CAP 6617 ADVANCED MACHINE LEARNING

NAME: SANKET KUMAR   UF ID: 4513-3352

ISOMAP

ISOMAP is a dimensionality reduction approach that combines the major algorithmic features of Principle Component Analysis(PCA) and Multidimensional Scaling(MDS) with the flexibility to learn a broad class of nonlinear manifolds.

- Many data sets contain essential nonlinear structures that are invisible to PCA and MDS. Previous attempts to extend PCA and MDS to nonlinear data sets fall into two broad classes, each of which suffers from limitations.
- Local linear techniques are not designed to represent the global structure of a data set within a single coordinate system. Nonlinear techniques based on greedy optimization procedures attempt to discover global structure, but lack the crucial algorithmic features that ISOMAP inherits from PCA and MDS.

Example taken in this paper is of a two-dimensional Swiss-roll manifold. Only the geodesic distances reflect the true low-dimen- sional geometry of the manifold, but PCA and MDS effectively see just the Euclidean structure; thus, they fail to detect the intrinsic two-dimensionality. The important part is estimating the geodesic distance between faraway points approximated by finding shortest paths in a graph. Following is the procedure:

1. Define a graph G by connecting all points if they are inside a radius ε or is one of K-nearest neighbors.

2. Compute shortest paths between all pairs of points in G and store in the matrix D.

3. Construct d-dimensional embedding. Set the p-th component of the d-dimensional coordinate vector y in Euclidean space Y equal to square-root of $\lambda_p v_p$ where λ_p is the p-th eigenvalue of matrix D and v_p the p-th eigenvector.

- ISOMAP's global coordinates provide a simple way to analyze and manipulate high-dimensional observations in terms of their intrinsic nonlinear degrees of freedom.
- The technique may be applied wherever nonlinear geometry complicates the use of PCA or MDS.
- ISOMAP may be combined with linear extensions of PCA based on higher order statistics.