Locality Preserving Projections
– Xiaofei He and Partha Niyogi

Mohit Sinha

March 27, 2017
The Problem Setup

The dimensionality reduction problem

The problem of linear dimensionality reduction is the following. Given a set $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$, find a mapping that maps these $m$ points to a set of points $y_1, y_2 \ldots y_m \in \mathbb{R}^l (l << n)$, such that $y_i$ “represents $x_i$.

- PCA finds a linear mapping that captures the maximum variance.

Here, the focus is on the special case where $x_1, x_2 \ldots x_m \in \mathcal{M}$ and $\mathcal{M}$ is a nonlinear manifold embedded in $\mathbb{R}^n$. 
Unrolling the manifold
The Algorithm

- **Constructing the Adjacency Graph:** Put an edge between $i$ and $j$ if $x_i$ and $x_j$ are “close”.
  - $\epsilon$-neighborhoods: $i$ and $j$ are connected by an edge if $||x_i - x_j||^2 < \epsilon$
  - $k$-nearest neighbors

- **Choosing the Weights:** Let $W_{ij}$ denote the weight of an edge between $i$ and $j$.
  - Heat Kernel: $W_{ij} = e^{-\frac{||x_i - x_j||^2}{t}} \quad t \in \mathbb{R}$
  - Simple Minded: $W_{ij} = 1$
**Eigenmaps:** Compute the eigenvalues and eigenvectors for the generalized eigenvector problem:

\[ XLX^T a = \lambda XDX^T a \]  

(1)

where \( D \) is diagonal with \( D_{ii} = \sum_j W_{ij} \), \( L = D - W \) is the Laplacian matrix and \( i \)th column of \( X \) is \( x_i \).

**Linear Embedding:** Let the column vectors \( a_0 \ldots a_{l-1} \) be the solutions of equation (1), ordered according to their eigenvalues, \( 0 < \cdots < l-1 \). Thus, the embedding is as follows:

\[ x_i \rightarrow y_i = A^T x_i, A = (a_0, \ldots a_{l-1}) \]  

(2)

where \( y_i \) is a \( l \)-dimensional vector, and \( A \) is a \( n \times l \) matrix.
Consider the optimization problem

\[
\begin{align*}
\text{minimize} & \quad (y_i - y_j)^2 W_{ij} \\
\text{subject to} & \quad y^T D y = 1
\end{align*}
\]

where \( D \) is a diagonal matrix with \( D_{ii} = \sum_j W_{ij} \)

This reduces to:

\[
\begin{align*}
\text{arg min}_a & \quad a^T X L X^T a \\
\text{subject to} & \quad a^T X D X^T = 1
\end{align*}
\]
Kernel LPP

Let $\phi : \mathbb{R}^n \to \mathcal{H}$ denote the map from the Euclidean space to the Hilbert space.

The eigenvalue problem in this new space:

$$ [\phi(X) L \phi^T(X)] \nu = \lambda [\phi(X) D \phi^T(X)] \nu $$  \hspace{1cm} (4)

Consider the kernel function:

$$ K(x_i, x_j) = \phi^T(x_i) \phi(x_j) $$ \hspace{1cm} (5)

The problem reduces to:

$$ KLK \alpha = \lambda KDK \alpha $$ \hspace{1cm} (6)

where $\nu = \phi(X) \alpha$. 
Advantages

- Preserves local structure and is thus different from other linear techniques.
- LPP is linear and therefore is computationally more tractable than nonlinear techniques like Eigenmaps.
- LPP is defined everywhere.
- LPP may be conducted in the original space or in the reproducing kernel Hilbert space.
LPP is derived by preserving local information, hence it is less sensitive to outliers than PCA.

LPP can have more discriminating power than PCA.
Synthetic Example 2
MNIST dataset

Laplacian Eigenmap

PCA

Locality Preserving Projection