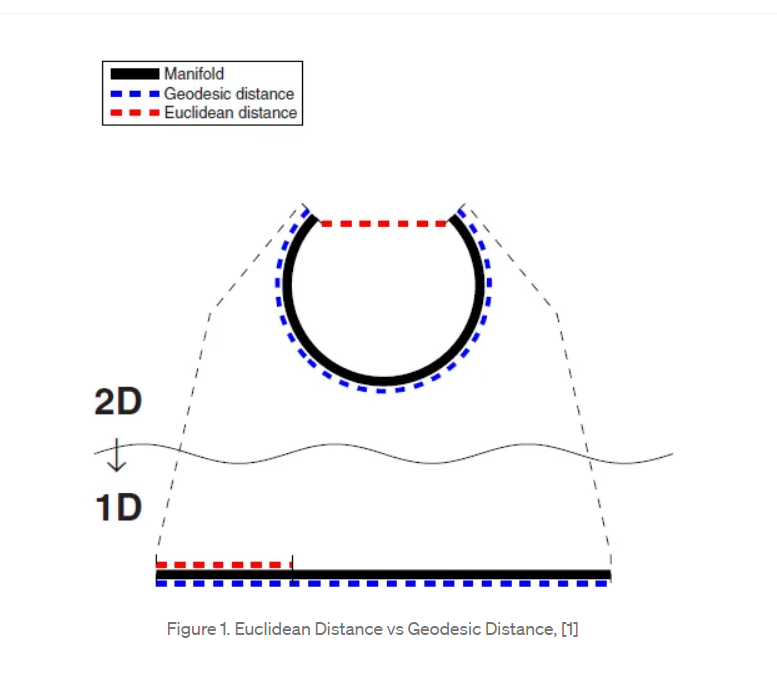
**ISOMAP.**

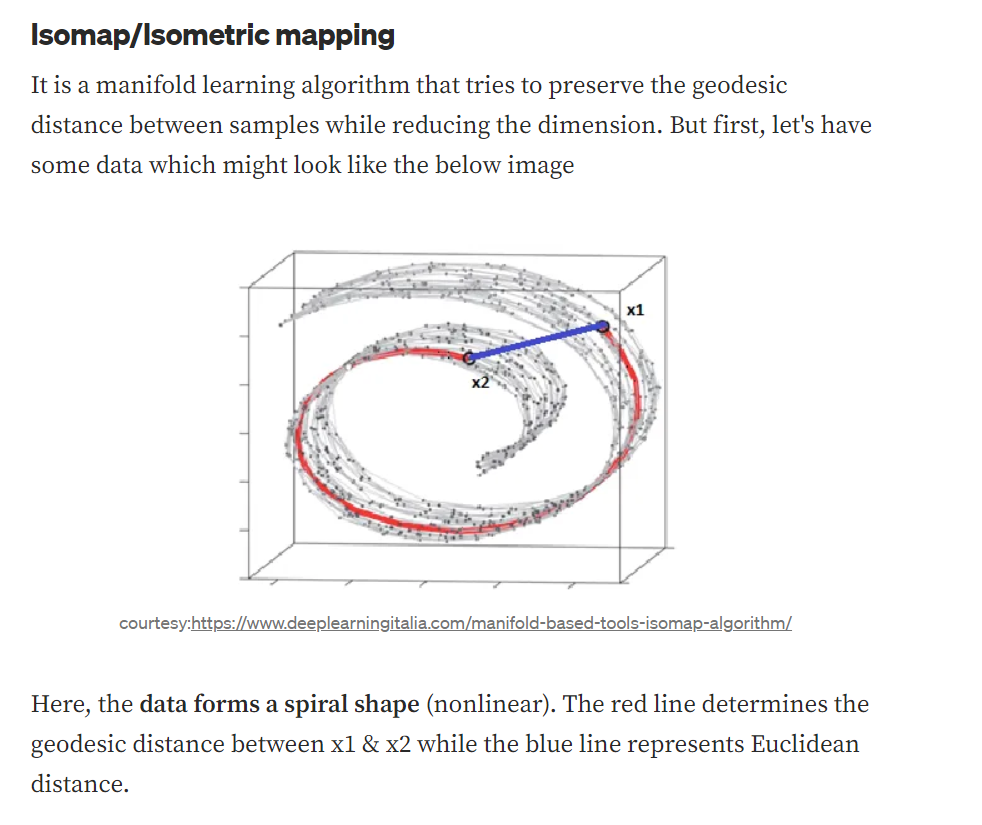
# Isomap Embedding — An Awesome Approach to Non-linear Dimensionality Reduction. Preserving Geodesic Distance for Non-Linear Datasets.

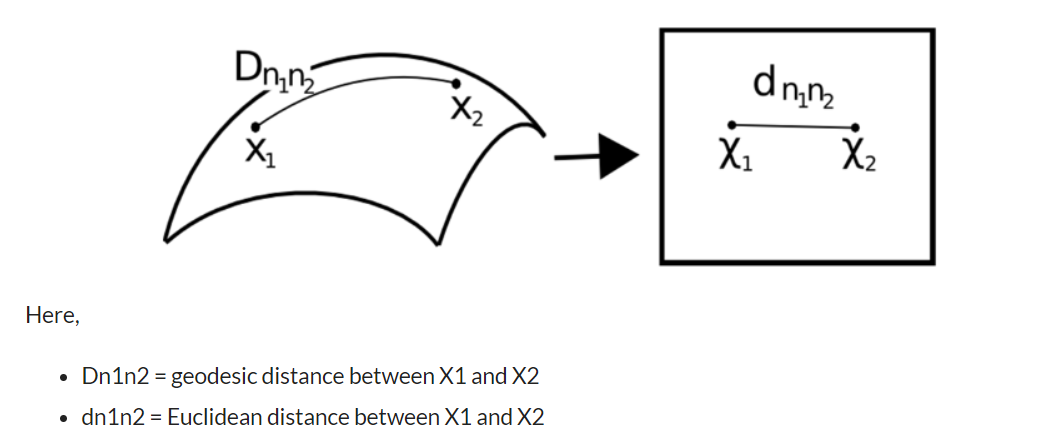
# Before we look at the example of Isomap and compare it to a linear method of Principal Components Analysis (PCA), let’s list the high-level steps that Isomap performs:

1. Use a KNN approach to **find the k nearest neighbors** of every data point. Here, “k” is an arbitrary number of neighbors that you can specify within model hyperparameters.
2. Once the neighbors are found, **construct the neighborhood graph** where points are connected to each other if they are each other’s neighbors. Data points that are not neighbors remain unconnected.
3. **Compute the shortest path** between each pair of data points (nodes). Typically, it is either Floyd-Warshall or Dijkstra’s algorithm that is used for this task. Note, this step is also commonly described as finding a **geodesic distance** between points. Not linear distance such as PCA.
4. **Use**[**multidimensional scaling (MDS)**](https://towardsdatascience.com/mds-multidimensional-scaling-smart-way-to-reduce-dimensionality-in-python-7c126984e60b)**to compute lower-dimensional embedding**. Given distances between each pair of points are known, MDS places each object into the N-dimensional space (N is specified as a hyperparameter) such that the between-point distances are preserved as well as possible.

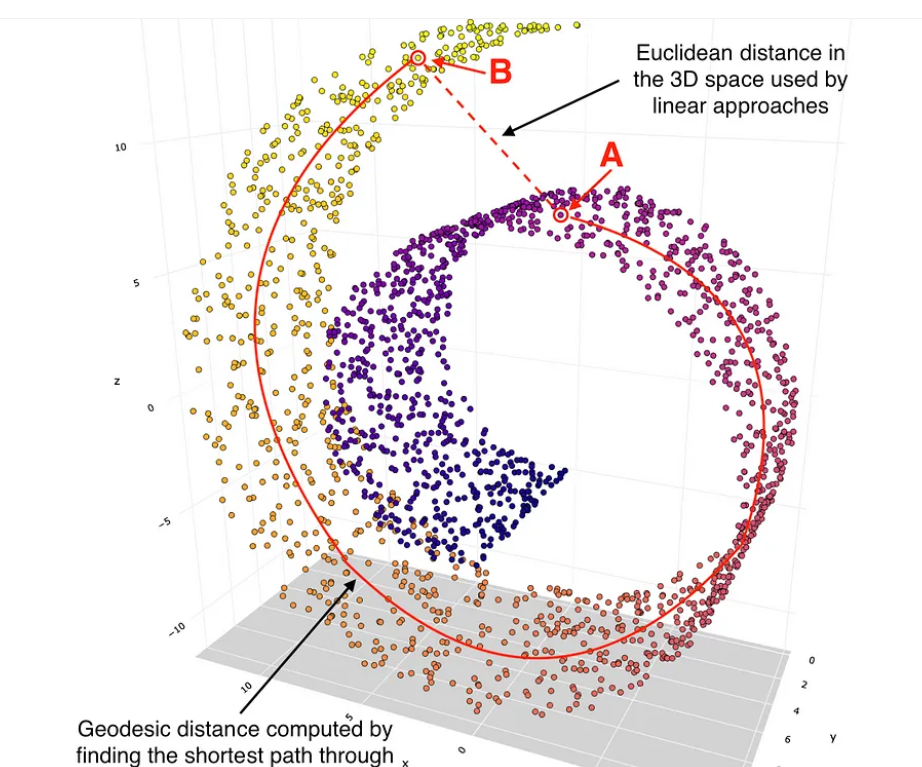
Isomap is a non-linear dimensionality reduction method, which is a different version of metric MDS, and reduces the dimensionality while preserving **geodesic distance**. The most obvious difference between Metric MDS and ISOMAP: Euclidean distance is preserved in Metric MDS, while geodesic distance is preserved in ISOMAP.







**the geodesic distance considers the data adjacent to these data**, in the Euclidean distance it is only calculated the shortest linear path, without taking into consideration the shape of the dataset. Of course, since our goal is to reduce the dimensionality of the dataset with the least loss, more effective results can be obtained with geodesic distance according to the dataset.



# PCA is not a good option for non-linear datasets.

