PROXIMITY GRAPHS FOR CLUSTERING AND MANIFOLD LEARNING

Miguel Á. Carreira-Perpiñán[1] and Richard S. Zemel[2]

[1]Dept. of CSEE, OGI, Oregon Health & Science University [2]Dept. of CS, University of Toronto, Canada

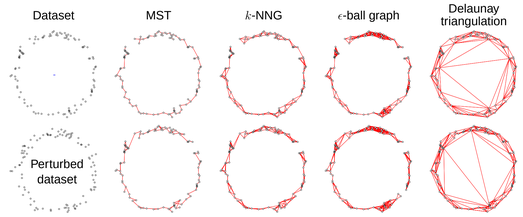
1 Abstract

Many machine learning algorithms for clustering or dimensionality reduction take as input a cloud of points in Euclidean space, and construct a graph with the input data points as vertices. This graph is then partitioned (clustering) or used to redefi ne metric information (dimensionality reduction). We propose new types of proximity graphs, based on ensembles of minimum spanning trees, that adapt locally to the structure of the data and are robust to noise. This improves the quality of the subsequent clustering or dimensionality reduction.

3 Problems with traditional graphs

- The graph parameter ϵ or k has to be chosen carefully to avoid:
  - connecting the wrong points (shortcuts)
  - not connecting the right points (disconnected graph, gaps)
- Local neighbourhoods are not adaptive (ϵ, k should depend on $\mathbf{x}_i$)
- The graphs are sensitive to small perturbations of the data
- Other types of graphs connect points nonlocally (DTG, RNG, GG)

After building the graph, the subsequent spectral algorithm is $\mathcal{O}(N^3)$, thus trial-and-error of graphs is very expensive.



2 Pairwise-distance methods

Given a cloud of data points $\{\mathbf{x}_i\}_{i=1}^N$, we want to learn statistical structure based on the pairwise distances $\{d_{ij}\}_{i,j=1}^N$. This implies a graph with vertices $\{\mathbf{x}_i\}$ which should represent the data manifold. E.g.:

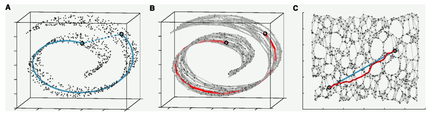
- Dimensionality reduction: preserve metric information (implied by the graph) in low-dimensional space.
- Clustering: partition the graph to optimise a cut criterion.

Most traditional proximity graphs are defined by a rule that tells each point what points to link to, e.g. ϵ-ball graph, k-nearest neighbours graph, relative neighbourhood graph, Gabriel graph, etc. Others are defined by an objective function, e.g. the minimum spanning tree.

Dimensionality reduction: Isomap

1. Build proximity graph (ϵ-ball or k-NNG)
2. Approximate geodesic distances $\{g_{ij}\}_{i,j=1}^N$ by shortest-path lengths in the graph
3. Use multidimensional scaling to obtain low-dimensional points whose Euclidean distances optimally preserve the geodesic ones

Related spectral methods: MDS, LLE, Laplacian eigenmaps, SDE.



Spectral clustering: normalised cuts

1. Use fully-connected graph or, for image segmentation, fixed-grid
2. Obtain affinities $w_{ij} = e^{-d_{ij}^2/2\sigma^2}$ for a "good" scale σ (needs search over σ)
3. Cluster leading eigenvectors of $\mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}$ where $\mathbf{D} = \operatorname{diag}\left(\sum_i w_{ij}\right)$

This approximates the (NP-complete) normalised cut cost function $\operatorname{ncut}(A,B) = \operatorname{cut}(A,B)\left(\frac{1}{\operatorname{vol} A} + \frac{1}{\operatorname{vol} B}\right)$. Other cost functions are possible.

4 Two new types of proximity graphs: PMSTs and DMSTs

Perturbed minimum spanning trees (PMSTs)

1. Estimate local noise model for each data point: uniform zero-mean isotropic with std dev rd_i (d_i = average distance to the k nearest neighbours of $\mathbf{x}_i$ and $r \in [0,1]$)
2. Generate T jittered copies of entire dataset according to this noise
3. For each copy, build its MST
4. Average all MSTs

Result:

- Stochastic graph with edges $e_{ij} \in [0,1]$
- Number of edges is linear in N
- Insensitive to noise by construction
- Essentially deterministic for large T

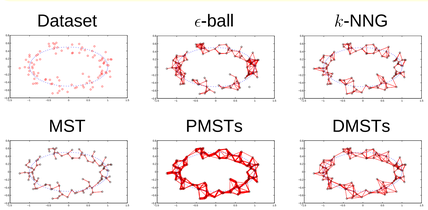
Disjoint minimum spanning trees (DMSTs)

Deterministic collection of t MSTs such that the nth tree (for $n = 1,\dots,t$) is the MST of the data subject to not using any edge already in the previous $1,\dots,t-1$ trees. Construction algorithm:

1. Sort edge list by increasing distance d_{ij}
2. Run Kruskal's algorithm t times picking edges without replacement

Result:

- Binary graph with edges $e_{ij} \in \{0,1\}$
- Number of edges is linear in N
- Relatively insensitive to noise
- Deterministic



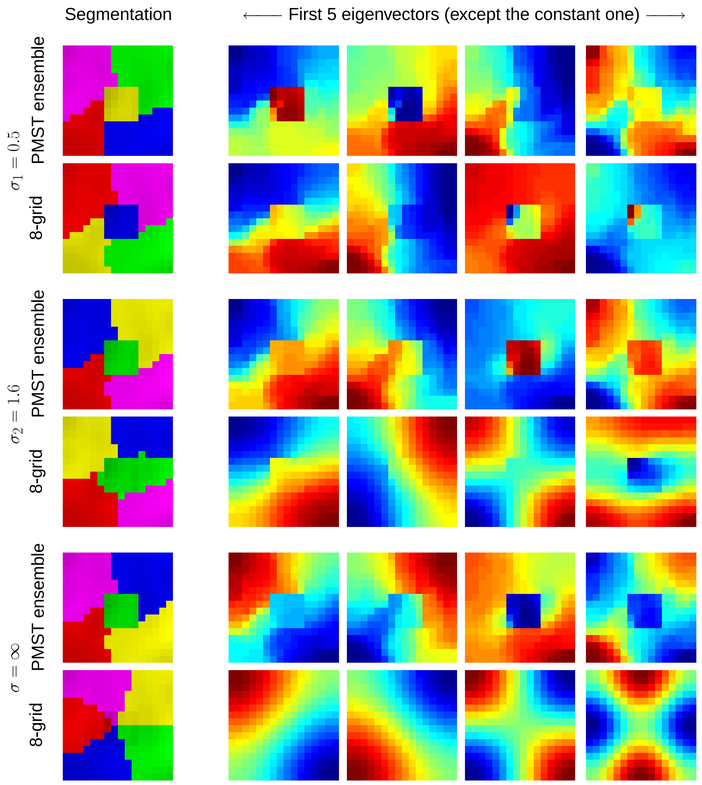
An ensemble of MSTs gives a good representation of the manifold. Each MST:

- uses short edges, thus avoiding shortcuts, and gives a good skeleton of the data
- is very sparse, but the combination fleshes out the graph

Computational complexity: PMSTs: $\mathcal{O}(TN^2 \log N)$; DMSTs: $\mathcal{O}(N^2(\log N + t))$

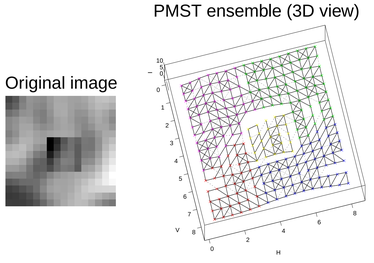
This is: Just a bit more than searching for nearest neighbours; Much less than the subsequent $\mathcal{O}(N^3)$ spectral algorithm

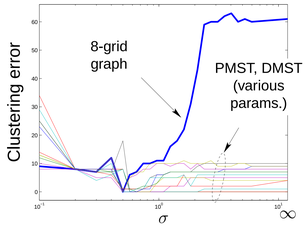
5 Results: spectral clustering (normalised cut)



PMST ensemble (3D view)

Original image



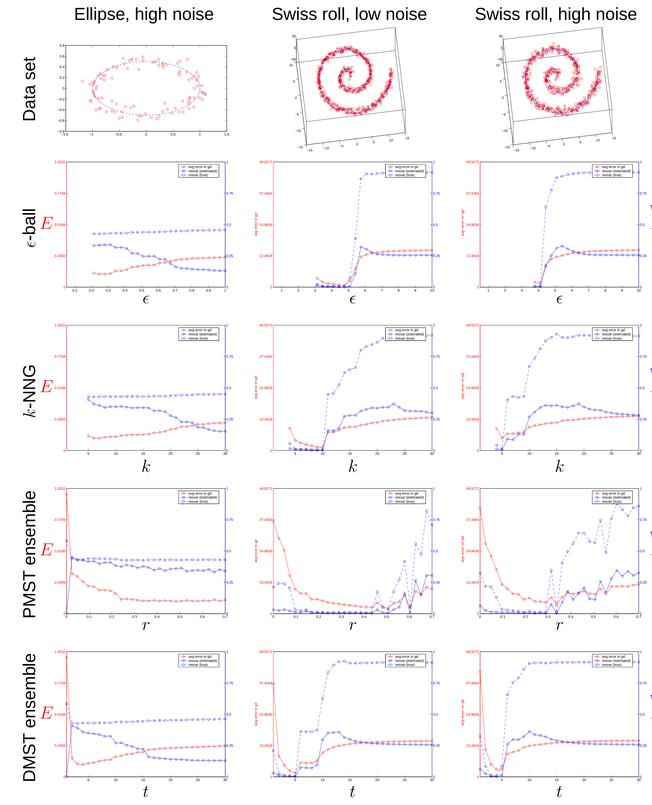


Task: segmenting the occluder from the background in the greyscale image, using normalised cut (asking for 5 clusters). The affinities are the componentwise product of the Gaussian affinity $w_{ij} = \exp(-\frac{1}{2}(d_{ij}/\sigma)^2)$ and the edge weight e_{ij}. We examine the performance over different scales σ.

- Using the 8-grid graph: good segmentation only for a narrow range of scales $\sigma \in [0.2, 1]$
- Using the PMST or DMST ensemble (under many settings of their parameters): almost perfect segmentation over a large range of scales $\sigma \in [0.2, \infty]$

Reason: the PMST or DMST ensemble represents the data manifold better (note non-edges around the occluder) and so facilitates the graph cut.
Thus, having a good graph can eliminate an expensive search over scales (each σ value costs $\mathcal{O}(N^3)$).

6 Results: dimens. reduction (Isomap)



For each dataset, we examine the preservation of geodesic distances for several graph types (each over a range of their parameters):

- Average error in the geodesic distances: $E = \frac{1}{N^2}\|\hat{\mathbf{G}} - \mathbf{G}\|$
- Isomap's estimated residual variance: $\hat{V} = 1 - R^2(\hat{\mathbf{G}}, \mathbf{D}_\mathcal{Y})$
- True residual variance: $V = 1 - R^2(\mathbf{G}, \mathbf{D}_\mathcal{Y})$

where the matrices $\hat{\mathbf{G}}$, $\mathbf{G}$ and $\mathbf{D}_\mathcal{Y}$ contain the true geodesic distances, estimated (graph shortest-path) geodesic distances and Euclidean distances in the low-dimensional embedding, respectively.
The PMST or DMST ensembles are more robust than the ϵ-ball or k-NNG graphs, particularly for high noise (note the narrow range of ϵ and k for which the performance is good). Again, this eliminates an expensive search (over ϵ or k).

7 Conclusions and future work

We have pointed out the importance of proximity graphs as scaffolds for clustering and manifold learning, and how the graph should represent the structure of the data manifold. We have introduced two new types of proximity graphs based on ensembles of minimum spanning trees, which are relatively inexpensive to compute and robust across many noise levels and parameter settings. This limits the required parameter search for clustering and manifold learning.
Although at present we do not have an objective function for unsupervised graph learning, the MST ensembles tend to reduce both bias and variance of the average error for the true geodesic distances.
Future work: manifold-aligned noise model; noise model for non-Euclidean data; study stochastic graphs.