Isomap

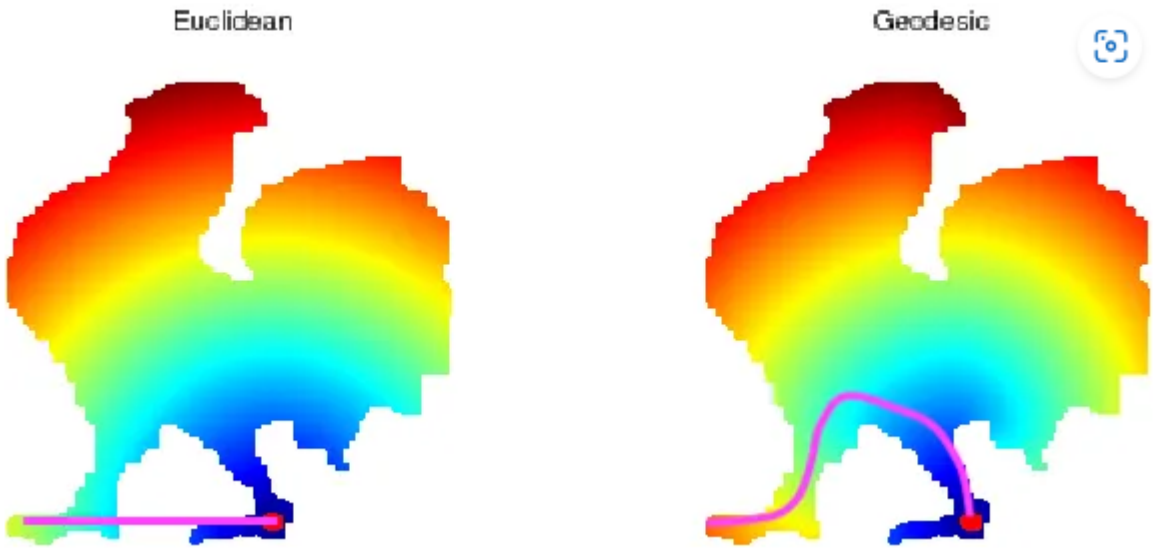
1. Introduction:

- Isomap stands for "Isometric Mapping". It is a non linear dimensionality reduction technique that aims to preserve the intrinsic geometric structure of high-dimensional data in a lower dimensional space. Unlike linear techniques like PCA, Isomap captures non 0linear relationships, which can be critical for datasets with complex structures.
- Isomap preserves Global Data Structure by considering the geodesic distance between data points. This is crucial for intricate patters. This is also suitable for Manifold Learning.
- When data is projected **non-linearly** after dimension reduction(Isomap). This class of algorithms is also called **manifold learning** algos.
- It is a manifold learning algorithm that tries to preserve the geodesic distance between the samples while reducing the dimensions.

---

2. Jargons:

2.1 Geodesic Distance:



2.2 Double Centering a matrix:

It means to transform a matrix such that

- mean for any row = 0
- mean for any column = 0

2.2.1 This is computed as follows:

Step 1: For a Matrix A, prepare Matrix B and C such that
Step 2: Matrix B

Mean(col_1)	**Mean(col_2)**	**Mean(col_3)**
Mean(col_1)	**Mean(col_2)**	**Mean(col_3)**
Mean(col_1)	**Mean(col_2)**	**Mean(col_3)**

Step 3: Matrix C

Mean(row_1)	**Mean(row_2)**	**Mean(row_3)**
Mean(row_1)	**Mean(row_2)**	**Mean(row_3)**
Mean(row_1)	**Mean(row_2)**	**Mean(row_3)**

Step 4: Double centered Matrix A = A-B-C+Mean(A)

2.3 Dissimilarity Matrix:

It is a matrix that represents dissimilarity between points in a dataset. This dissimilarity can be calculated using any measure. Though the most common measure is the distance between points. The more the distance, the more dissimilar the samples are.

	A	B	C	D	E	F
A	0	16	47	72	77	79
B	16	0	37	57	65	66
C	47	37	0	40	30	35
D	72	57	40	0	31	23
E	77	65	30	31	0	10
F	79	66	35	23	10	0

---

3. Isometric Mapping:

Step 1: Calculating Geodesic Distance:

1. We can calculate the adjacency matrix for all the points in the dataset using KNN where K= [3,4].
2. Then using this weughted matrix, we apply dijkstra algorithm to find the shortest distance.
3. Hence we can calculate the geodesic distance between 2 points.

Step 2: Dissimilarity matrix:

Step 3: Square the Dissimilarity matrix and double center it.

Step 4: Eigendecomposition & choosing ‘k’ eigenvectors:

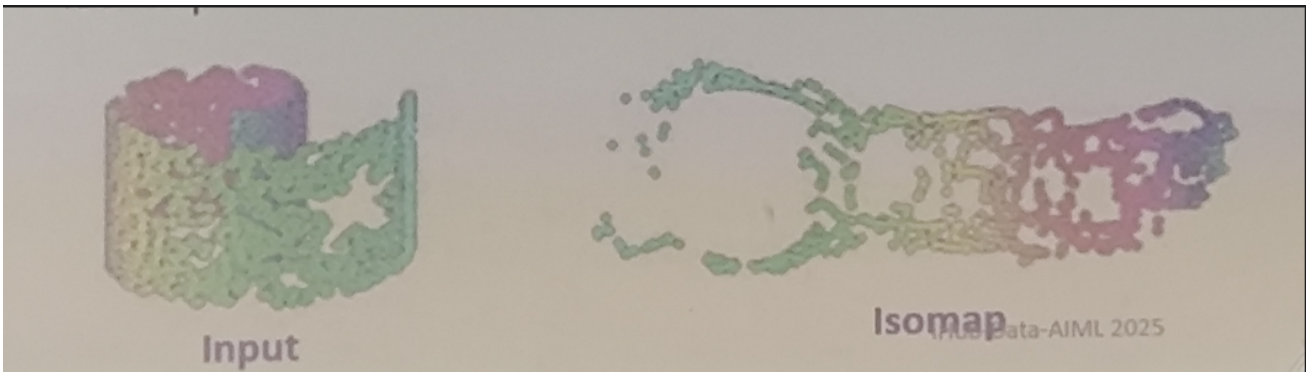
This is something similar to what we do in PCA after calculating the correlation matrix.

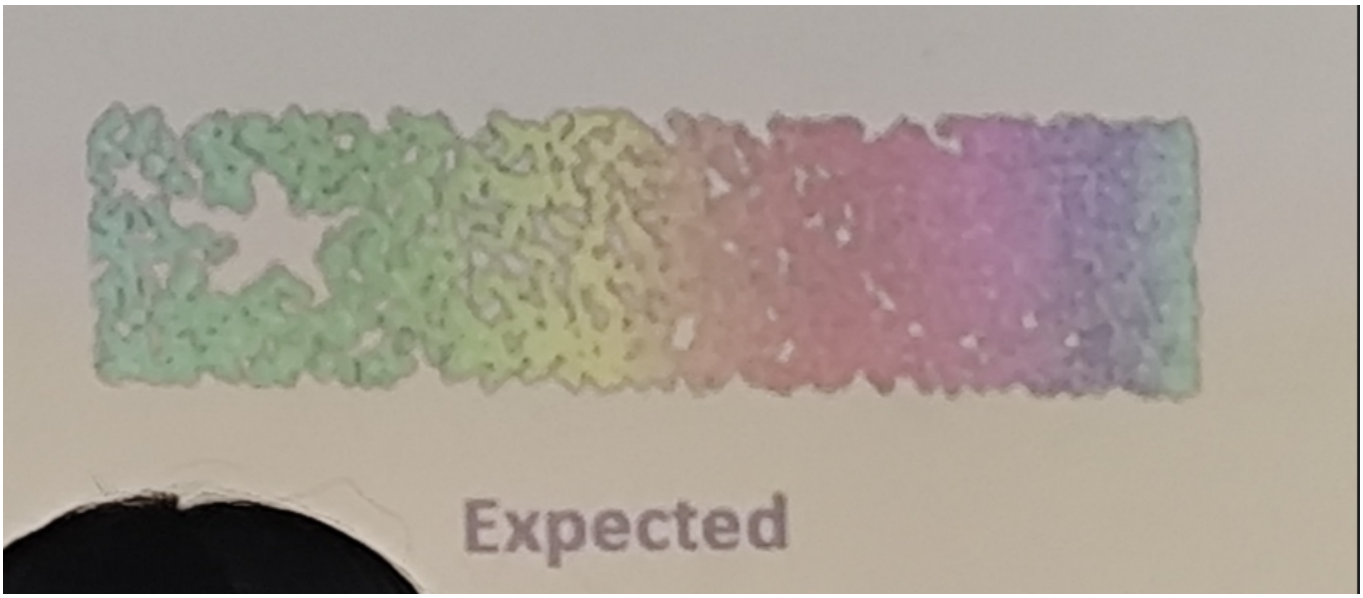
---

4. Drawbacks:

A few drawbacks always exist

- This version of Isomap is computationally heavy. Though other versions exist which are comparatively light.
- Parameter tuning for KNN is important as a wrong selection of ’n’ can be devastating. We can use Radius nearest neighbors as well for forming the neighborhood graph
- Isomap suffers from non-convexity such as holes on manifolds.





---

5. References:

1. Dimension Reduction using Isomap. Something you need for nonlinear data | by Mehul Gupta | Data Science in Your Pocket | Medium