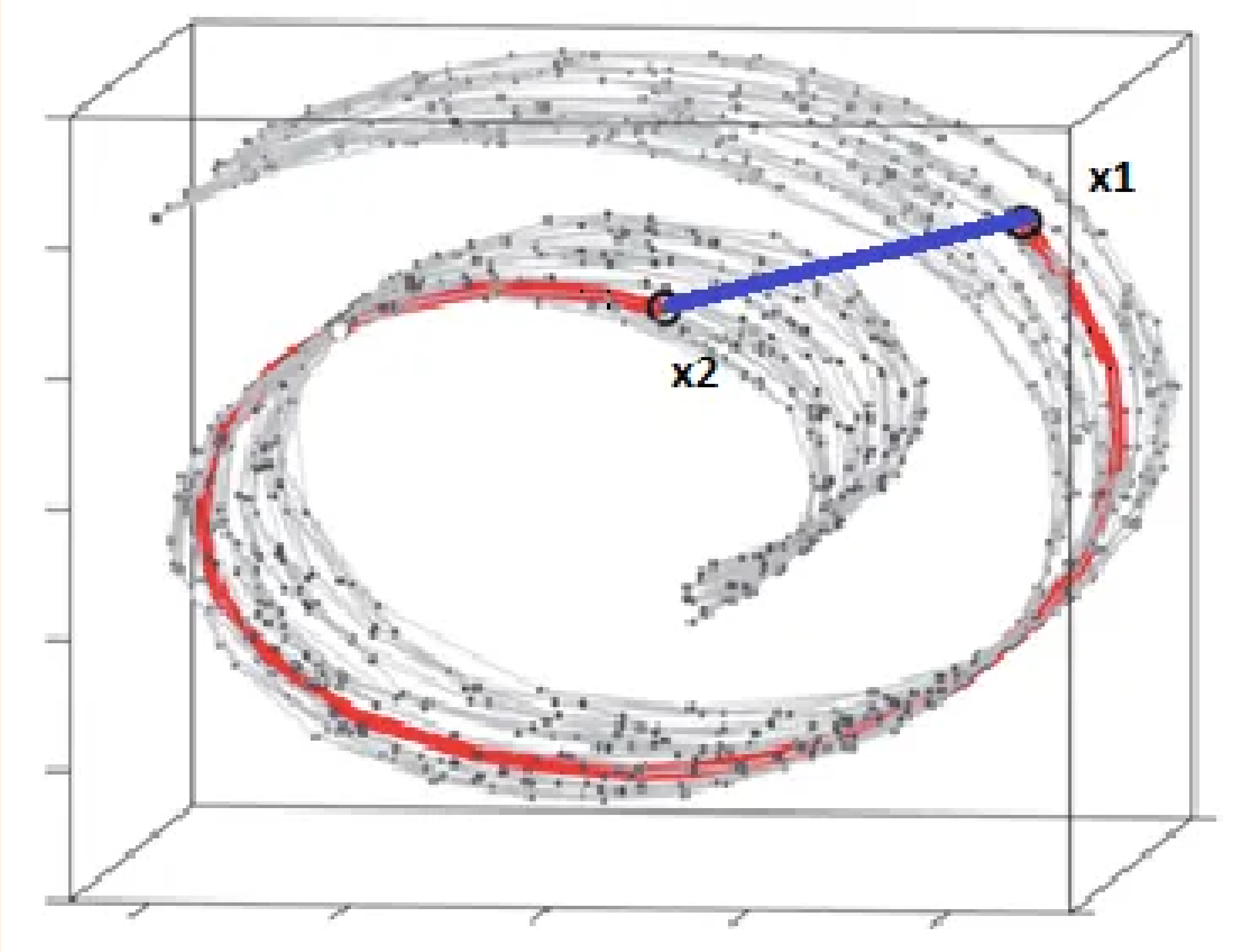
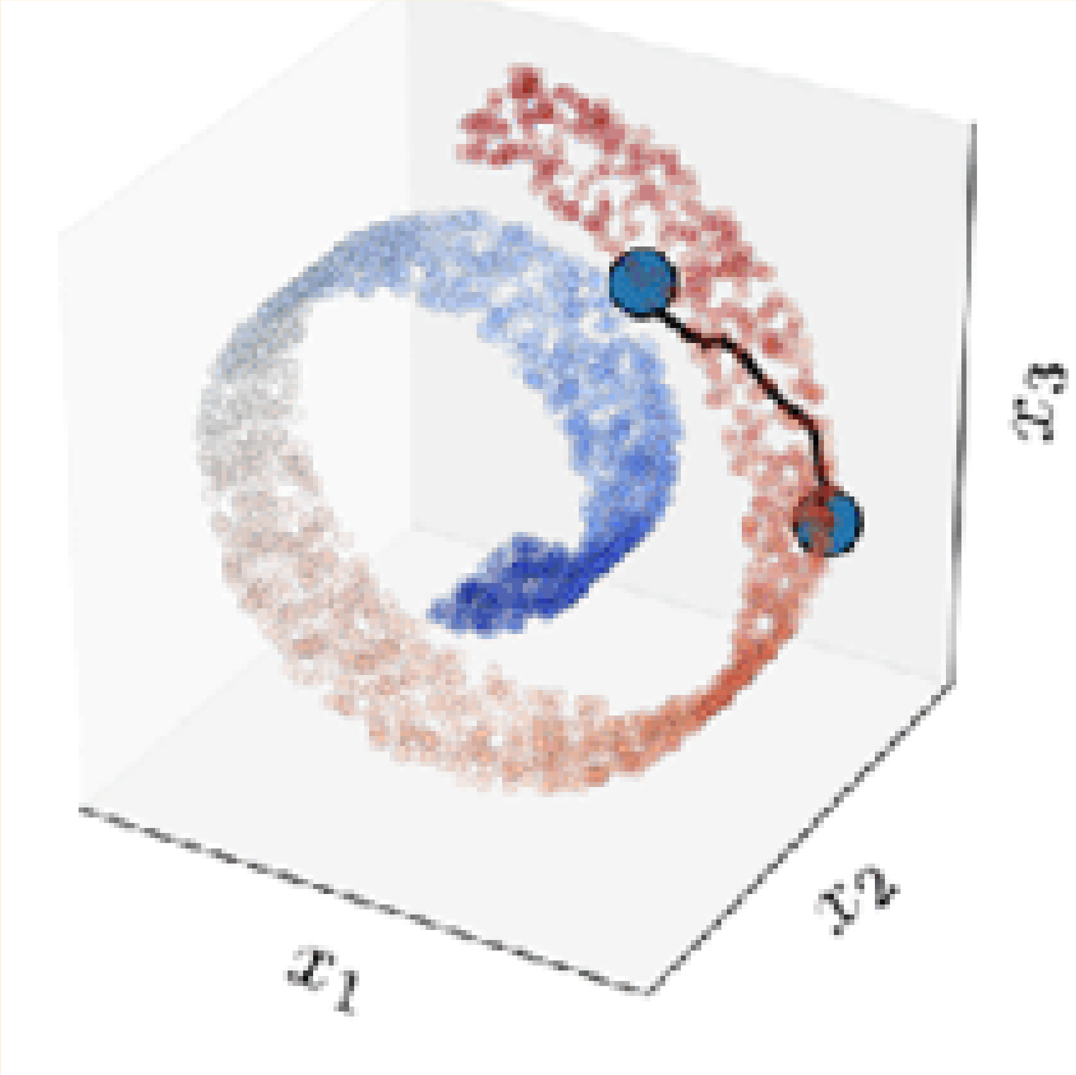
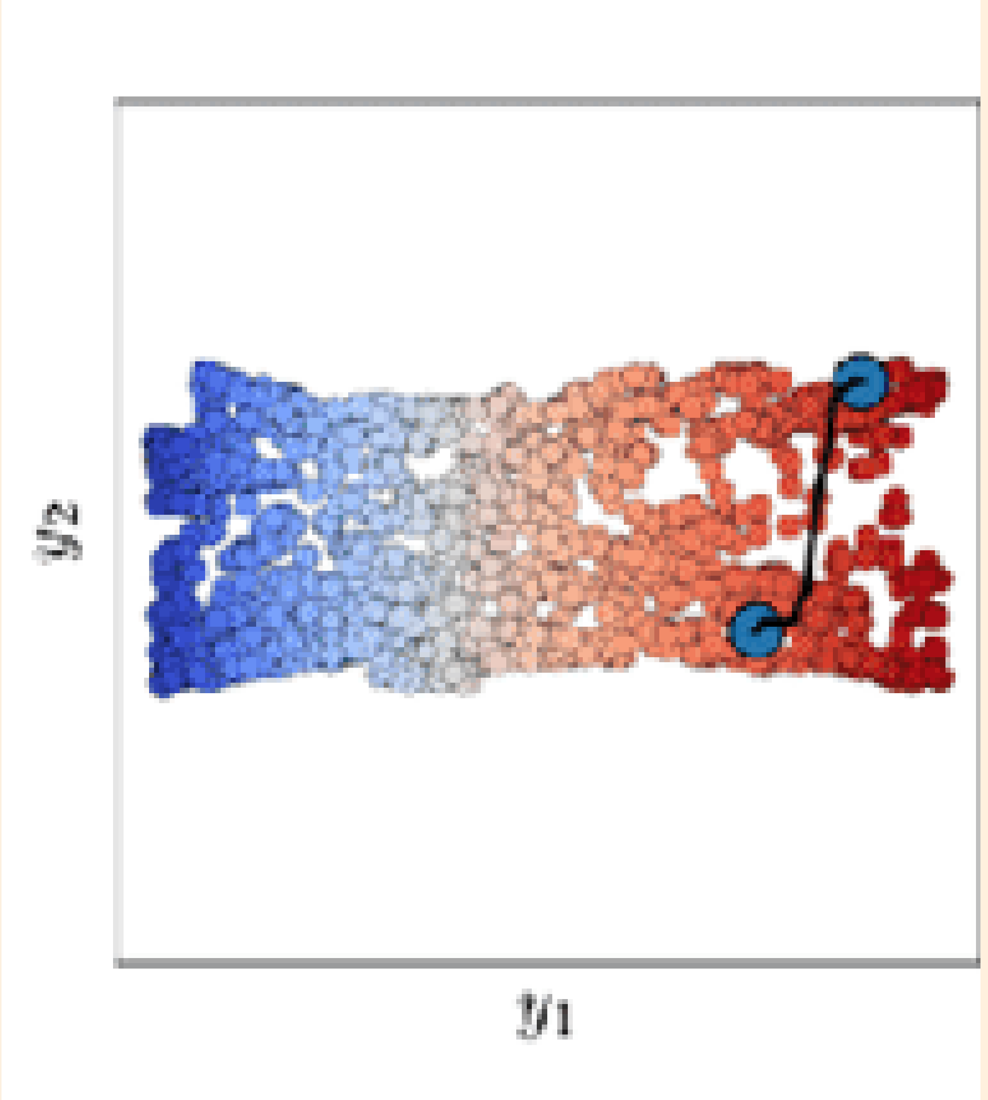
ISOMAP FOR DIMENSIONAL REDUCTION

WHAT IS IT?



WHAT IS IT?

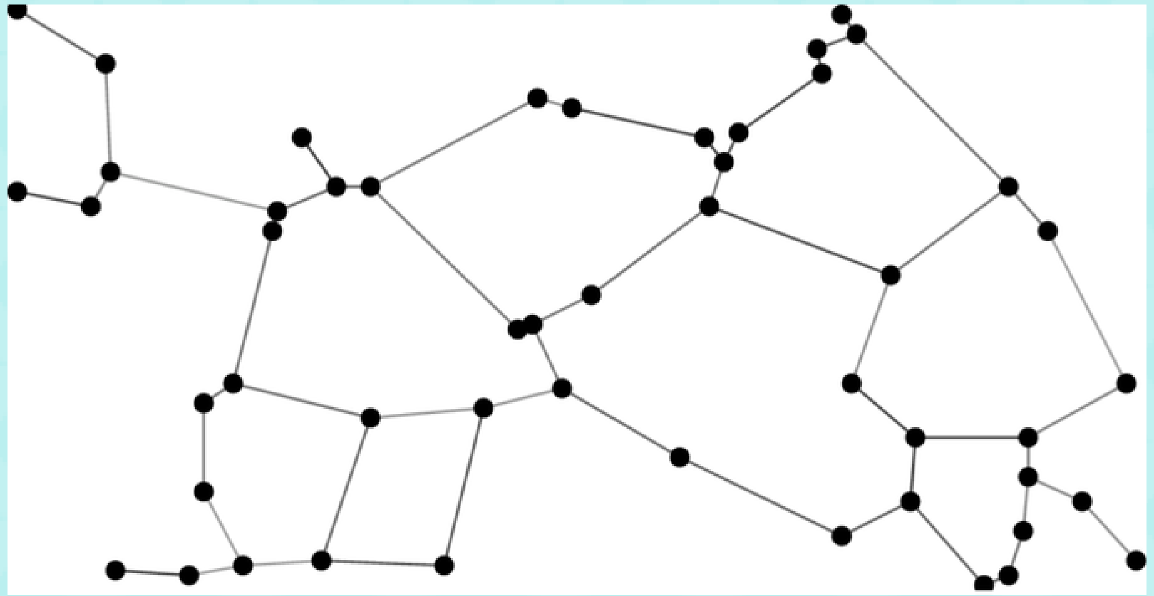




STEPS

NEIGHBORHOOD GRAPH

It represents the local relationships between data points in the high-dimensional space.
Iit helps capture the intrinsic geometry of the data. Isomap then uses this graph to find low-dimensional representations of the data points that best preserve the pairwise distances along the graph.



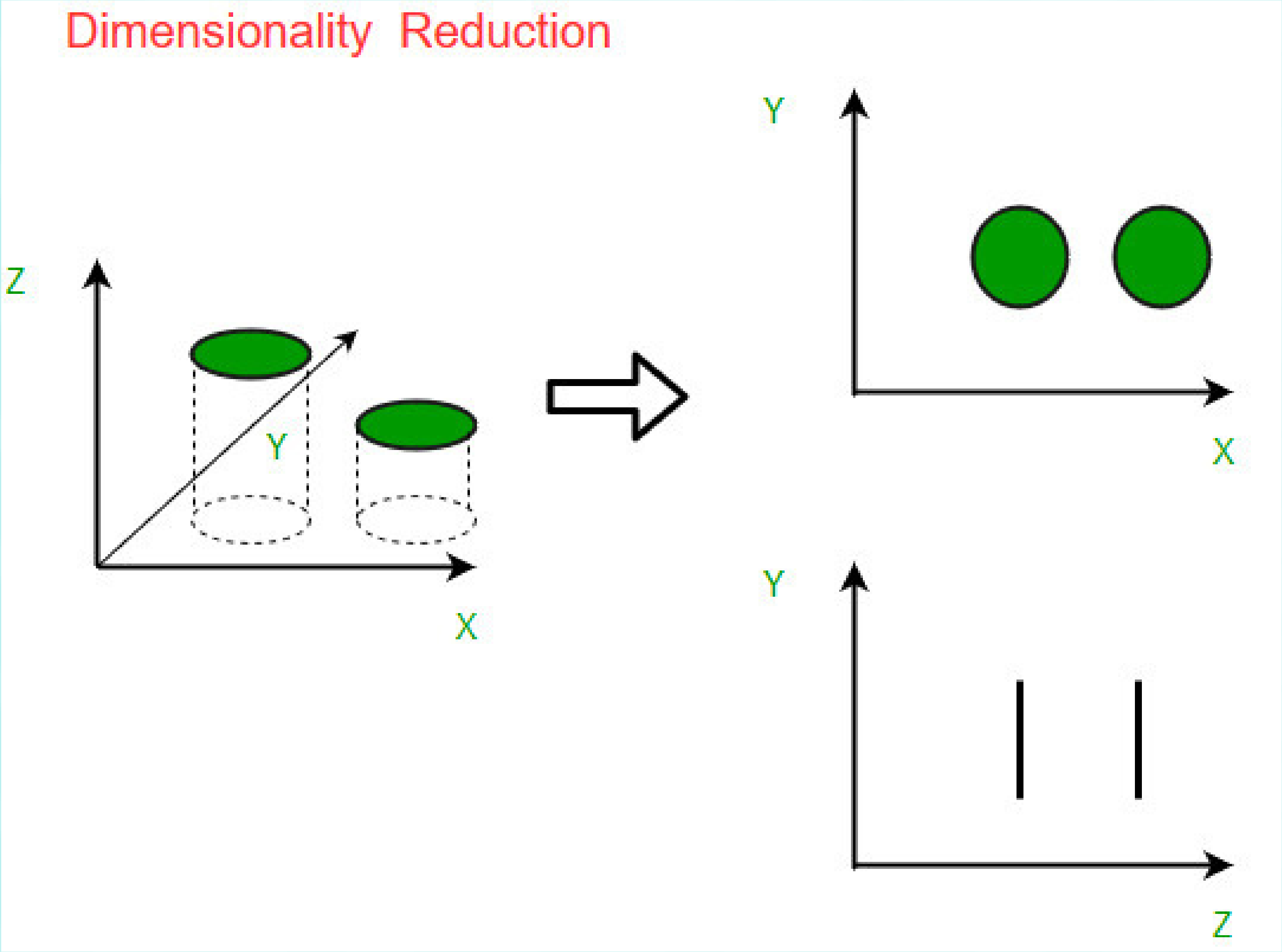
DISSIMILARITY MATRIX

The geodesic distances between points need to be calculated. While creating the neighborhood network, the resulting graph has to be a single connected component.

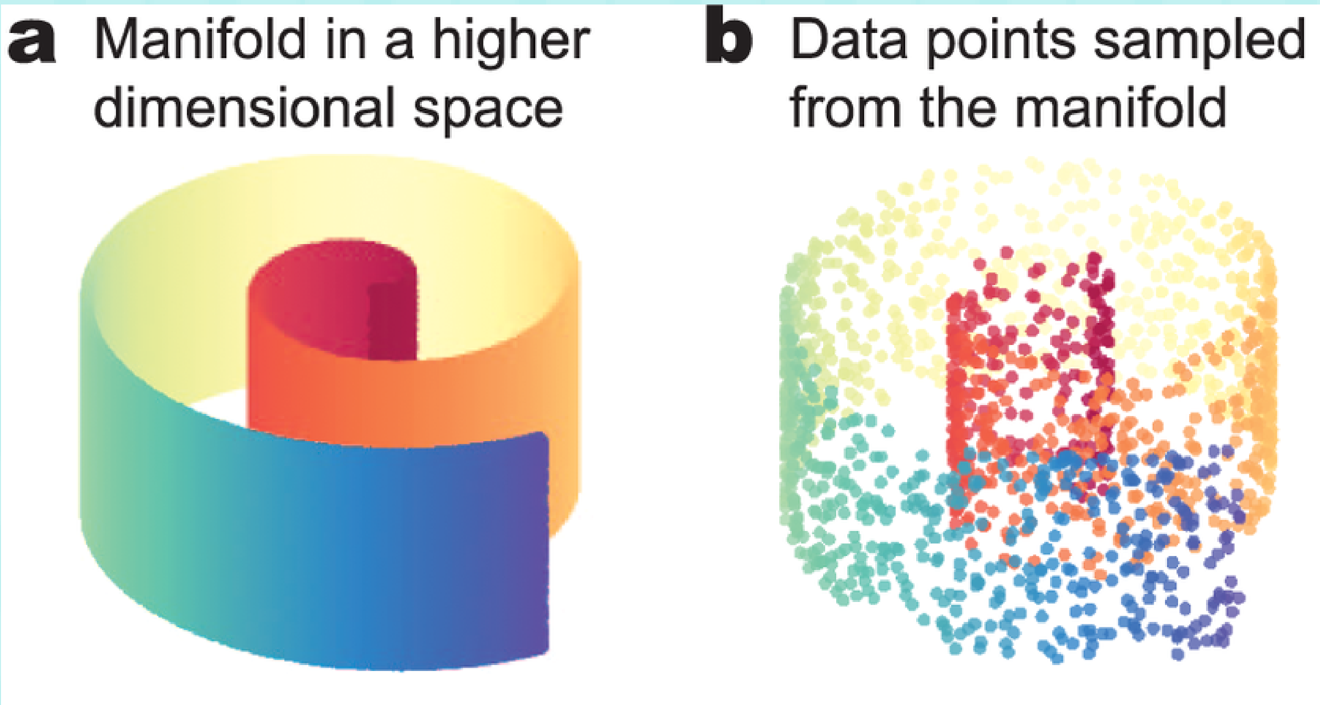
STEPS OF ISOMAP

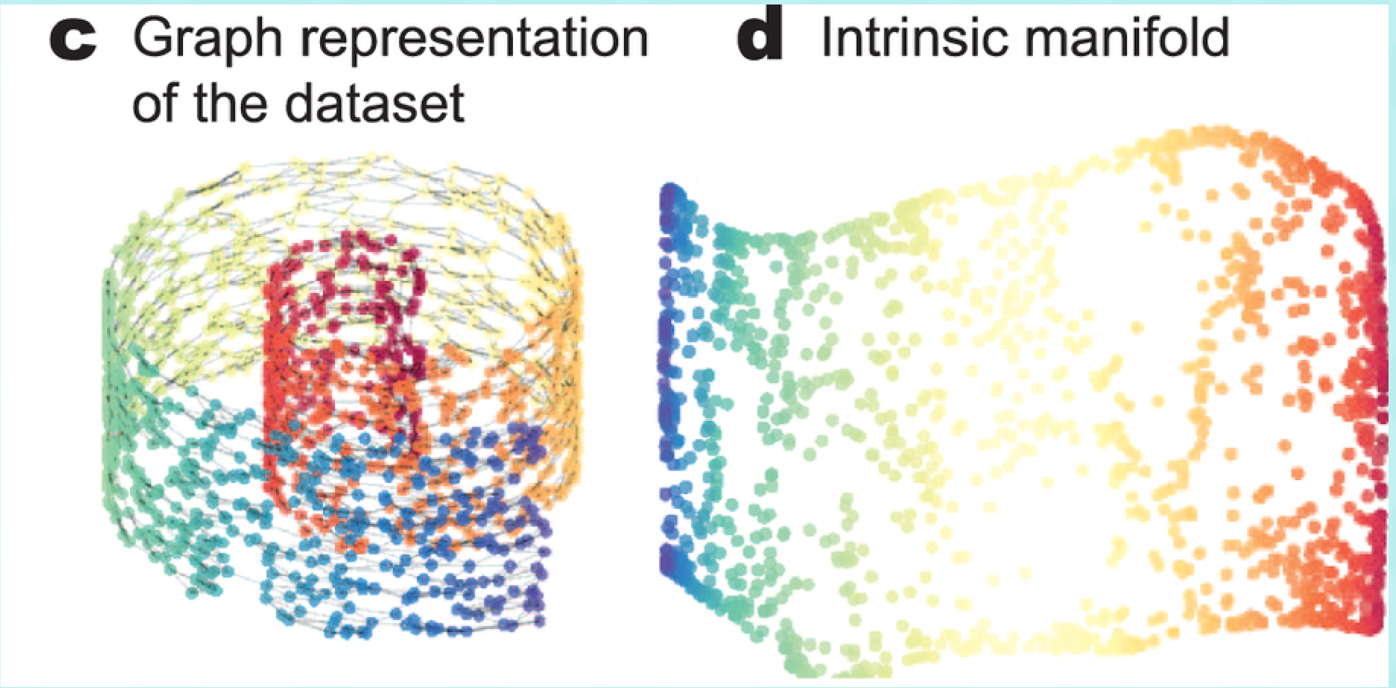
CONSTRUCTING THE LOW-DIMENSIONAL EMBEDDING

It uses dimensionality reduction techniques, such as classical multidimensional scaling (MDS), to find an embedding in a lower-dimensional space
Lower Dimensional Results: The result of this process is a representation in a lower dimensional space (often two-dimensional or three-dimensional) of the original data.



STEPS OF ISOMAP

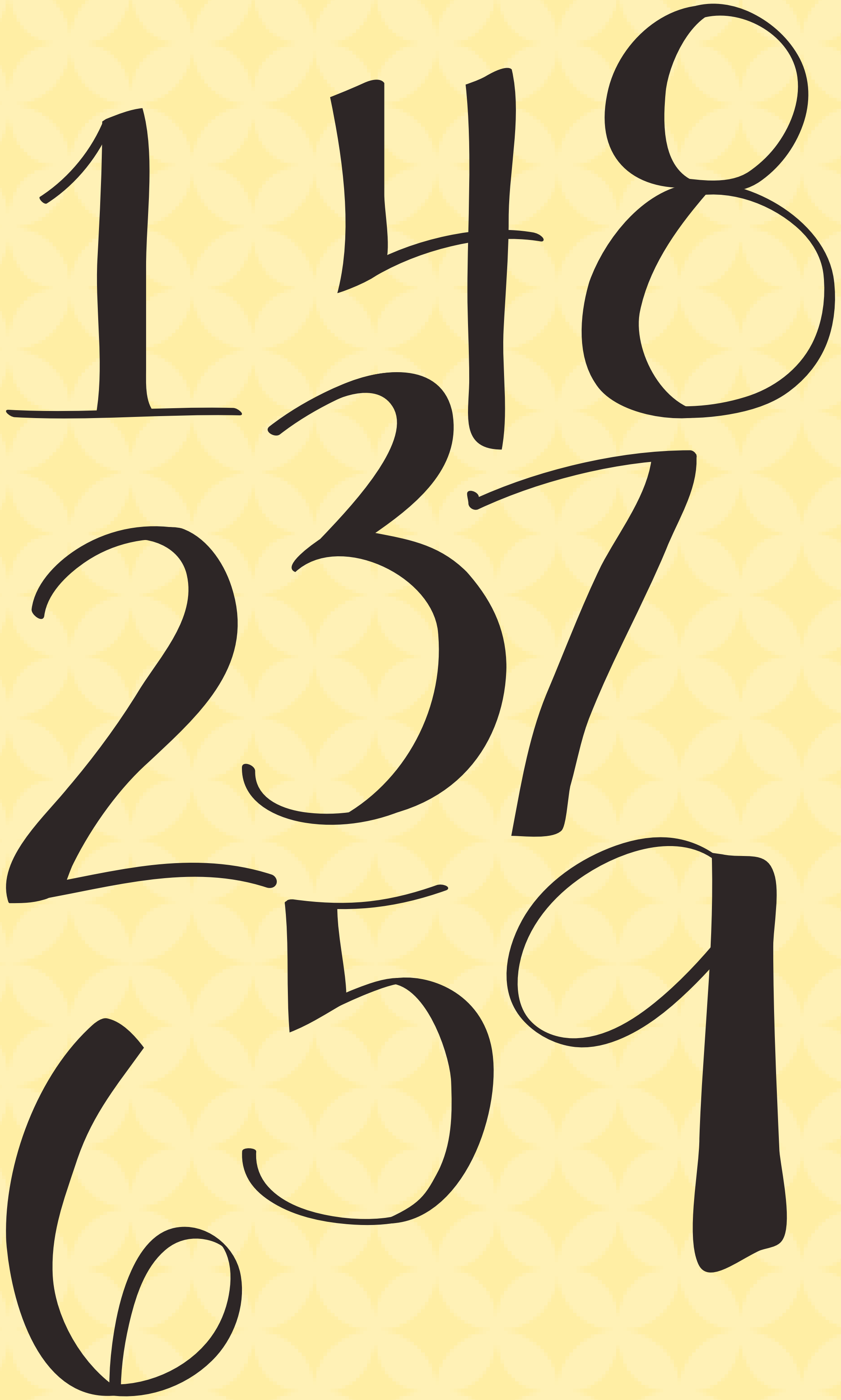




MAPPING NEW DATA POINTS

Once the lower-dimensional embedding is computed, ISOMAP can map new, unseen data points to this lower-dimensional space by finding their approximate locations based on the pairwise distances obtained.

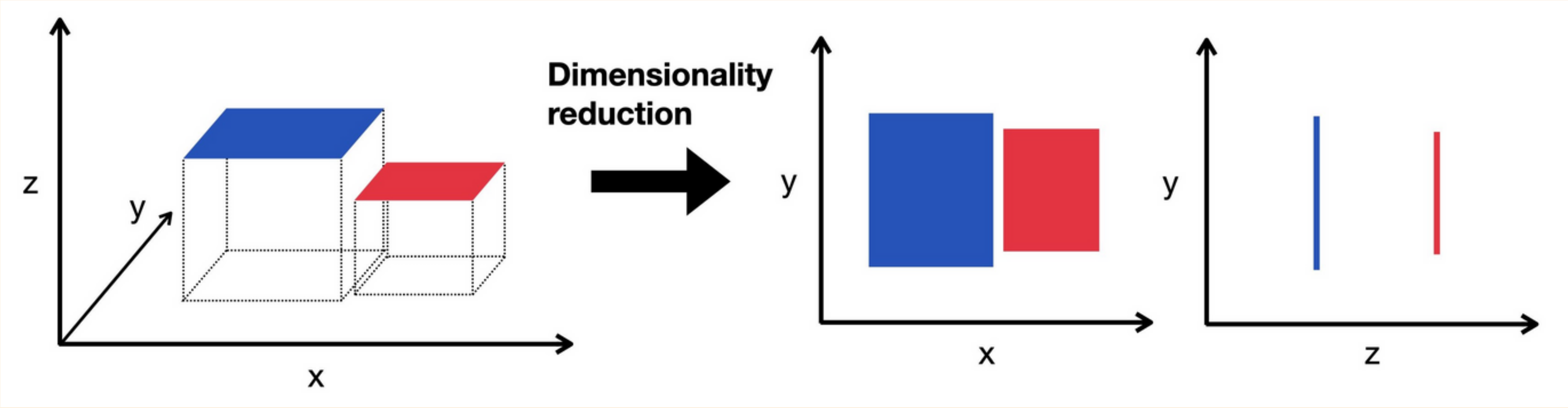
EXAMPLE



1. *Data Preparation: Dataset with thousands of these images, each represented by a vector of 28x28 = 784 pixel values. This makes your dataset high-dimensional.*
2. *Isomap Dimensionality Reduction:*
   a. *Neighbor Finding: Find the nearest neighbors for each image based on some similarity measure, like the pixel values' differences.*
   b. *Geodesic Distance Calculation: Calculate the geodesic distances between all pairs of images by considering the connections in the neighbor graph. It figures out the shortest path through these connections to measure how similar or dissimilar each pair of images is.*

EXAMPLE

3. Lower-Dimensional Representation: creates a lower-dimensional representation of your dataset, typically in two or three dimensions. This new representation keeps the most important relationships between the images while reducing the number of features dramatically.



LIMITATIONS

It is worth mentioning that the algorithm has been criticized for its ability to estimate surface connectivity accurately.

- It assumes that the data lies on a single, connected manifold. It may not perform well when the data consists of multiple disconnected manifolds.
- Computationally Intensive: Calculating pairwise distances and geodesic distances for large datasets can be computationally expensive.
- Sensitivity to Hyperparameters: The choice of the number of neighbors and distance metric can impact the results, parameter tuning is important.

REFERENCES

[1] Mlearnere, W. by. (2021, mayo 24). What is isomap? Towards Data Science. https://towardsdatascience.com/what-is-isomap-6e4c1d706b54

[2] Mlearnere. (2022, 6 enero). What is ISOMAP? - towards data science. Medium. https://towardsdatascience.com/what-is-isomap-6e4c1d706b54

[3] Tseng, J. C.-H. (2022). An ISOMAP analysis of sea surface temperature for the classification and detection of El Niño & La Niña events. Atmosphere, 13(6), 919. https://doi.org/10.3390/atmos13060919