Data-dependent representations: Laplacian Eigenmaps, continued

Wojciech Czaja and Brian Hunt

April 16, 2015

Norbert Wiener Center for Harmonic Analysis and Applications

Outline

Norbert Wiener Center for Harmonic Analysis and Applications

Outline

Norbert Wiener Center
for Harmonic Analysis and Applications

Data Organization and Manifold Learning

- There are many techniques for Data Organization and Manifold Learning, e.g., Principal Component Analysis (PCA), Locally Linear Embedding (LLE), Isomap, genetic algorithms, and neural networks.
- We are interested in a subfamily of these techniques known as *Kernel Eigenmap Methods*. These include Kernel PCA, LLE, Hessian LLE (HLLE), and Laplacian Eigenmaps.
- Kernel eigenmap methods require two steps. Given data space X of N vectors in $\mathbb{R}^D$.
  1. Construction of an $N \times N$ symmetric, positive semi-definite kernel, K, from these N data points in $\mathbb{R}^D$.
  2. Diagonalization of K, and then choosing $d \leq D$ *significant* eigenmaps of K. These become our new coordinates, and accomplish dimensionality reduction.

We are particularly interested in diffusion kernels K, which are defined by means of transition matrices.

Kernel Eigenmap Methods for Dimension Reduction - Kernel Construction

- Kernel eigenmap methods were introduced to address complexities not resolvable by linear methods.
- The idea behind *kernel methods* is to express correlations or similarities between vectors in the data space X in terms of a symmetric, positive semi-definite kernel function $K : X \times X \to \mathbb{R}$. Generally, there exists a Hilbert space $\mathbb{K}$ and a mapping $\Phi : X \to \mathbb{K}$ such that

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle.$$

  Then, diagonalize by the spectral theorem and choose significant eigenmaps to obtain dimensionality reduction.
- Kernels can be constructed by many kernel eigenmap methods. These include Kernel PCA, LLE, HLLE, and Laplacian Eigenmaps.

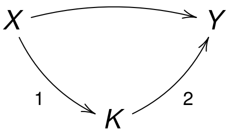
Norbert Wiener Center for Harmonic Analysis and Applications

Kernel Eigenmap Methods for Dimension Reduction - Kernel Diagonalization

- The second step in kernel eigenmap methods is the diagonalization of the kernel.
- Let e_j, $j = 1, \dots, N$, be the set of eigenvectors of the kernel matrix K, with eigenvalues λ_j.
- Order the eigenvalues monotonically.
- Choose the top $d << D$ significant eigenvectors to map the original data points $x_i \in \mathbb{R}^D$ to $(e_1(i), \dots, e_d(i)) \in \mathbb{R}^d$, $i = 1, \dots, N$.

Norbert Wiener Center for Harmonic Analysis and Applications

Data Organization



There are other alternative interpretations for the steps of our diagram:

1. Constructions of kernels K may be independent from data and based on principles.
2. Redundant representations, such as frames, can be used to replace orthonormal eigendecompositions.

We need not select the target dimensionality to be lower than the dimension of the input. This leads, to data expansion, or data organization, rather then dimensionality reduction.

Operator Theory on Graphs

- Presented approach leads to analysis of operators on data-dependent structures, such as graphs or manifolds.
- Locally Linear Embedding, Diffusion Maps, Diffusion Wavelets, Laplacian Eigenmaps, Schroedinger Eigenmaps
- Mathematical core:
    - Pick a positive semidefinite bounded operator A as the infinitesimal generator of a semigroup of operators, e^{tA}, $t > 0$.
    - The semigroup can be identified with the Markov processes of diffusion or random walks, as is the case, e.g., with Diffusion Maps and Diffusion Wavelets
    - The infinitesimal generator and the semigroup share the common representation, e.g., eigenbasis

Norbert Wiener Center for Harmonic Analysis and Applications

Laplacian Eigenmaps

- M. Belkin and P. Niyogi, 2003.
- Points close on the manifold should remain close in $\mathbb{R}^d$.
- Use Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ to control the embedding.
- Use discrete approximations for practical problems.
- Proven convergence (Belkin and Niyogi, 2003 – 2008).
- Gave rise to Diffusion Maps and Diffusion Wavelets, among others.

Laplacian Eigenmaps - Implementation

1. Put an edge between nodes i and j if x_i and x_j are close. Precisely, given a parameter $k \in \mathbb{N}$, put an edge between nodes i and j if x_i is among the k nearest neighbors of x_j or vice versa.
2. Given a parameter $t > 0$, if nodes i and j are connected, set $W_{i,j} = e^{-\frac{\|x_i - x_j\|^2}{t}}$.
3. Set $D_{i,i} = \sum_j W_{i,j}$, and let $L = D - W$. Solve $Lf = \lambda Df$, under the constraint $y^\top Dy = Id$. Let $f_0, f_1, \ldots, f_d$ be $d + 1$ eigenvector solutions corresponding to the first eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_d$. Discard f_0 and use the next d eigenvectors to embed in d-dimensional Euclidean space using the map $x_i \to (f_1(i), f_2(i), \ldots, f_d(i))$.

Swiss Roll

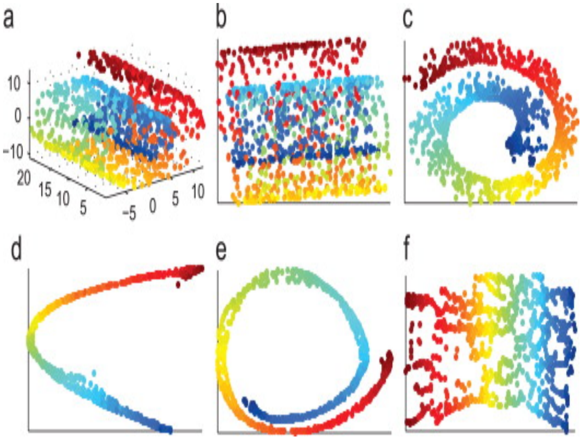


Figure : a) Original, b) PCA, c–f) LE, J. Shen et al., Neurocomputing, Volume 87, 2012

Norbert Wiener Center for Harmonic Analysis and Applications

Swiss Roll

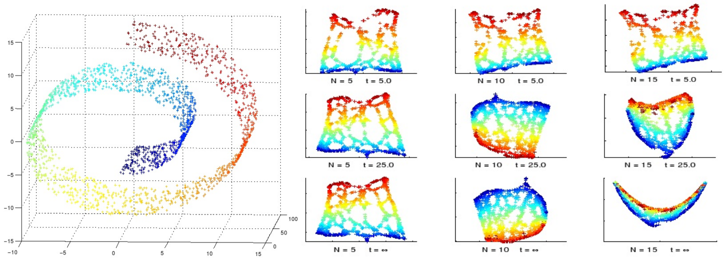


Figure : a) Original, b) LE

From eigenproblems to optimization

Consider the following minimization problem, $y \in \mathbb{R}^d$,

$$\min_{y^\top Dy = Id} \frac{1}{2} \sum_{i,j} \|y_i - y_j\|^2 W_{i,j} = \min_{y^\top Dy = E} tr(y^\top Ly).$$

Its solution is given by the d minimal non-zero eigenvalue solutions of $Lf = \lambda Df$ under the constraint $y^\top Dy = Id$.

Norbert Wiener Center for Harmonic Analysis and Applications