Giang Vu
STAT 2221
Fall 2021

Homework 7

A summary of "Rehabilitating Isomap: Euclidean Representation of Geodesic Structure"

In practice, there might be instances that we suspect our high dimensional feature data lie on a low-dimensional manifold (not a Euclidean space), so several manifold learning methods were proposed to help with this problem. The seminar talk from Dr. Michael Trosset discussed this topic of nonlinear dimensionality reduction, specifically the Isomap method as a manifold learning tool.

Firstly, Dr. Trosset did a brief review of Isomap:

- Goal: To represent the geodesic structure of an intrinsically non-Euclidean object (the manifold) in a Euclidean space (turning geodesics into straight lines, basically). (Geodesic distance = the shortest distance between two points)
- The Isomap algorithm has 3 steps (as introduced in Izenman as well):
    - constructing a graph of the points that we assume to lie on an unknown manifold
    - constructing the shortest path distances on the graph
    - embedding the shortest path distances using multidimensional scaling (MDS)
- Because we do not know the actual manifold, we only get access to the points on the manifold, we construct a graph in which each of the points is a vertex of the graph and we connect nearby points to form edges. Then we approximate geodesics by finding the shortest path between two points, which are of course not perfectly accurate geodesics. After that, we compute the entire set of those "imperfectly" approximated geodesics and feed those distances into MDS. MDS will then average out all those "imperfections" or redundant information about the geodesic structure, so we end up with a reasonable low dimensional Euclidean representation of the actual manifold itself.
- The literature so far, as in Izenman as well, has set two requirements for Isomap, where we have to have a global isometry to a convex set for the method to work.

Then, Dr. Trosset continued to argue against the requirement of "convexity" by examining the classic example where we sample from a rectangle with a missing rectangular strip punched out in the center. This example has been used in the literature as evidence that Isomap provides unsatisfactory result without convexity, because the Swiss roll obtained from Isomap for this example misses the corresponding strip.

- Dr. Trosset argued that while Isomap does fail to do a parameterization recovery with this example, but it still does what it is supposed to do, which is find an Euclidean representation of the manifold's geodesic structure.
- According to him, parameterization recovery is a very specialized problem for a very limited variety of manifolds (Swiss roll is one of them), so it is not that important in the context of manifold learning.
- Manifold learning boils down to finding an Euclidean representation of the Riemannian distances, and Isomap achieves that goal by adding an extra step in between, it approximates Riemannian distances with shortest path distances, then approximates the latter with Euclidean distances.

He also pointed out that in the original paper where Isomap was proposed, Bernstein et. al (2000), the assumption of convexity was only stated but not really used by the actual authors themselves when they derived the result of Isomap. That paper result actually stated that over any two points on the manifold, the worst case of the best approximation for shortest path distance on your graph would always converge to the true manifold Riemannian distance in probability, and this result did not require any convexity assumption.

The takeaway from this talk is that the use case for Isomap is actually far more general than we think it is, and we do not have to limit ourselves with the convexity requirement like suggested in the literature.