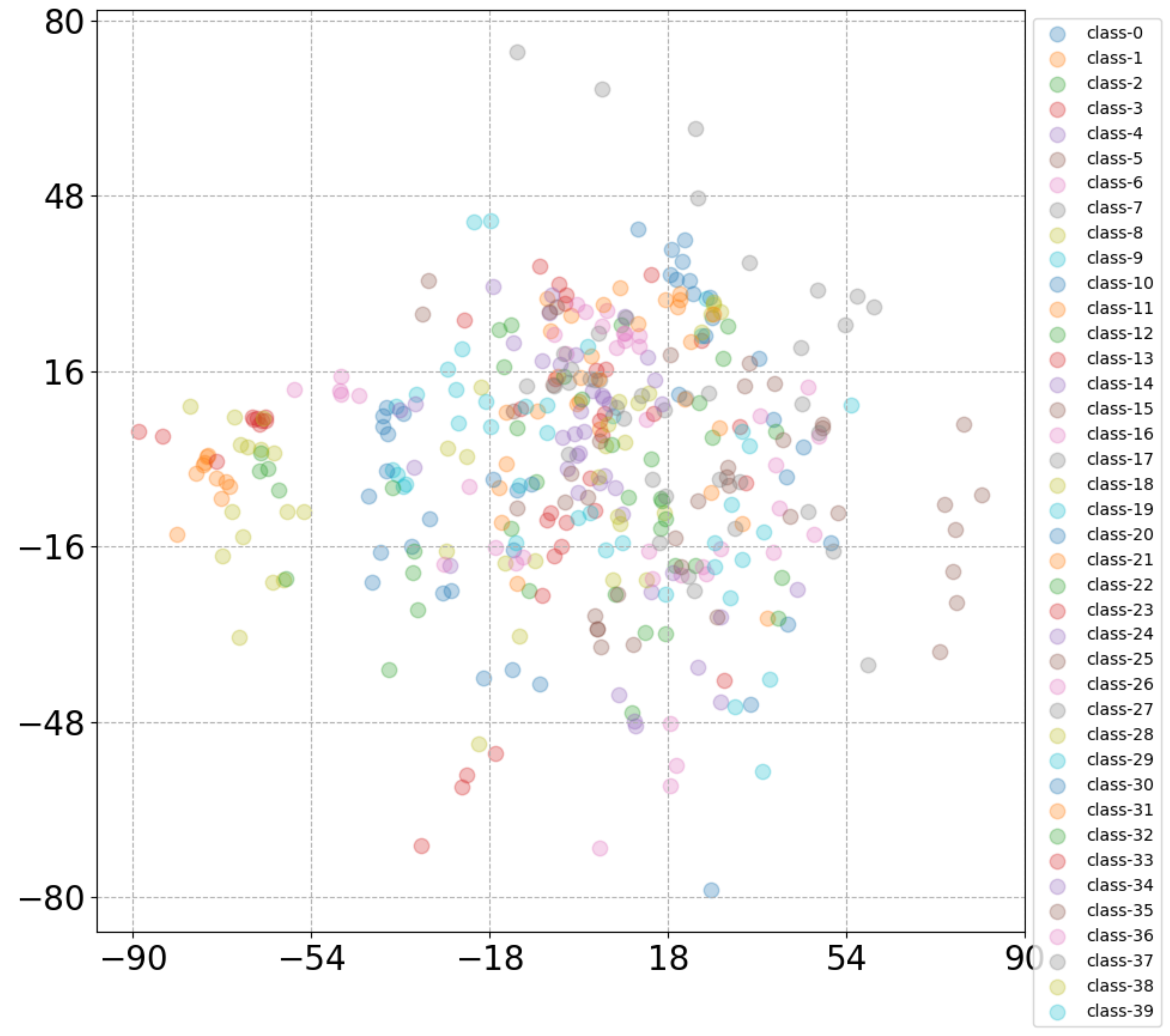


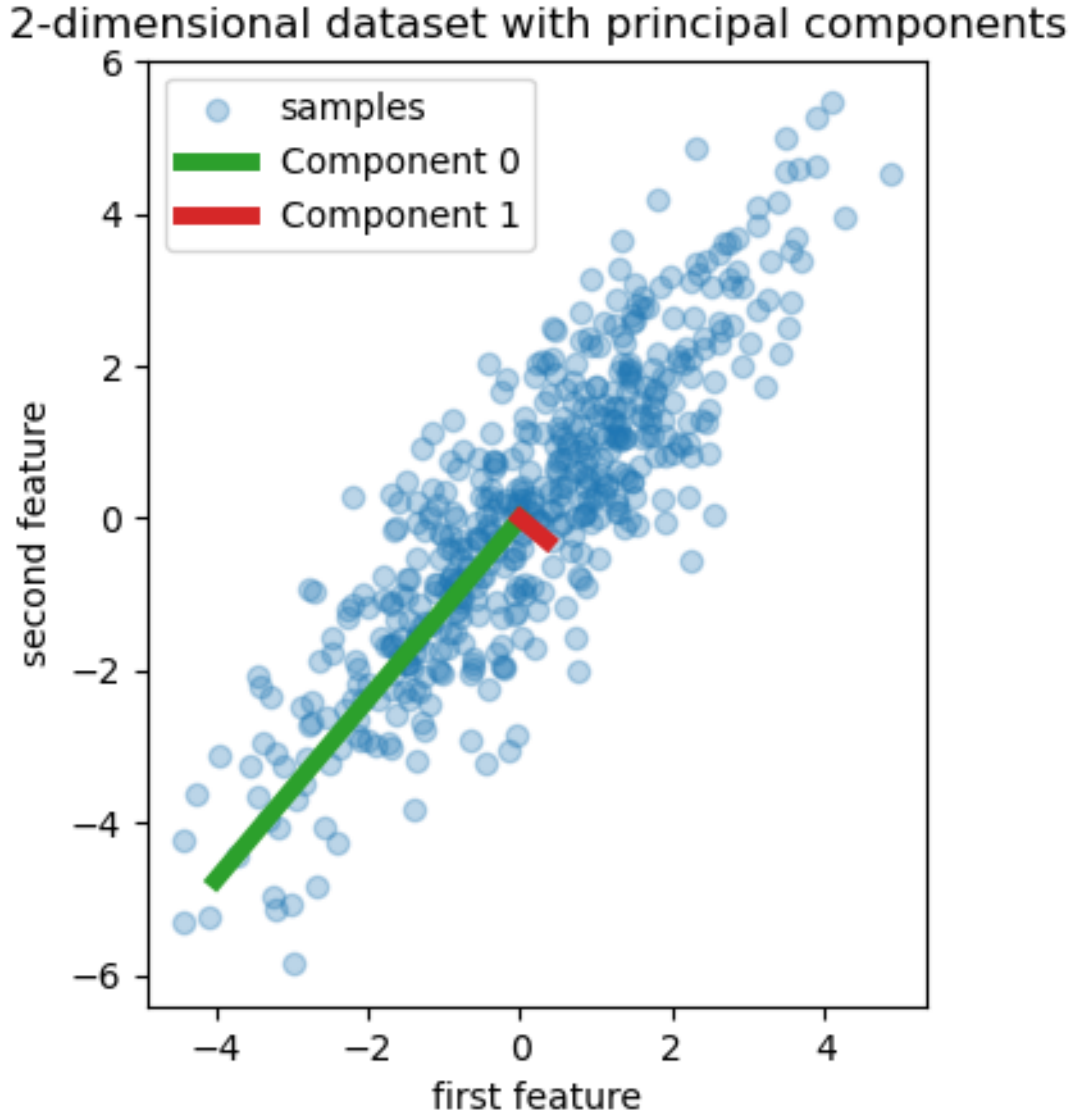


Images: Tenenbaum et al. A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, 2020

Principle component analysis (PCA)

- Let $X_1, X_2, \ldots, X_n$ be given data points (observation), $X_i \in R^d$
- Find a linear transformation $P : R^d \rightarrow R^l$ such that $Z_i = P^T X_i;\;\; i = 1,\cdots,n$
- Objective $f_{X_i;i=1,\cdots,n}(:)$
  - Maximize the variance of the projected data
  - Minimize the reconstruction error





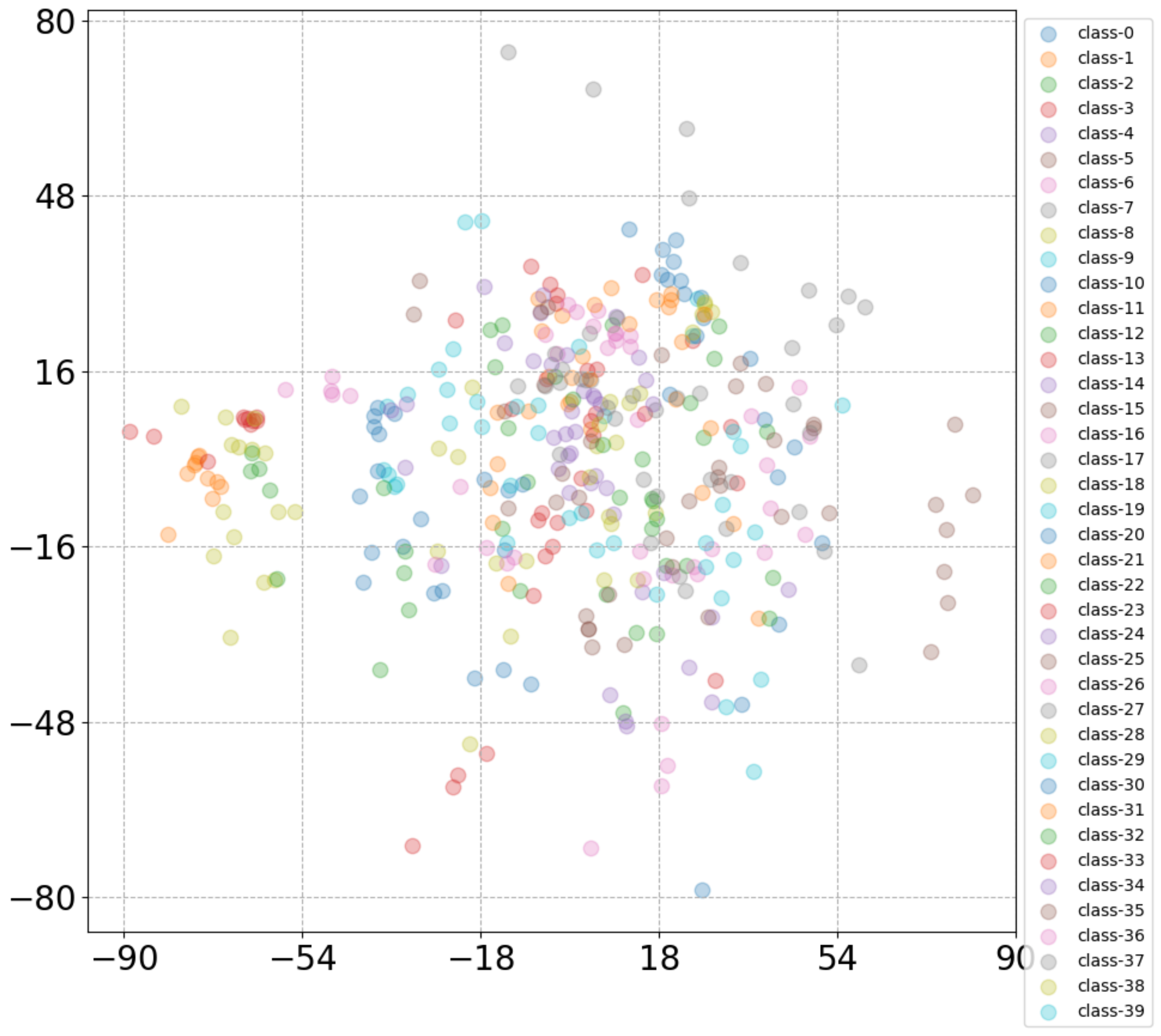
Images: https://scikit-learn.org/

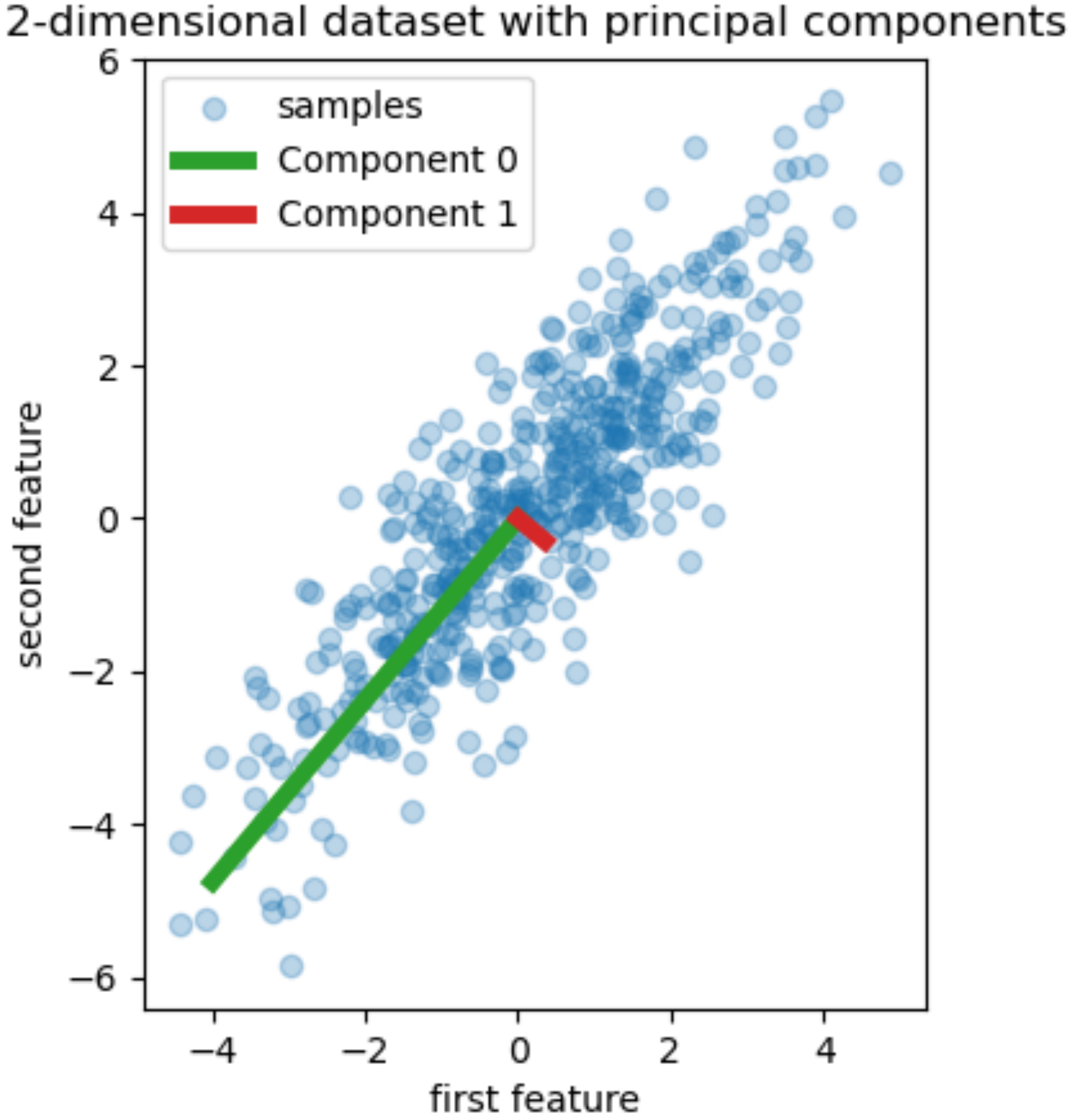
When PCA does not give an expected answer?

30-03-2024

Principle component analysis (PCA)

- Let $X_1, X_2, ..., X_n$ be given data points (observation), $X_i \in R^d$
- Find a linear transformation $P : R^d \rightarrow R^l$ such that $Z_i = P^T X_i;\ \ i = 1,\cdots,n$
- Objective $f_{X_i;i=1,\cdots,n}(:)$
  - Maximize the variance of the projected data
  - Minimize the reconstruction error





Images: https://scikit-learn.org/

Principle component analysis (PCA)

- Let $X_1, X_2, \ldots, X_n$ be given data points (observation), $X_i \in R^d$
- Find a linear transformation $P : R^d \rightarrow R^l$ such that $Z_i = P^T X_i;\ \ i = 1, \cdots, n$
- Objective $f_{X_i; i=1,\cdots,n}(:)$
  - Maximize the variance of the projected data
  - Minimize the reconstruction error
- What about computational complexity?
  - $n > d$?
  - $n < d$?

Multi-dimensional scaling (MDS)

- Given a distance matrix $D \in R^{n \times n}$, such that $d_{ij} = ||x_i - x_j||$
- Can we recover $X_1, X_2, \ldots, X_n; X_i \in R^d$?
- Classical MDS
  - $D \in R^{n \times n}$ is Euclidean
- Metric MDS
  - $D \in R^{n \times n}$ is any distance matrix
- Non-metric MDS
  - Not given $D \in R^{n \times n}$ as distance but in ordinal form
    - $d_{ij} < d_{ik}$

Images: U K Luxburg

Classic MDS

- $D \in R^{n \times n}$ is Euclidean
  - $d_{ij} = \|x_i - x_j\|^2 = \langle x_i - x_j, x_i - x_j \rangle$
  - $\langle x_i, x_j \rangle = \frac{1}{2}(\langle x_i, x_i \rangle + \langle x_j, x_j \rangle - d_{ij}^2)$
- Define Gram matrix $S \in R^{n \times n}$ using $s_{ij} = \langle x_i, x_j \rangle = \frac{1}{2}(\langle x_i, x_i \rangle + \langle x_j, x_j \rangle - d_{ij}^2)$
- Compute the eigenvalue decomposition $S = V \Lambda V^T$
- Now fix $d \leq n$ and let V_d be the first d column of V and Λ_d is the first d eigenvalues on the diagonal
- $X = V_d \sqrt{\Lambda_d}$; i-th row gives $X_i \in R^d$

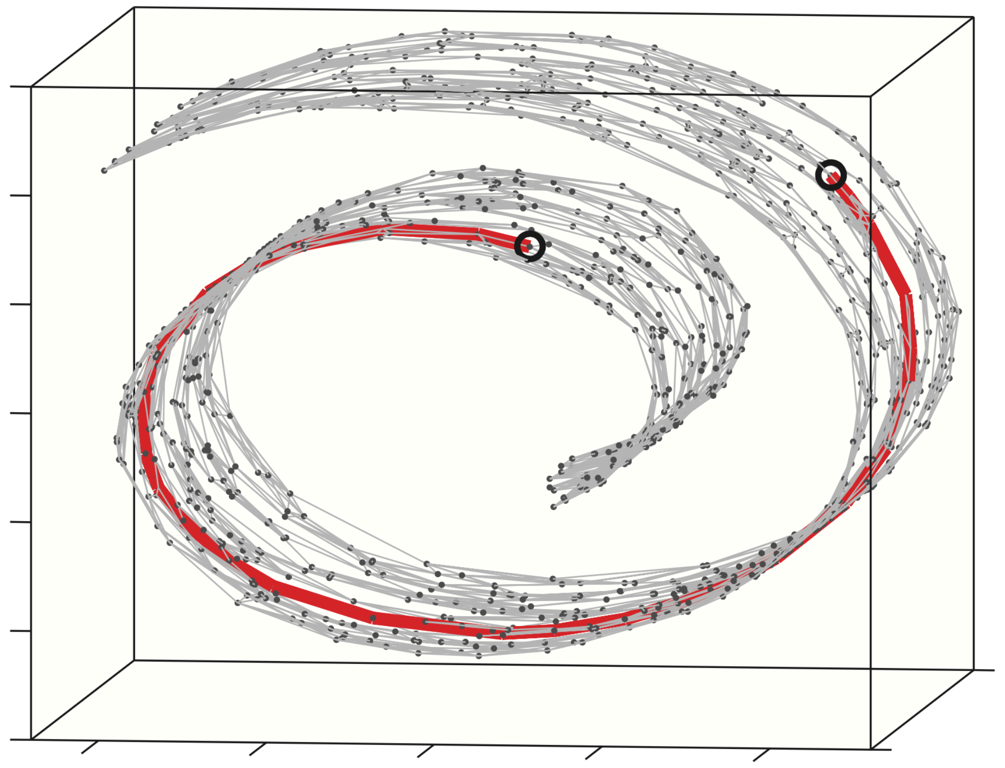
Metric MDS

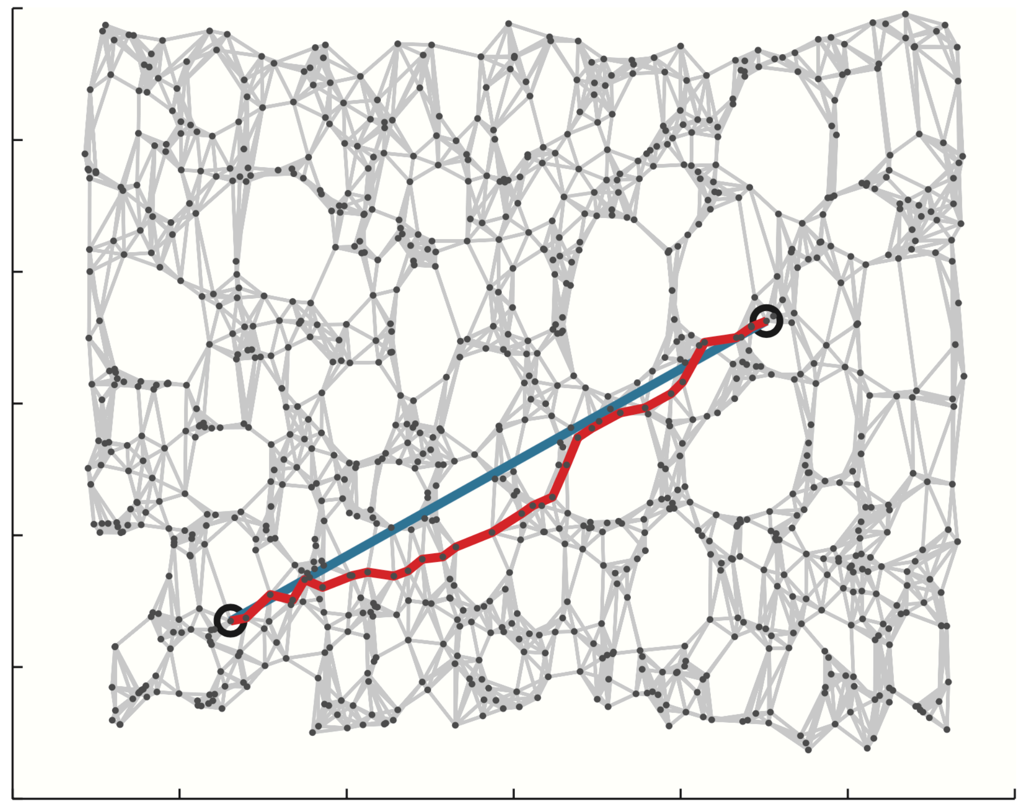
- If $D \in R^{n \times n}$ is not Euclidean
  - $$stress(embedding) = \frac{\sum_{ij} (\|x_i - x_j\| - d_{ij}^2)^2}{\sum_{ij} \|x_i - x_j\|}$$
- Optimize (gradient descent) stress with respect to an embedding

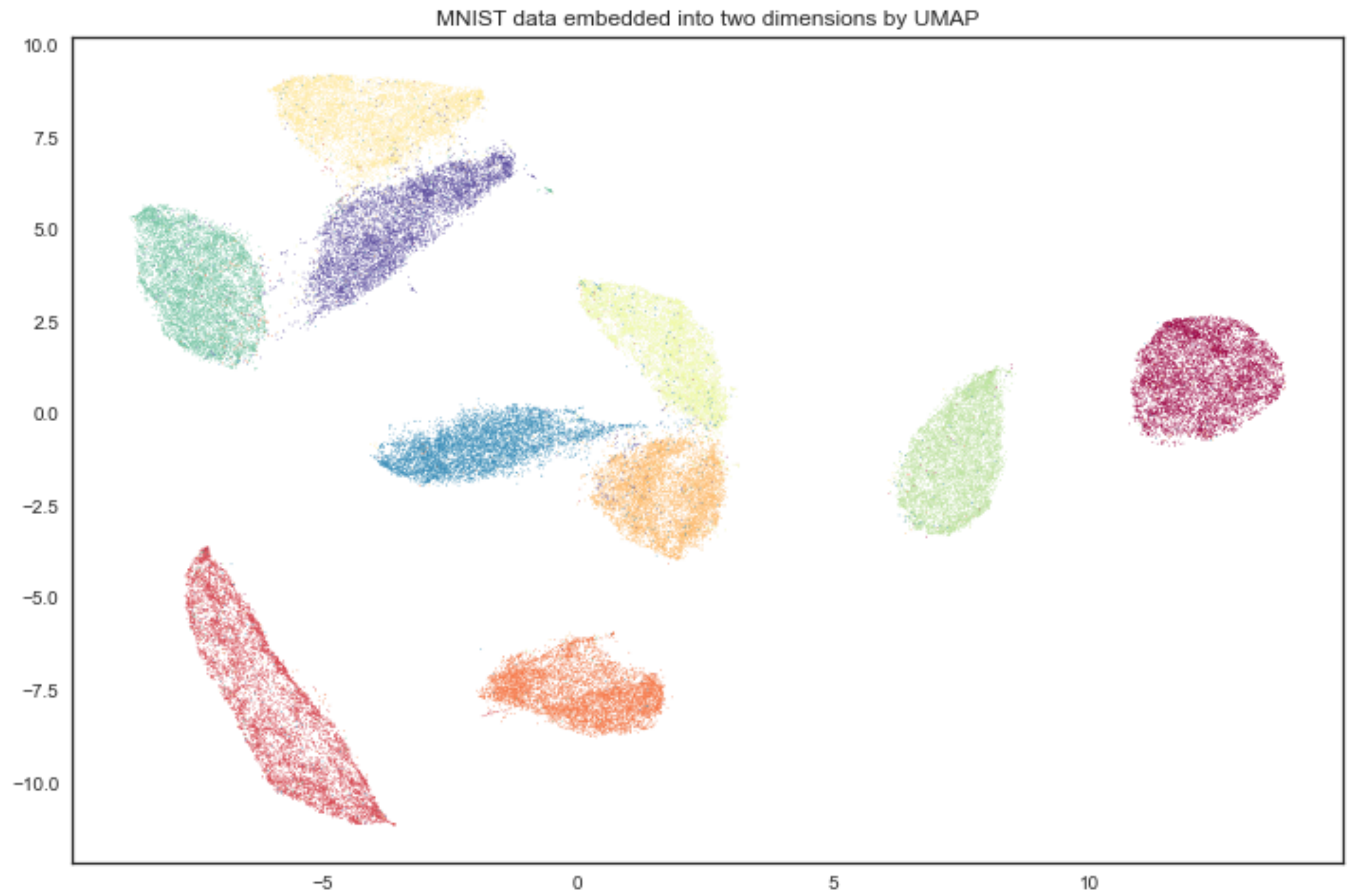
01-04-2024

Non-linear dimensionality reduction (NLDR)

- Try to capture some spatial structure of the data when projected into lower dimension
  - Manifold learning
- Different types of NLDR
  - Isomap
  - Uniform Manifold Approximation and Projection (UMAP)
  - T-distributed stochastic embedding (t-SNE)





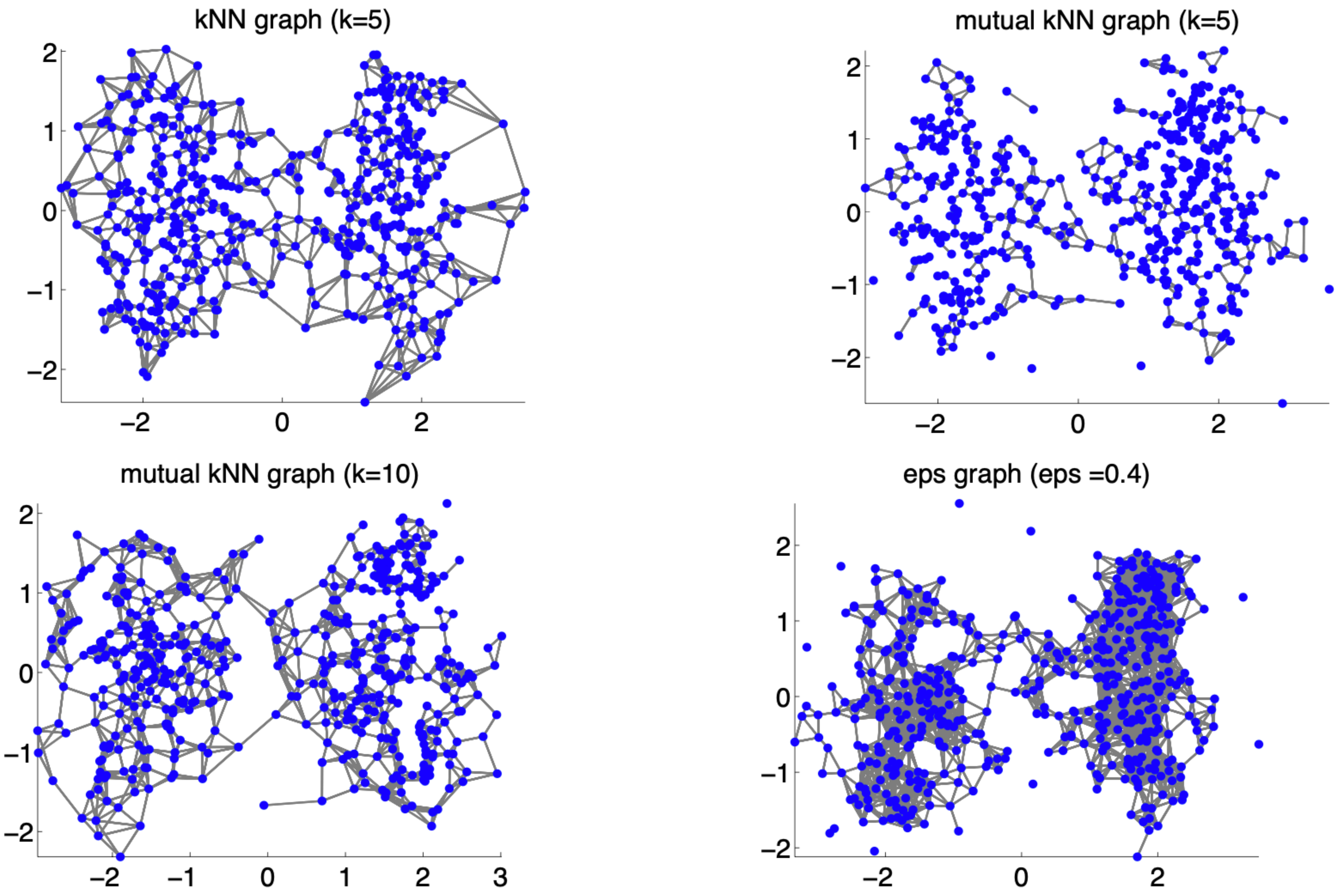


Images: Tenenbaum et al. A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, 2000

Neighbourhood graph (NG)

- Connect data points with edges based on their neighbourhood
- Depending on the neighbourhood different types of NGs:
- Directed k-nearest neighbour graph
  - Connect a data point to its k-nn by directed edges
  - Is it symmetric?
    - No
- k-nearest neighbour graph
  - Connect X_i and X_j if X_j is a k-nn of X_i OR vice versa
- Mutual k-nearest neighbour graph
  - Connect X_i and X_j if X_j is a k-nn of X_i AND vice versa
- ϵ-neighbour graph
  - Connect all the points within the distance of ϵ-neighbourhood

Neighbourhood graph

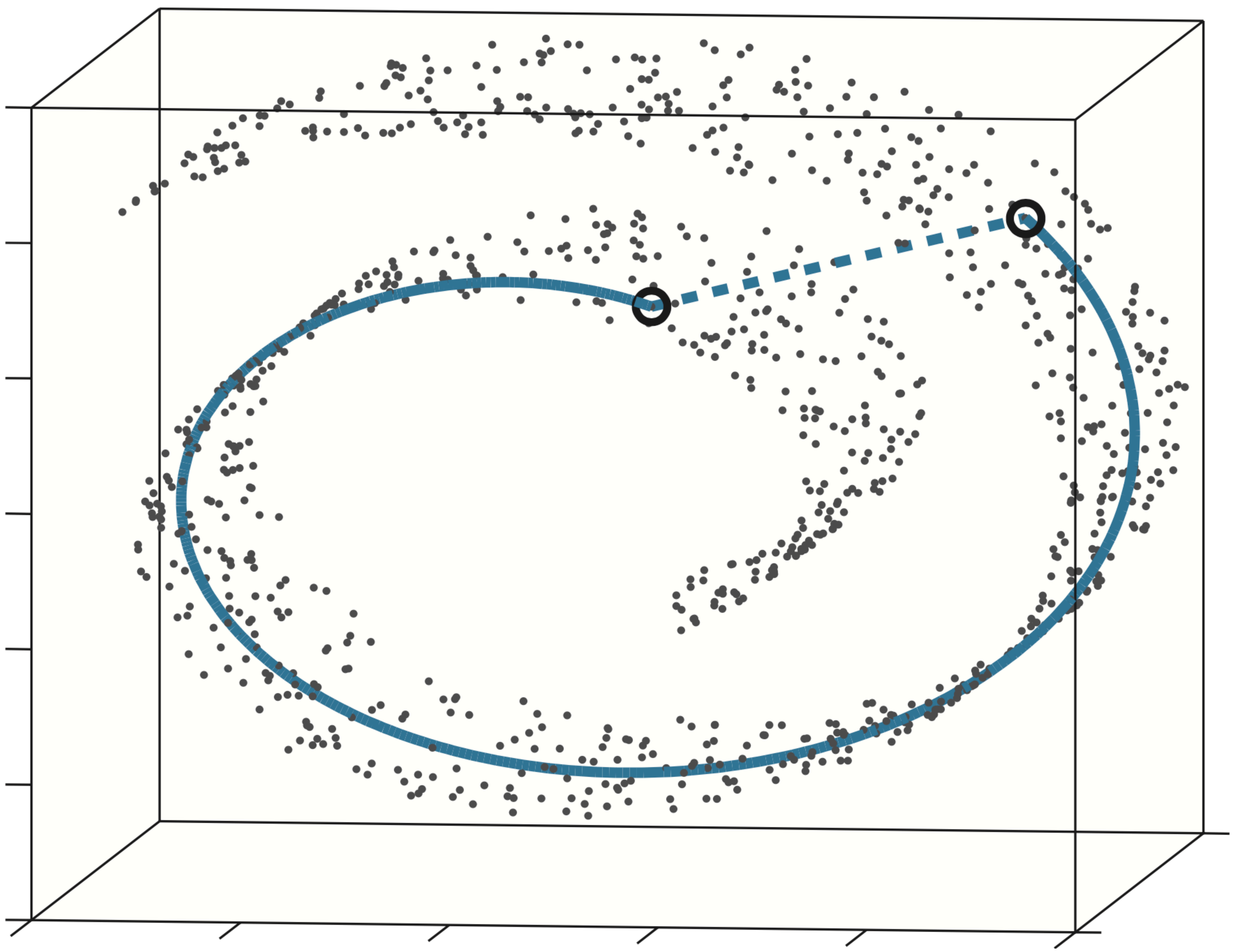


Images: U V Luxburg's slide

Neighbourhood graph (cont…)

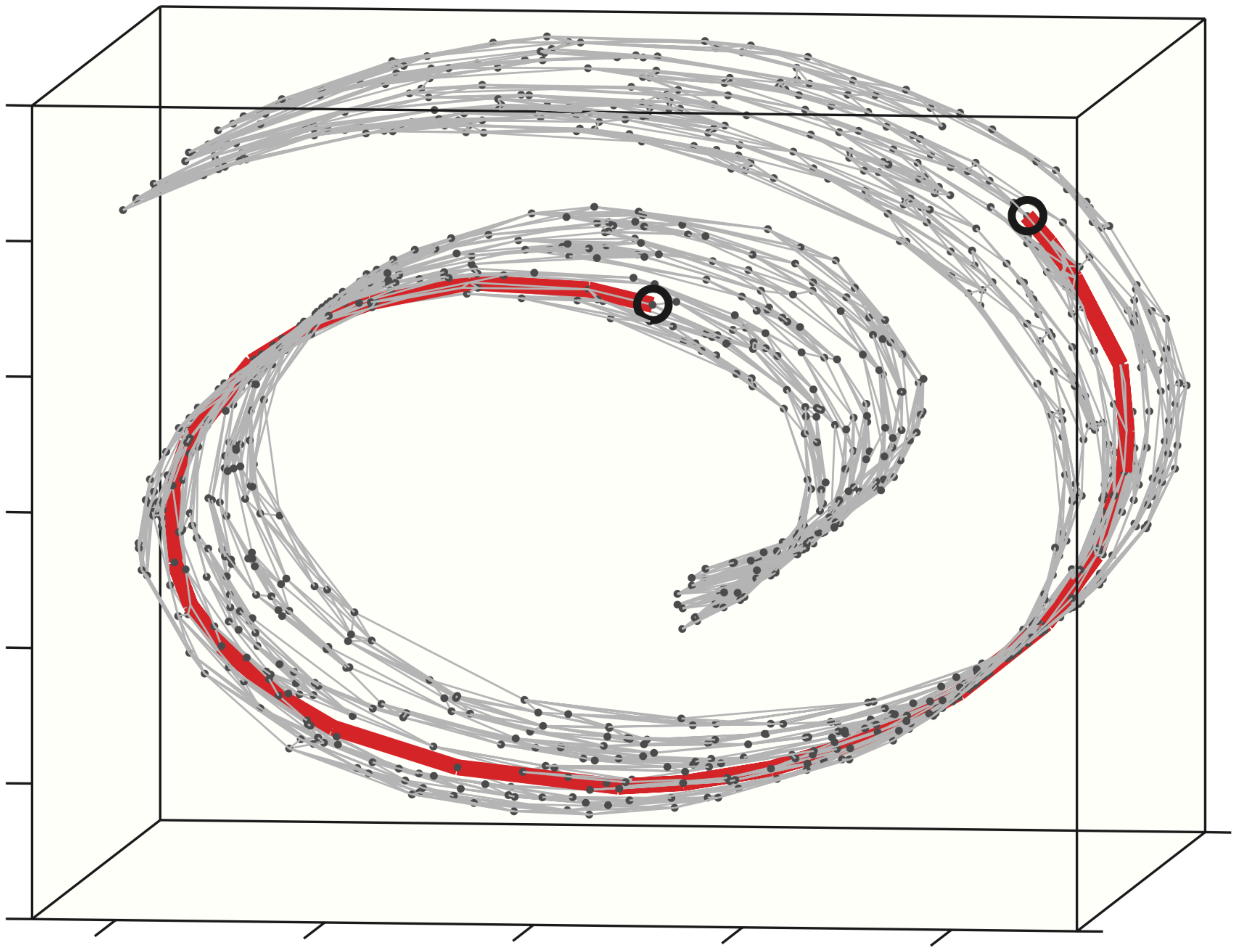
- Why similarity graph ?
  - Graphs is a weel studied structure
  - Capture local structures
- What about edge weights?
  - Similarity between X_i and X_j?
  - In ϵ-neighbour graph, edge weights do not add extra informations

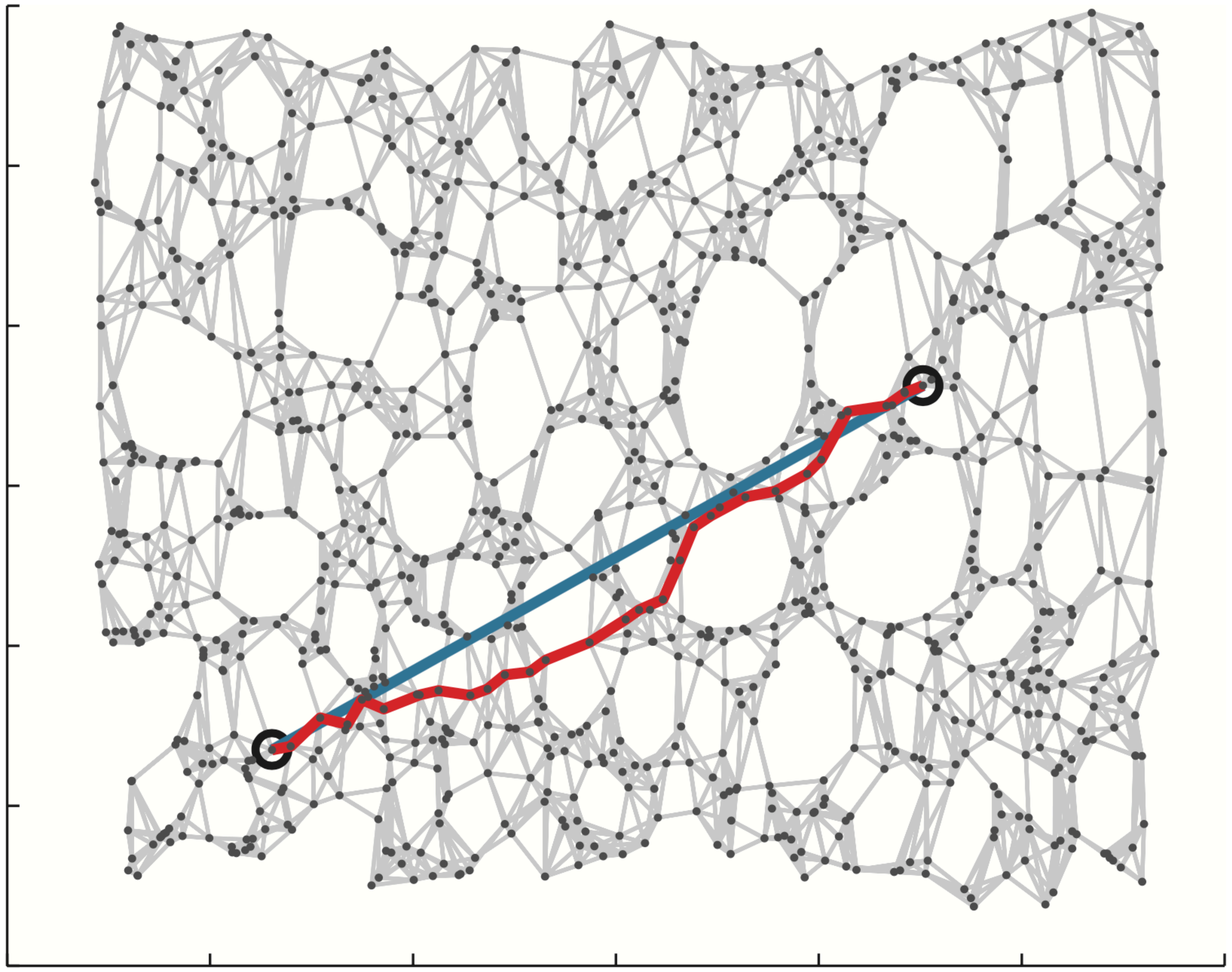
Isomap: geodesic distance



Source: Tenenbaum et al. A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, 2000

Isomap



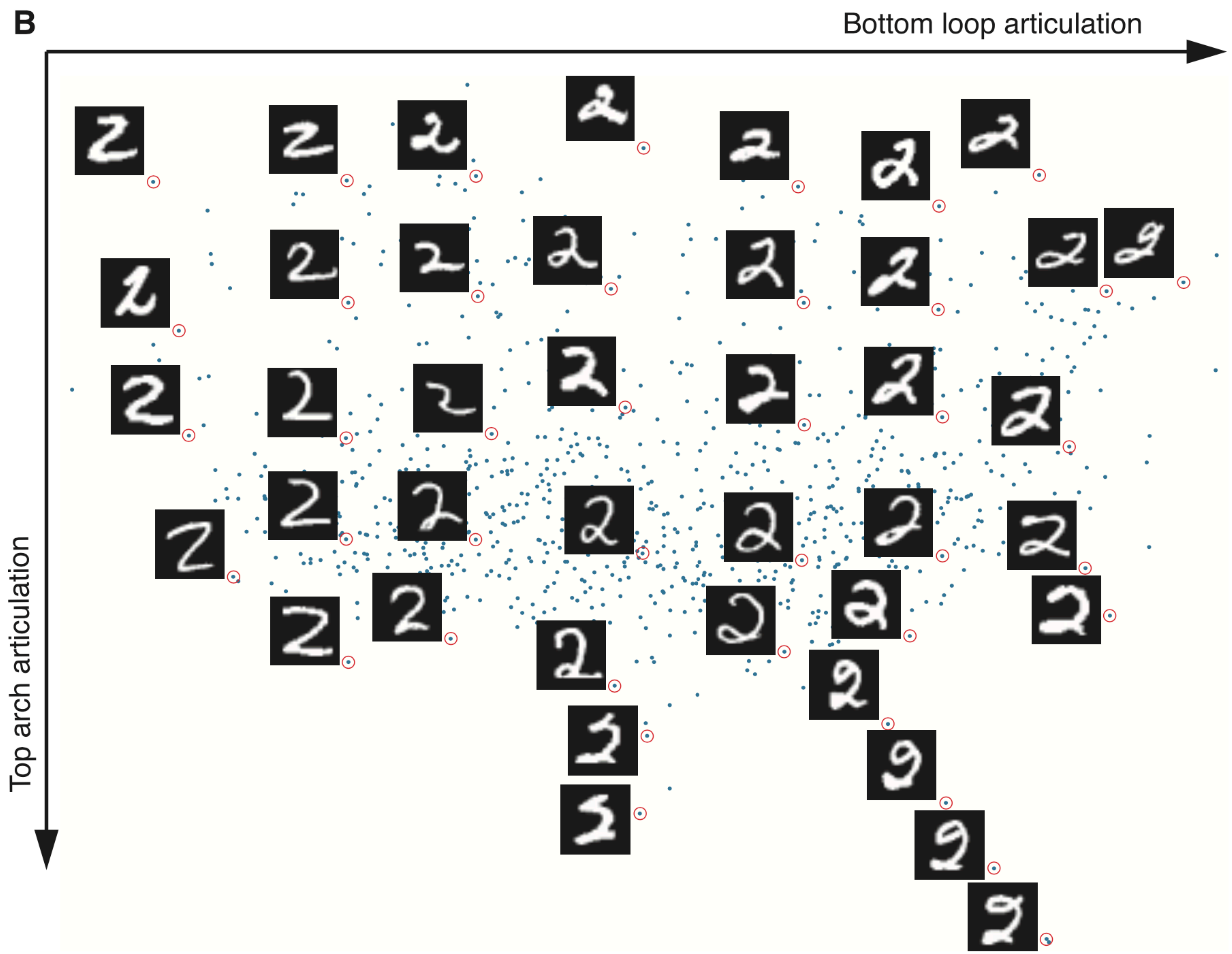


Source: Tenenbaum et al. A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, 2020

Isomap algorithm

- Input:
  - Let $X_1, X_2, \ldots, X_n$ be given data points (observation), $X_i \in R^d$
  - A distance function $f : (X_i, X_j) \rightarrow R$
  - l: the reduced dimension
- Step-1: Construct k-nn graph considering $X_i \in R^d$ as vertices and edge weights by the distance defined by $f : (X_i, X_j) \rightarrow R$
- Step-2: Compute all-pairs shortest path and create a distance matrix $D \in R^{n \times n}$
- Step-3: Apply metic MDS on D

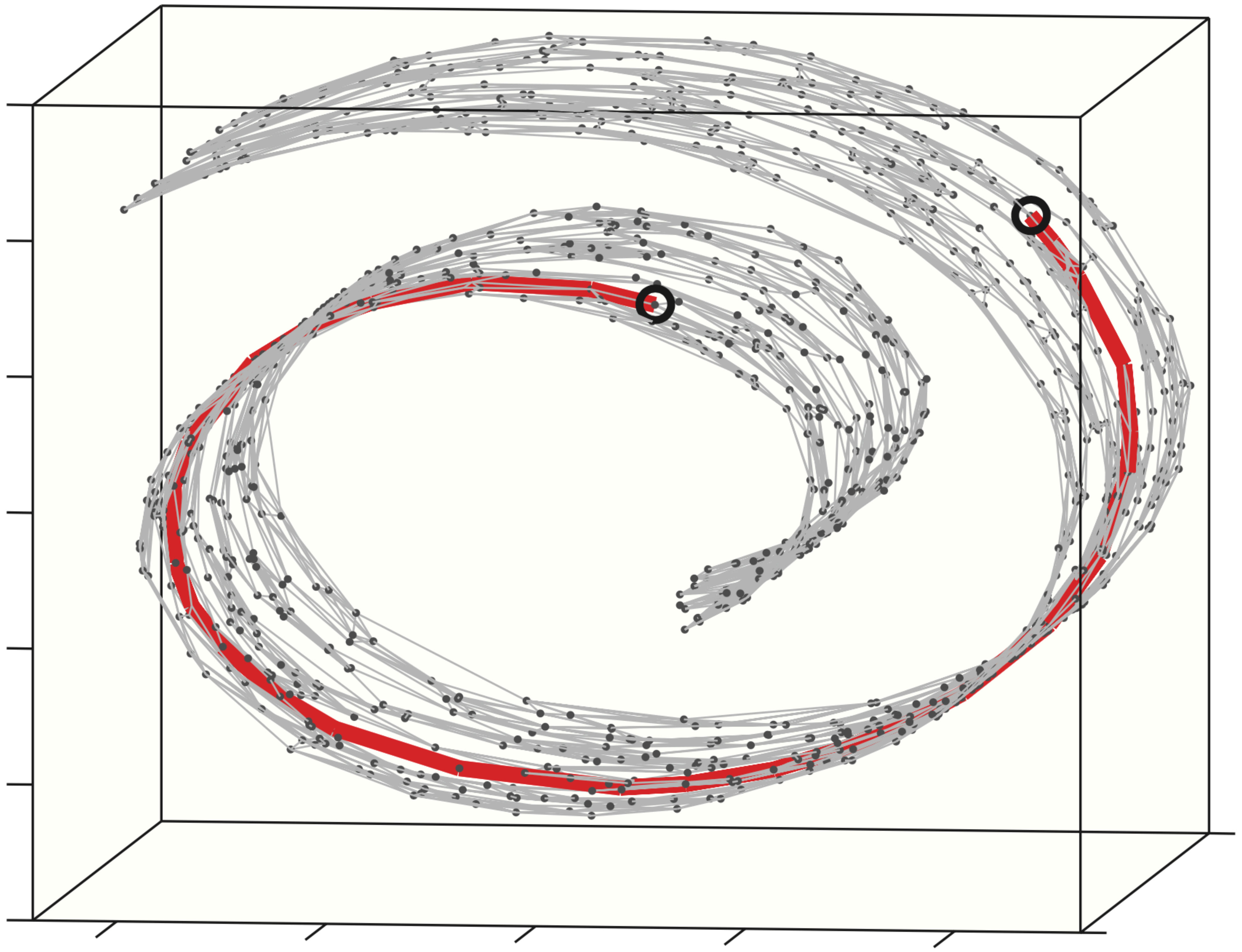
Results

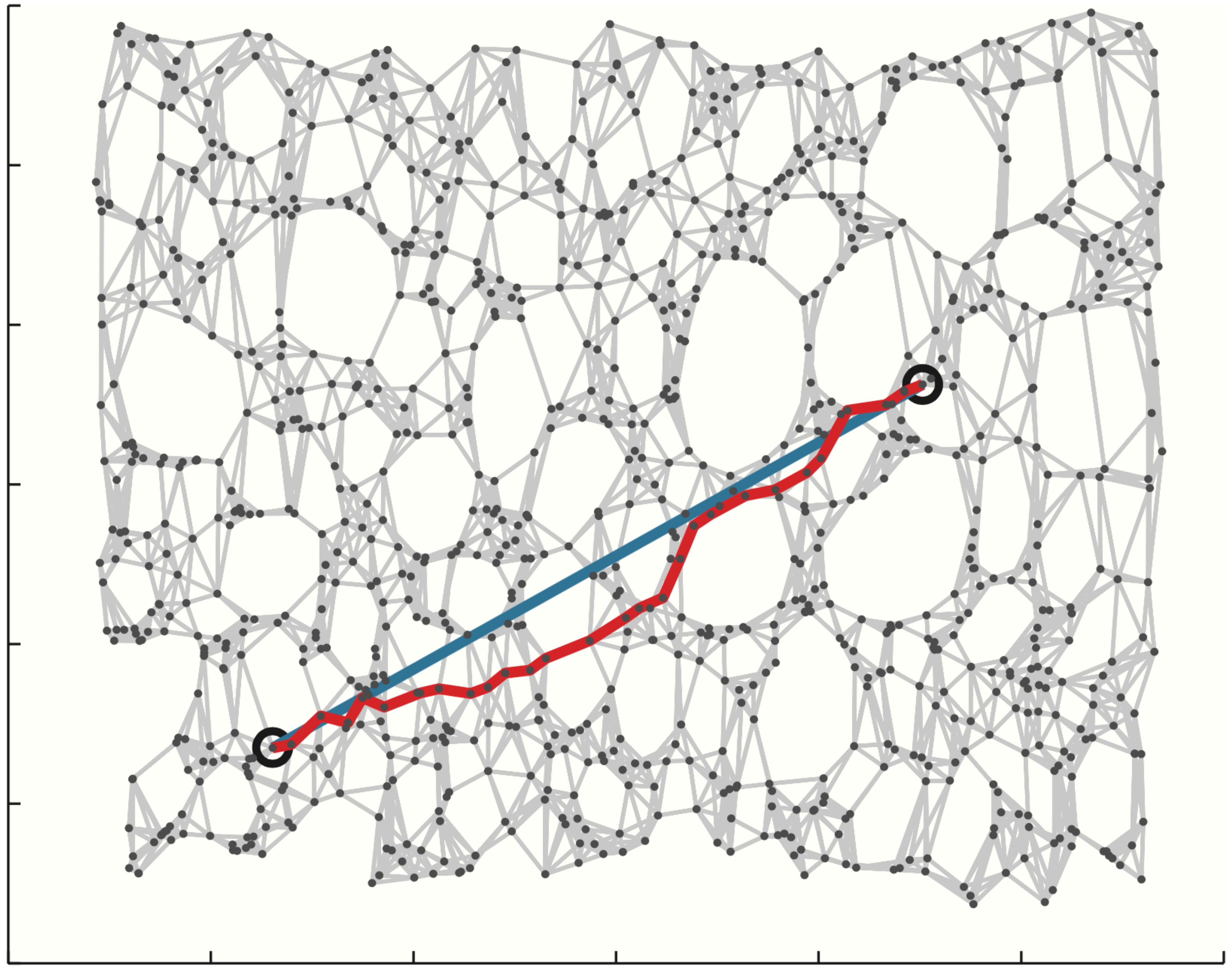


Source: Tenenbaum et al. A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, 2000

Isomap theory

- If $X_i \in R^d$ are uniformly samples from a nice manifold and as $n \to \infty$ and $k \approx \log n$, then the shortest path distance pproximate with the geodesic distance on manifold





Random projection (Johnson-Lindenstrauss)

- Let $X_1, X_2, \ldots, X_n$ be given data points (observation), $X_i \in R^d$
- Find a mapping $P : R^d \rightarrow R^l$ such that $Z_i = P^T X_i;\ i = 1, \cdots, n$, such that
  - $\|x_i - x_j\|_{R^d} = \|z_i - z_j\|_{R^l}$
  - Here $P : R^d \rightarrow R^l$ is a random mapping

Random projection (cont…)

- How can we define a random projection matrix $P \in R^{d \times l}$?
- Step-1: Draw a d-dimenional vector according to the normal distribution and normalize to a unite vector
- Step-2: Draw l vectors using step-1
- Step-3: It is highly likely that you will get l linearly independent vectors
  - If not repeat Step-1 & 2
- Step-4: Orthonormalize the vectors and return that as $P \in R^{d \times l}$

Random projection (cont…)

- Achlioptas -2003
  - Construct a projection matrix $P \in R^{d \times l}$ with independent entries ± 1 with probability $\frac{1}{2}$
- If d is large
  - All column vectors will have length $\sqrt{d}$
  - All column vectors are nearly orthogonal

How can we believe a random mapping?

- Johnson-Lindenstrauss lemma:
  - For any $0 < \epsilon < 1$ and any integer n and let l be a positive integer such that $l \geq \frac{24}{3\epsilon^2 - 2\epsilon^3} \log n$, then for any n points $X_1, X_2, \ldots, X_n$, $X_i \in R^d$ there exists a map $P : R^d \to R^l$ such that for all $i \neq j$
    $(1-\epsilon)\|x_i - x_j\|_{R^d}^2 \leq \|P(x_i) - P(x_j)\|_{R^l}^2 \leq (1+\epsilon)\|x_i - x_j\|_{R^d}^2$

Assignment-5

- Implement multi-dimensional scaling (MDS) with the follows:
  - Consider the MNIST dataset and take 600 random samples from each class
  - Consider the original metric as Euclidean
  - Project the MNIST data in two-dimension using MDS
- Submission deadline: 08-04-2024