UNIVERSIDAD POLITÉCNICA DE YUCATÁN

Machine Learning

1LU - Task: Research on Dimensionality Reduction
ISOMAP

Team 4:
Oswaldo Josue Gomez Gonzalez
Israel Prado Pineda
Karen Valeria Villanueva Novelo

ISOMAP

Isomap stands for Isometric mapping. It is another distance-preserving nonlinear dimensionality reduction technique that is based on spectral theory. This technique solves nonlinear dimensionality reduction by preserving the geodesic distances between the points in the lower dimension. Firstly, it finds the approximate distance between all the pairs of points on the higher dimensional manifold. Then it finds a set where distances between pairs of points in the low dimensional space are roughly the same as the distances between the pairs of points on the manifold.

ISOMAP ALGORITHM

1. For every point in the original dataset, find it is k-Nearest Neighbor (with respect to the actual distance in the high-dimensional space).

2. Plot the k-nearest neighbor graph G = (V, E). Every point of the dataset is a vertex in the graph (hence, there n-vertices in total), and every point in this graph is connected to its k-nearest neighbors by an edge.

3. Compute the pairwise distances between all pairs of points in the graph using the graph's geodesic distance as the metric and represent it using a matrix (say A).

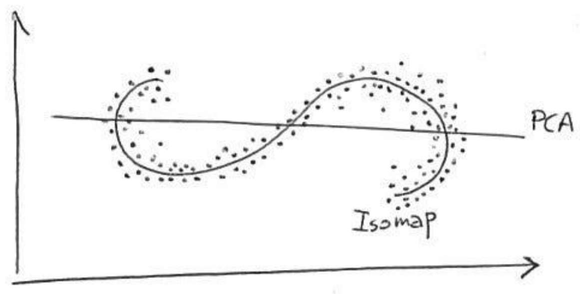
4. Find points in the lower-dimensional space such that pairwise distances between points are approximately the same as distances between the points on the graph.

The algorithm's output is the low dimensional representation of all the points computed in the final step. The idea is that if we have enough points on the high dimensional manifold, that is if they are packed tightly, and because the manifold is locally Euclidean, the k-nearest neighbors on the original high dimensional space are also the manifold's k-nearest neighbors. We then form the nearest neighbor graph based on the idea that the shortest distances on the graph correspond to the manifold.

After constructing the neighborhood graph, you want to understand how points are connected and how far apart they are. This is crucial for determining which paths are shorter or longer in the high-dimensional space.

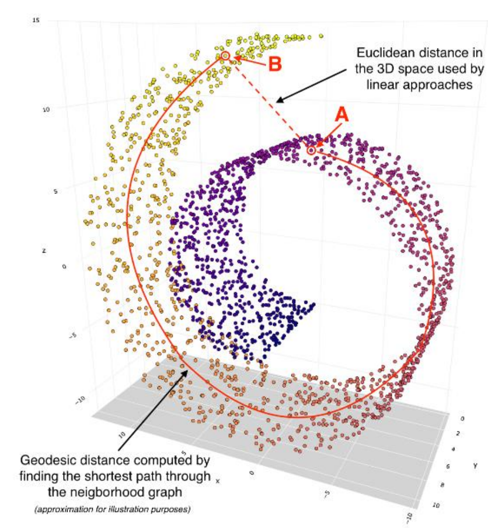
We either use the Floyd Warshall algorithm or Dijkstra's algorithm between all pairs to get the pairwise distance between all pairs of points on the manifold. Now we simply need to find points on low dimensional space such that Euclidean distances in this lower dimensional space are approximately equal to the shortest distance on the manifold between these points.

You project the data points from the high-dimensional space into a lower-dimensional space (d dimensions), preserving the relationships and distances you calculated in the previous steps as accurately as possible.



For most types of high dimensional data, there is likely a non-linear relationship and therefore we need to maintain this shape when we reduce the dimensions. This is where Manifold Learning techniques like Isomap come into play.

Geodesic distance is the distance between 2 points following the path available/possible between the two points whereas Euclidean distance doesn't have a path constraint to follow. It is the length of a straight line from point 'a' to 'b'. The geodesic distances between two points can be approximated by graph distance between the two points, so, even though the Euclidean metric does a relatively poor job in approximating the distance between two points in nonlinear manifolds, the geodesic metric of distances gives satisfactory results. Hence, while dealing with finding the approximated distance between points on a nonlinear manifold, using the geodesic metric of distances is a better fit. Isomap uses the concept of geodesic distance to solve the problem of dimensionality reduction.



The goal of Isomap is to maintain a geodesic distance between two points. Geodesic is more formally defined as the shortest path on the surface itself. By understanding the pair-wise geodesic distances, Isomap aims to approximate the geometry of the data before projecting it down into the specified dimension. It is a manifold learning algorithm that tries to preserve the geodesic distance between samples while reducing the dimension.

ISOMAP technique combines the best algorithmic features of PCA and MDS–computational efficiency, global optimality, and asymptotic convergence guarantees-with a flexibility to learn structures in a broad class of non-linear manifolds.

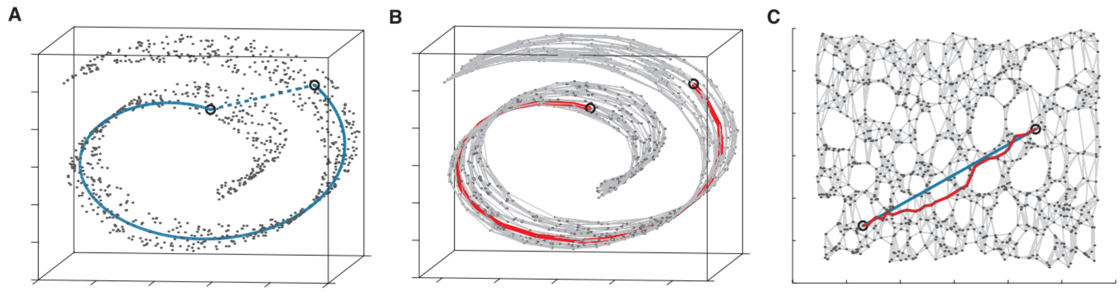


Fig. 3. The "Swiss roll" data set, illustrating how Isomap exploits geodesic paths for nonlinear dimensionality reduction. **(A)** For two arbitrary points (circled) on a nonlinear manifold, their Euclidean distance in the high-dimensional input space (length of dashed line) may not accurately reflect their intrinsic similarity, as measured by geodesic distance along the low-dimensional manifold (length of solid curve). **(B)** The neighborhood graph *G* constructed in step one of Isomap (with $K = 7$ and $N = 1000$ data points) allows an approximation (red segments) to the true geodesic path to be computed efficiently in step two, as the shortest path in *G*. **(C)** The two-dimensional embedding recovered by Isomap in step three, which best preserves the shortest path distances in the neighborhood graph (overlaid). Straight lines in the embedding (blue) now represent simpler and cleaner approximations to the true geodesic paths than do the corresponding graph paths (red).

ISOMAP guarantees a global optimum and is also guaranteed asymptotically to recover the true dimensionality and geometric structure of a strictly larger class of non-linear manifolds. Like the swiss roll example above, there are many non-linear manifolds whose intrinsic geometry is that of a convex region in Euclidean space, but whose ambient geometry in high dimensional data is very folded, twisted, or curved. For non-Euclidean manifolds, such as a hemisphere or the surface of the donut, ISOMAP is still able to produce a globally optimum low-dimensional Euclidean representation.

These guarantees of asymptotic convergence rests on a proof that as the number of data point increases, the graph distances Dg(i,j) provide increasingly better approximations to the intrinsic geodesic distances DM(i,j), becoming arbitrarily accurate in the limit of the input data. How quickly Dg(i,j) converges to DM(i,j), depends on the number on certain parameters of the manifold as it lies within high dimensional space (radius of curvature and branch separation) and on the density of points.

Isomap's global coordinates provide a simple way to analyze and manipulate high dimensional observations in terms of their intrinsic non-linear degrees of freedom.

EXAMPLES

Example: Viewing face data in 3D

Suppose we have a data set that represents different points on a three-dimensional face. When we try to display this data in a two-dimensional space (such as a computer screen), the information is lost or distorted due to the curvatures and non-linear features of the face.

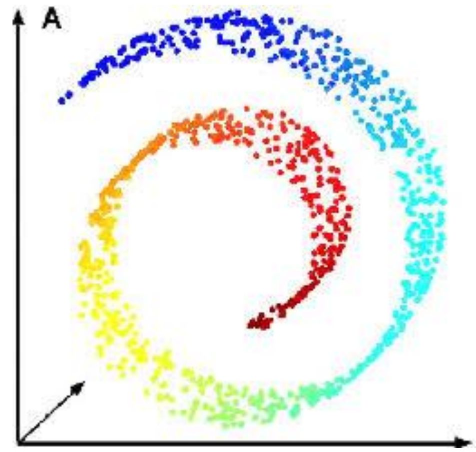
<u>Data collection:</u> Collection of 3D data of a face through 3D scanning.

<u>Nearest neighbor graph construction:</u> For each point (or node) on the face, you find a set of nearest neighbor nodes. This can be done using the k-nearest neighbors method.

Calculating Geodesic Distances: Instead of calculating direct Euclidean distances between points, we use the nearest neighbors graph to calculate geodesic distances, which are the shortest distances along the surface of the face.

Dimensionality reduction with ISOMAP: we use ISOMAP to project the 3D data into a 2D space. ISOMAP does this by preserving the geodesic distances between points, resulting in a two-dimensional representation that retains the inherent structure of the original data.

Visualization: We can now visualize the structure of the face in 2D in a way that closely reflects the relationships between points in the original 3D space



Example: Visualization of hand postures

Supose that we are conducting a study on ergonomics and we want to understand how the human hand moves in different activities. You have a special glove with sensors that can capture the position and orientation of each joint in your hand. This generates a significant amount of high-dimensional data for each posture or movement.

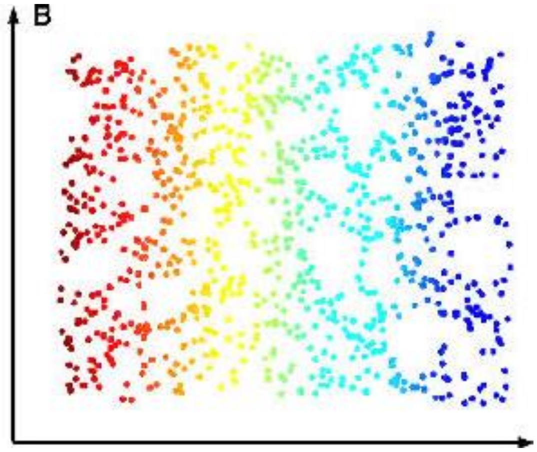
Data Collection: We have several participants perform a series of tasks, such as grabbing objects of different shapes, writing with a pencil, pointing at something, etc. Each of these activities is recorded in terms of the position and orientation of the hand joints.

Nearest Neighbor Graph Construction: Since each hand pose is represented as a point in a high-dimensional space, we need to identify nearest neighbors for each point using the k-nearest neighbors method.

Calculating geodesic distances: Instead of looking at direct distances between poses (which may not capture the true "distance" between two poses), we can calculate geodesic distances using the nearest neighbors graph.

Dimensionality reduction with ISOMAP: ISOMAP can be applied to project this high-dimensional data (many joints with multiple degrees of freedom) into a 2D or 3D space.

Visualization and interpretation: In the small space, similar hand postures will be close to each other. For example, all variants of a clenched fist could be grouped in one region, while pointing postures would be in another region.



References

Dobilas, S. (2022, 5 febrero). ISOMAP embedding — an awesome approach to non-linear dimensionality reduction. *Medium*. https://towardsdatascience.com/isomap-embedding-an-awesome-approach-to-non-linear-dimensionality-reduction-fc7efbca47a0

Face recognition using extended ISOMaP. (2002). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/1039901

Ghojogh, B. (2020, 17 septiembre). *Multidimensional scaling, Sammon Mapping, and ISOMap: tutorial and survey*. arXiv.org. https://arxiv.org/abs/2009.08136

ISOMAP. (2017, July 15). https://thegreedychoice.github.io/machinelearning/isomap

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

Pal, A. K. (2020). Tutorial: Dimension Reduction - IsoMap. *Paperspace Blog*. https://blog.paperspace.com/dimension-reduction-with-isomap/

Tripathy, B. K., Anveshrithaa, S., & Ghela, S. (2021). ISOMAP. In CRC Press eBooks (pp. 53–65). https://doi.org/10.1201/9781003190554-7