OM-S20-19: OM for ML - II Manifold Learning

C. V. Jawahar

IIIT Hyderabad

http://preon.iiit.ac.in/om_quiz

27 Mar 2020

Problems of Interest

- **Visualization** Given a set of N points in R^d, create a lower (k) dimensional (often 2 or 3) representation of the data in R^k. When human wants to look at the data!!.
- **Clustering** Given a set of N points in R^d, partition the set into K (often 2) subsets.
  There is no clear "goal" or "desirability" and the task could get poorly defined.

PCA and Limitations

- Well known dimensionality reduction in ML
- Compute Covariance matrix of the data
- Choose k dimensions (or principal components) and then project.

1. Does reduce dimensions which could help in visualization. But that is not the goal.
2. Reduction in dimension might have removed important cues/structure in the data.
3. Not all data lie in lower dimensional space.

K-Means for Clustering

Minimize:

$$J = \sum_{j=1}^{K} \sum_{i=1}^{N_j} ||x_j - \frac{1}{N_j} \sum_{p=1}^{N_j} x_p||$$

NP-hard to minimize in general. Iteratively optimize as:

- Choose k centers at random.
- Assign points to closest center.
- Compute new centers are clusters centroids.

1. Challenege of local minima in optimization (non-convex). k-means++ with smarter initialization.
2. Assumption of Clusters being compact and well separated.
3. How do we separate elongated clusters?

Ratio Cuts and Clustering (from L14)

$$Cut(A, \bar{A}) = \sum_{i \in A} \sum_{j \in \bar{A}} w_{ij}$$

can lead to unwanted solution (such as one tiny cluster corresponding to an outlier).

Ratio Cut: Objective:

$$\frac{Cut(A, \bar{A})}{|A|} + \frac{Cut(A, \bar{A})}{|\bar{A}|} \Leftrightarrow \sum_{ij} w_{ij}(f_i - f_j)^2 \Leftrightarrow f^T L f$$

where f_i is $\frac{\sqrt{|\bar{A}|}}{\sqrt{|A|}}$ if $i \in A$ and $-\frac{\sqrt{|A|}}{\sqrt{|\bar{A}|}}$ if $i \in \bar{A}$

This is equivalent to optimization of :

$$f^T L f \text{ subject to: } f^T f = 1$$

Note: the constraint on f is a "relaxation" of what we really want to optimize.

Clustering/Cut Objective (Revisit from L14)

Ratio Cut:

$$\frac{Cut(A,B)}{|A|} + \frac{Cut(B,A)}{|B|}$$

Normalized Cut:

$$\frac{Cut(A,B)}{Cut(A,A) + Cut(A,B)} + \frac{Cut(B,A)}{Cut(B,A) + Cut(B,B)}$$

Min-Max-Cut:

$$\frac{Cut(A,B)}{Cut(A,A)} + \frac{Cut(B,A)}{Cut(B,B)}$$

Comments: If clusters are

- well separated, all three give very similar and accurate results.
- marginally separated, NormCut and MinMaxCut give better results
- overlapping significantly, MinMaxCut tend to give more compact and balanced clusters.

Extensions to Multi-Way Cuts. (beyond this course)

MDS and Objective

Problem: Given the set X, find the set Y.
Given a pair of distances D^X, find a set of points in lower dimension that preserves the distances best.

$$\min_{Y} ||D^X - D^y||^2$$

Hints:

- What we have is the original distance matrix. May be of "full rank"
- Find a lower rank approximation (remember SVD) for the new matrix.

ISOMAP

- Construct a K nearest neighbour graph.
- Compute the shortest path between all points (as an estimation of Geodesic distance) D_G
- $$K = -\frac{1}{2}HDH$$

  (H is a centering matrix)
  - Find Eigen Vec of K as V.
  - Find top p Eigen Val of K as Λ

$$Y = \Lambda^{\frac{1}{2}}V^T$$

ISOMAP (from Science Dec 2000)

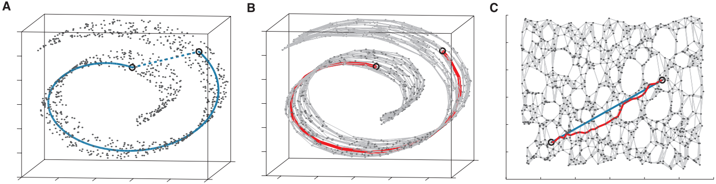

Fig. 3. The "Swiss roll" data set, illustrating how Isomap exploits geodesic paths for nonlinear dimensionality reduction. (**A**) For two arbitrary points (circled) on a nonlinear manifold, their Euclidean distance in the high-dimensional input space (length of dashed line) may not accurately reflect 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).

LLE (from Science Dec 2000)

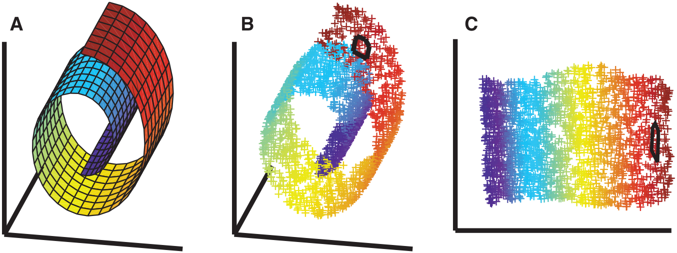

Fig. 1. The problem of nonlinear dimensionality reduction, as illustrated (*10*) for three-dimensional data (B) sampled from a two-dimensional manifold (**A**). An unsupervised learning algorithm must discover the global internal coordinates of the manifold without signals that explicitly indicate how the data should be embedded in two dimensions. The color coding illustrates the neighborhood-preserving mapping discovered by LLE; black outlines in (**B**) and (**C**) show the neighborhood of a single point. Unlike LLE, projections of the data by principal component analysis (PCA) (*28*) or classical MDS (*2*) map faraway data points to nearby points in the plane, failing to identify the underlying structure of the manifold. Note that mixture models for local dimensionality reduction (*29*), which cluster the data and perform PCA within each cluster, do not address the problem considered here: namely, how to map high-dimensional data into a single global coordinate system of lower dimensionality.

Locally Linear Embedding (LLE)

- Assumption of local linearity
- Start with the K-NN Graph. (similar to ISOMAP)
- Find W such that

$$\epsilon(W) = \sum_{i=1}^{t} ||x_i - \sum_{j=1}^{K} W_{ij} x_j||^2$$

- Now find Y such that

$$\min \phi(Y) = \sum_{i=1}^{t} ||y_i - \sum_{j=1}^{t} W_{ij} y_j||^2$$

Laplacian EigenMap

- Similar to LLE and ISOMAP but for dimensionality reduction
- Create a weiighted graph with weight as connectivity or similarity.
- $$\min_{Y} w_{ij}||y_i - y_j||^2$$
- Two points in high dimension should also be toghether in the lower dimensional space also.
- Connection to spectral clustering and LLE/ISPMAP/MDS.

Summary of Space

- PCA
- MDS
- K Means
- Ratio Cuts
- Normalized Cuts
- ISOMAP
- LLE
- Laplacian Eigen map