1 Non-linear Dimensionality Reduction

High-dimensional data, meaning data that requires more than two or three dimensions to represent, can be difficult to interpret. One approach to simplification is to map it to a space of low enough dimension, the data can be visualized in the low-dimensional space. PCA (principal component analysis) is one way of doing this which was discussed previously. However such methods involve multiplication by a matrix which is a linear function. Most data in real life require more sophisticated non-linear data-driven approaches for giving a proper visualization.

2 IsoMap

IsoMap is a nonlinear dimensionality reduction method which tries to map points on a high-dimensional non-linear manifold to a lower dimensional set of coordinates.

For example, when we display the structure on the left below with PCA, all the color dots are meshed together even though the 3D image shows a clear spectrum of color on a S curve shape. IsoMap is a MDS method that use geodesic to measure distance so it can capture manifold structure. On the right, it is the 2D projection of the 3D S-shape manifold. In the 2D projection, we can see the color transition in the original S shape curve.

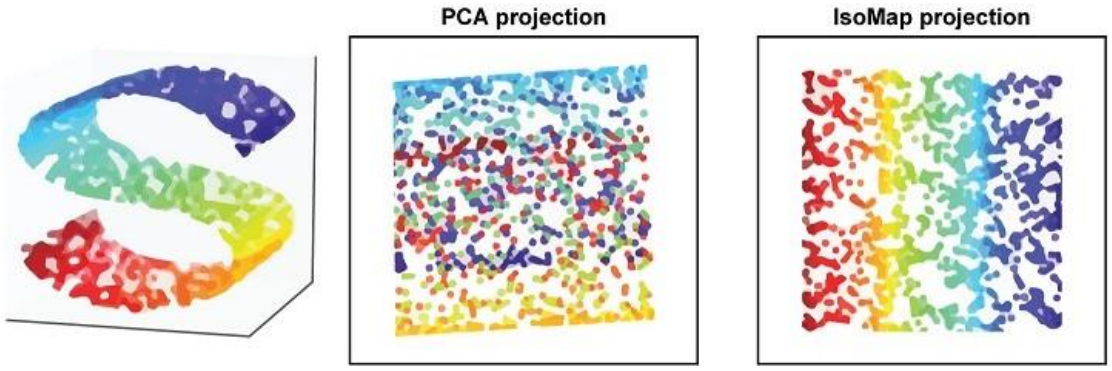

Figure 1

IsoMap uses geodesic rather than Euclidean space to measure distance

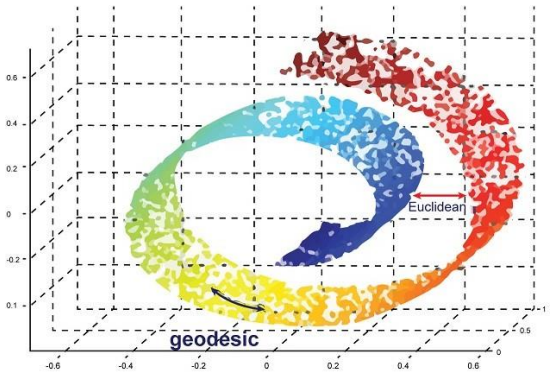

Figure 2

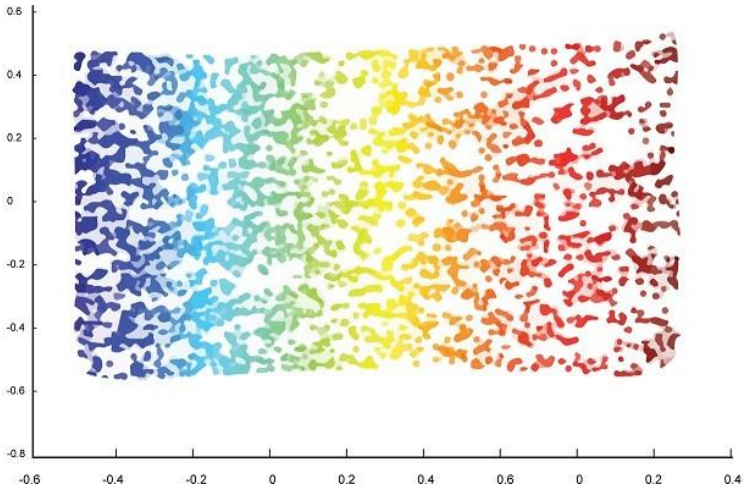

Figure 3

To visualize the “swiss-roll” manifold in 2D, we measure the geodesic distance on the manifold. We represent datapoints as nodes in a weighted graph with edges defined as the geodesic distance between 2 points.

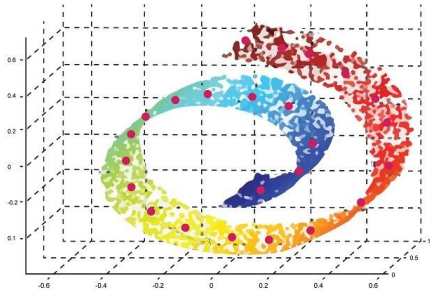

Figure 4

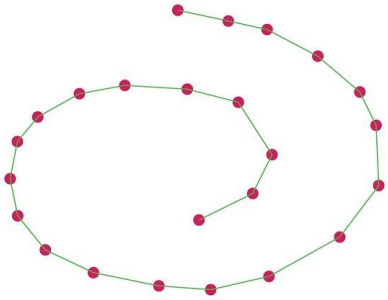

Figure 5

3 IsoMap Algorithm

- Find the neighbors of each point
  - Points within a fixed radius
  - K nearest neighbors
- Construct a graph with those nodes
- Compute neighboring edge weights (distance between neighbors)
- Compute weighted shortest path between all points
- Run MDS using the computed distance above

4 References:

For Graph theory(Edges and nodes),
https://en.wikipedia.org/wiki/Graph_theory
For classical MDS,
https://en.wikipedia.org/wiki/Multidimensional_scaling
For more details on Isomap,
https://en.wikipedia.org/wiki/Isomap