Introduction to Manifold Learning

A Geometry View on Machine Learning

Xiaoyu Xue

Manifold Learning

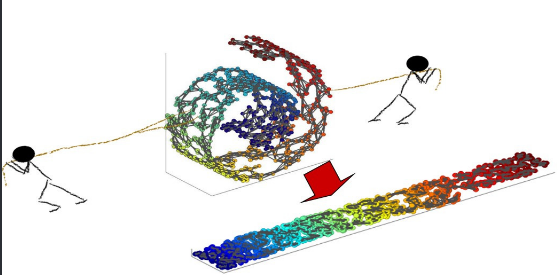
1. Manifold

2. Manifold based Dimensionality Reduction
    2.1 Principal Component Analysis
    2.2 Multidimensional scaling
    2.3 ISOMAP
    2.4 Local Linear Embedding
    2.5 Laplacian Eigenmaps
    2.6 Vector Field based Dimensionality Reduction

3. 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.

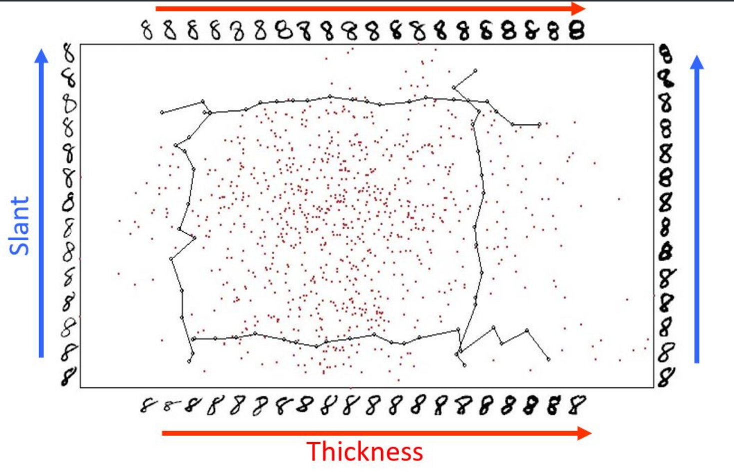
Definition of Manifold Learning

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

- The manifold is unknown! We have only samples!
- How to compute the distance on $\mathcal{M}$?
- 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 Principal Component Analysis
    2.2 Multidimensional scaling
    2.3 ISOMAP
    2.4 Local Linear Embedding
    2.5 Laplacian Eigenmaps
    2.6 Vector Field based Dimensionality Reduction

3. 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

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

- 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

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

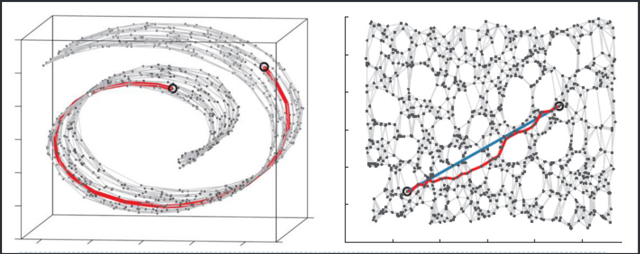
The δ is the Euclidean distance of every two points $\delta_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||$

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

MDS and ISOMAP

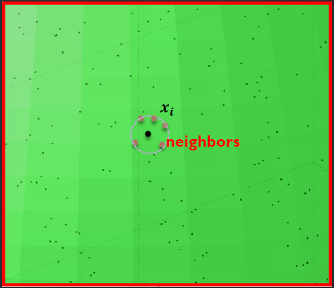
ISOMAP tries to keep the geodesic distances instead of the Euclidean distances.

- 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.

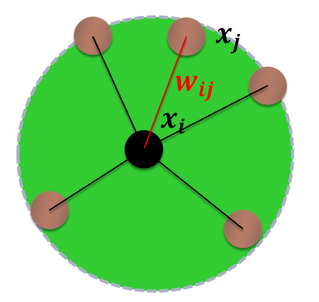


$$\min \epsilon(W) = \min \sum_i \|\mathbf{x}_i - \sum_j W_{ij}\mathbf{x}_j\|$$

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

Local Linear Embedding

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



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

Laplacian Eigenmaps

In Laplacian Eigenmaps, a conclusion has been proofed:

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

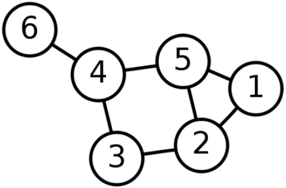
- If $\mathbf{x}_i$ and $\mathbf{x}_j$ are close to each other and the gradient of map f is small, we can sure that $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ preserve local structure.

Construct Laplace matrix and get object function

$$\min_{\|f\|_{L^2(\mathcal{M})}=1} \int_{\mathcal{M}} \|\nabla f(\mathbf{x})\|^2 \Rightarrow \min \sum_{i,j} (y_i - y_j)^2 W_{ij} \Rightarrow L\mathbf{y} = \lambda D\mathbf{y}$$

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 large amount of computation

Out of sample problem

- LE and LLE can not applied on new samples.
- Use linear projection: $y = \mathbf{p}^T \mathbf{x}$
- LE$\Rightarrow$ LPP : $XLX^T\mathbf{p} = \lambda XDX^T\mathbf{p}$
- LLE$\Rightarrow$ NPE : $XMX^T\mathbf{p} = \lambda XX^T\mathbf{p}$
- Projection Matrix $P = [\,\mathbf{p}_1, \ldots, \mathbf{p}_d\,]$
- Given a new sample $\mathbf{y}$, we can project it to low-dim space $\mathbf{z} = P^T\mathbf{y}$

Vector Field based Dimensionality Reduction

Parallel Vector Field Embedding

The goal to find a map $F = (f_1, \ldots, f_d) : \mathcal{M} \subset \mathbb{R}^n \to \mathbb{R}^d$

- F is a local isometry
- dF is orthonormal
- df_i is parallel vector fields $\Rightarrow \nabla df_i = 0$

Steps:

1. $E(V) = \int_{\mathcal{M}} \|\nabla V\|^2 dx, \quad s.t. \int_{\mathcal{M}} \|V\|^2 = 1$
2. $\min \Phi(f) = \int_{\mathcal{M}} \|\nabla f - V\|^2 dx$

A new perspective for manifold learning

Swiss Roll

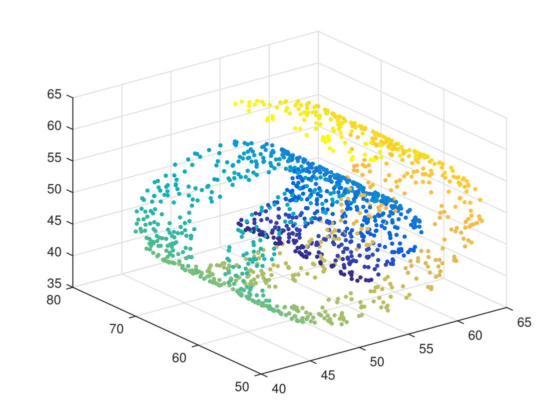


Figure: Swiss Roll

PFE

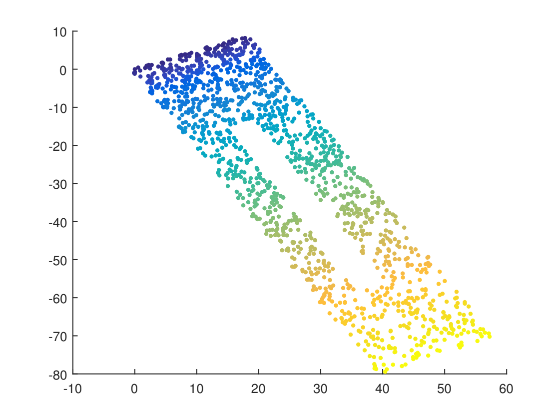


Figure: PFE

PCA

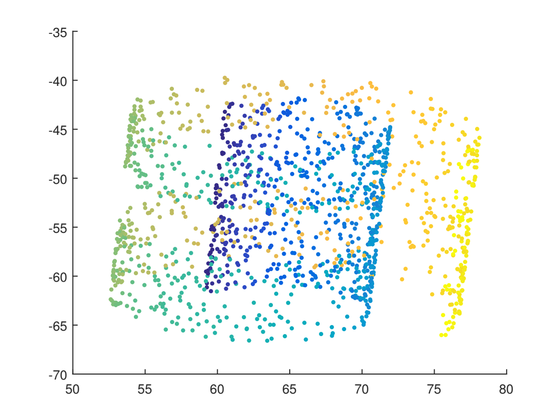


Figure: PCA

ISOMAP

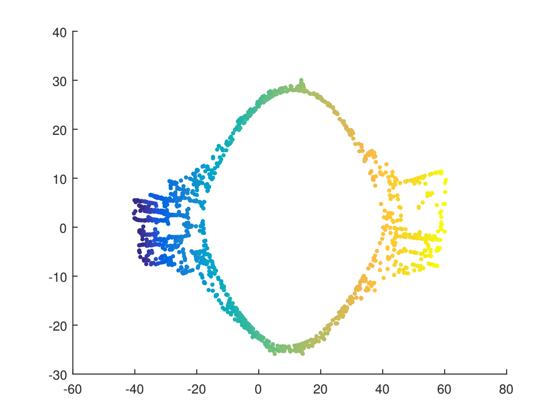


Figure: ISOMAP

LLE

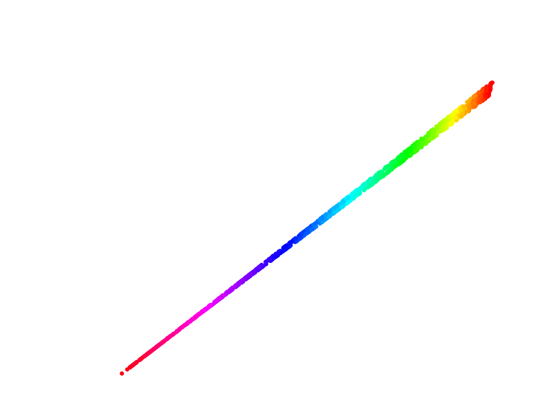


Figure: LLE

LE

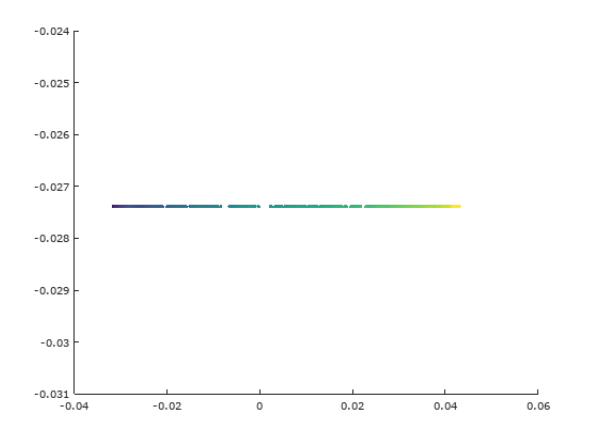


Figure: LE

NPE

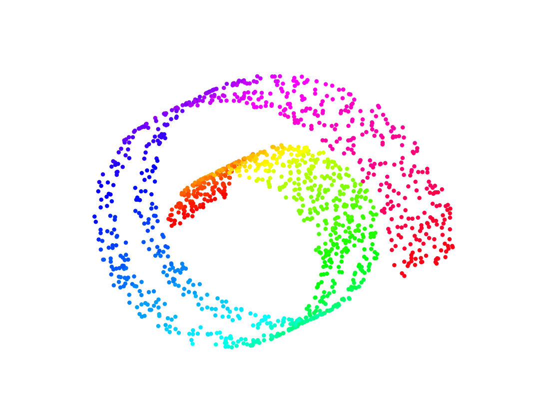


Figure: NPE

LPP

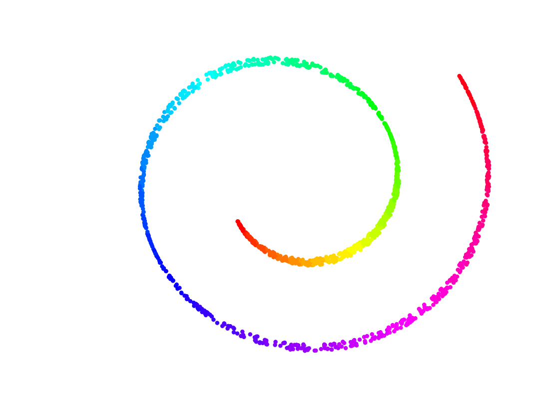


Figure: LPP

Manifold Learning

1. Manifold

2. Manifold based Dimensionality Reduction
   - 2.1 Principal Component Analysis
   - 2.2 Multidimensional scaling
   - 2.3 ISOMAP
   - 2.4 Local Linear Embedding
   - 2.5 Laplacian Eigenmaps
   - 2.6 Vector Field based Dimensionality Reduction

3. Manifold Regularization : Semi-Supervised Setting

Manifold Regularization

- Measured (labeled) points: discriminant structure

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

- Unmeasured (unlabeled) points: geometrical structure

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

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

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

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

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

Parallel Field Regularization

$$\arg\min_{f,V} E(f,V) = \frac{1}{m}\sum_{i=1}^{m} R_0(\mathbf{x}_i, y_i, f) + \lambda_1 R_1(f,V) + \lambda_2 R_2(V)$$

Where

- $R_1(f,V) = \int_M \|\nabla f - V\|^2, \qquad R_2(V) = \int_M \|\nabla V\|^2$
- Low cost of estimation.
- Insensitive to noise.

Resources

Conference:

- Computer Vision: CVPR, ICCV, ECCV
- Machine Learning: NIPS, ICML, IJCAI, AAAI
- NLP : ACL

Journal:

- IEEE Trans on Pattern Analysis and Machine Intelligence(TPAMI)
- Journal of Machine Learning Research(JMLR)
- International Journal of Computer Vision(IJCV)

Tools:

- Scikit-learn: http://scikit-learn.org/
- OpenCV: https://opencv.org/

Thanks !

Q&A