Introduction to Manifold Learning

A Geometry View on Machine Learning

Xiaoyu Xue

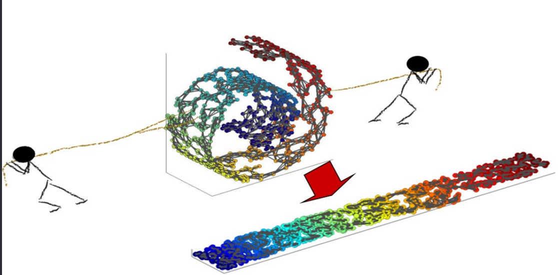
Manifold Learning

1. Manifold

2. Manifold based Dimensionality Reduction
   - 2.1 PCA
   - 2.2 MDS
   - 2.3 ISOMAP
   - 2.4 LLE
   - 2.5 LE
   - 2.6 Vector Field based Dimensionality Reduction
   - 2.7 Manifold Regularization : Semi-Supervised Setting

Manifold Learning

- The data space may not be a Euclidean space, but a nonlinear manifold
- Unfold a manifold, and preserve the geometry structure.
- Euclidean distance $\Longrightarrow$ geodesic distance



Manifold Learning

Find a Euclidean embedding, and then perform traditional learning algorithms in the Euclidean space.

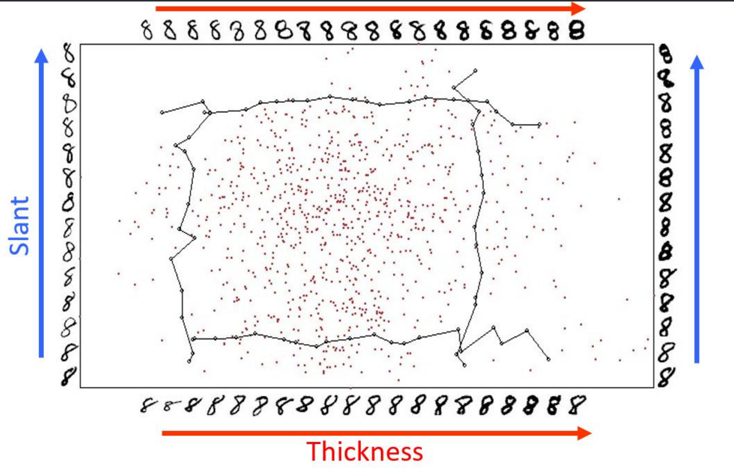
Definition of Manifold Learning

Given data points $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathscr{M} \subset \mathbb{R}^n$, try to find a map $f : \mathscr{M} \to \mathbb{R}^d$, $d \ll n$, where $f = (f_1, \dots, f_n)$, $f_i : \mathscr{M} \to \mathbb{R}$

- The manifold is unknown! We have only samples!
- How to compute the distance on $\mathscr{M}$?
- How to find the mapping function f

Manifold of Face Images

Manifold of Handwritten Digits



Manifold Learning

1. Manifold

2. Manifold based Dimensionality Reduction

   2.1 PCA

   2.2 MDS

   2.3 ISOMAP

   2.4 LLE

   2.5 LE

   2.6 Vector Field based Dimensionality Reduction

   2.7 Manifold Regularization : Semi-Supervised Setting

PCA: Traditional Dimensionality Reduction Method

Principal Component Analysis using linear projection to project data to some directions which have maximum variances

$$\begin{aligned}
\mathbf{p}_{opt} &= \arg\max_{\mathbf{p}} \sum_{i=1}^{m} (y_i - \bar{y})^2 \\
&= \arg\max_{\mathbf{p}} \mathbf{p}^T C \mathbf{p} \\
& \quad s.t. \mathbf{p}^T \mathbf{p} = 1
\end{aligned}$$

- If the manifold is linear, PCA can find the optimal result
- PCA can not process nonlinear manifold

MDS and ISOMAP

Multidimensional scaling tries to preserve the Euclidean distances

$$\Delta := \begin{pmatrix} \delta_{1,1} & \delta_{1,2} & \dots & \delta_{1,m} \\ \delta_{2,1} & \delta_{2,2} & \dots & \delta_{2,m} \\ \vdots & \vdots & & \vdots \\ \delta_{m,1} & \delta_{m,2} & \dots & \delta m, m \end{pmatrix}$$

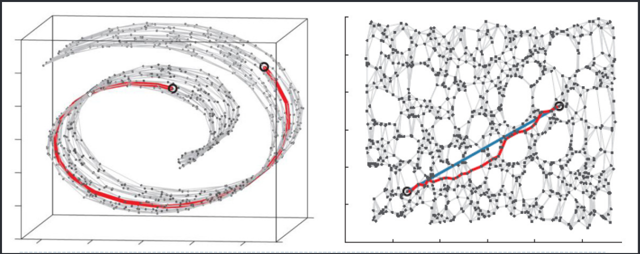
The δ is the Euclidean distance of every two points $\delta_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||$

$$\min_{\mathbf{y}_1,\dots,\mathbf{y}_m} \sum_{i<j} \left(||\mathbf{y}_i - \mathbf{y}_j|| - \delta_{i,j} \right)^2, \quad \dim(\mathbf{y}_i) \ll \dim(\mathbf{x}_i)$$

MDS and ISOMAP

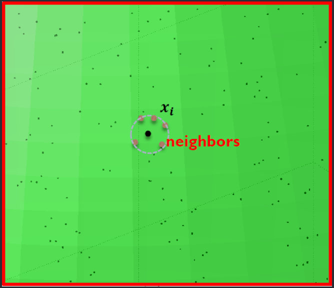
ISOMAP tries to keep the geodesic distances instead of the Euclidean distances.

- How to evaluate the geodesic distances with limited samples?
- Construct the adjacency Graph, and calculate the shortest distances (Dijstra or Floyd algorithm)



Local Linear Embedding

Local Linear Embedding(2000 Science) is another famous manifold learning method. It tries to preserve the local linear relationship.

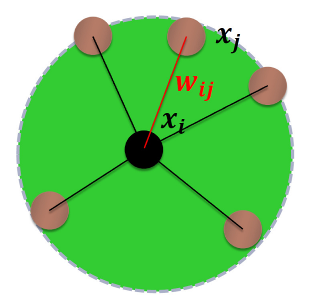


$$\min \epsilon(W) = \min \sum_i \|\mathbf{x}_i - \sum_j W_{ij}\mathbf{x}_j\|$$

$$s.t. \sum_j W_{ij} = 1$$

Local Linear Embedding

Local Linear Embedding(2000 Science) is another famous manifold learning method. It tries to preserve the local linear relationship.



$$\min \Phi(\mathbf{y}) = \min \sum_i \|\mathbf{y}_i - \sum_j W_{ij}\mathbf{y}_j\|^2$$

Laplace Eigen Map

In Laplace Eigen Map, a conclusion has been proofed:

$$|f(\mathbf{z}) - f(\mathbf{x})| < \|\nabla f(\mathbf{x})\| \cdot \|\mathbf{z} - \mathbf{x}\| + o(\|\mathbf{z} - \mathbf{x}\|)$$

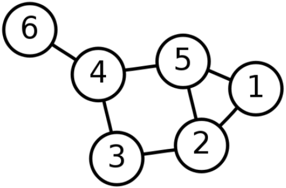
- If $\mathbf{x}_i$ and $\mathbf{x}_j$ are close to each other and the gradient of map f is small, we can sure that $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ preserve local structure.

Construct Laplace matrix and get object function

$$\min \sum_{i,j} (y_i - y_j)^2 W_{ij} \Rightarrow L\mathbf{y} = \lambda D\mathbf{y}$$

Laplace Operator on Graph

- L is the Laplace operator, which measures the smooth of the function on manifold
- On Graph, L is Laplace matrix



$$
\begin{array}{c c}
 & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} &
\begin{pmatrix}
2 & -1 & 0 & 0 & -1 & 0 \\
-1 & 3 & -1 & 0 & -1 & 0 \\
0 & -1 & 2 & -1 & 0 & 0 \\
0 & 0 & -1 & 3 & -1 & -1 \\
-1 & -1 & 0 & -1 & 3 & 0 \\
0 & 0 & 0 & -1 & 0 & 1
\end{pmatrix}
\end{array}
$$

Global vs Local

- Global method : ISOMAP
- Local method : LLE, LE
- Global method can keep more informations of data
- But the amount of computation of Global methods is huge

Out of sample problem

- LE and LLE can not applied new samples.
- Use linear projection: $y = \mathbf{p}^T\mathbf{x}$
- LE$\longrightarrow$ LPP : $XLX^T\mathbf{p} = \lambda D\mathbf{p}$
- LLE$\longrightarrow$ NPE : $XMX^T\mathbf{p} = \lambda XX^T\mathbf{p}$

Vector Field based Dimensionality Reduction

Manifold Regularization

- Measured (labeled) points: discriminant structure

$$\min \sum_{i=1}^{k} (y_i - f(\mathbf{z}_i))^2$$

- Unmeasured (unlabeled) points: geometrical structure

$$\min \sum_{i,j} \big(f(\mathbf{x}_i) - f(\mathbf{x}_j)\big)^2 S_{ij}$$

$$\min_f \sum_{i=1}^{k} (y_i - f(\mathbf{z}_i))^2 + \frac{\lambda}{2} \sum_{i,j}^{m} \big(f(\mathbf{x}_i) - f(\mathbf{x}_j)\big)^2 S_{ij}$$

Laplacian Regularized Least Square

- Linear objective function

$$\min_{\mathbf{w}} \sum_{i=1}^{k} \left(y_i - \mathbf{w}^T \mathbf{z}_i\right) + \frac{\lambda_1}{2} \sum_{i,j=1}^{m} \left(\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j\right)^2 S_{ij} + \lambda_2 \|\mathbf{w}\|$$

- Solution

$$\mathbf{w} = \left(ZZ^T + \lambda_1 XLX^T + \lambda_2 I\right)^{-1} Z\mathbf{y}$$

- $Z = (\mathbf{z}_1, \ldots, \mathbf{z}_k)$: labeled points
- $X = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$: all points