Unsupervised Learning: Neighbor Embedding
Manifold Learning

Suitable for clustering or following supervised learning
Locally Linear Embedding (LLE)

\( w_{ij} \) represents the relation between \( x^i \) and \( x^j \)

Find a set of \( w_{ij} \) minimizing

\[
\sum_i \left\| x^i - \sum_j w_{ij} x^j \right\|_2
\]

Then find the dimension reduction results \( z^i \) and \( z^j \) based on \( w_{ij} \)
LLE

\[ z^i, z^j \]

\[ w_{ij} \]

\[ x^i, x^j \]

Source of image:
http://feetsprint.blogspot.tw/2016/02/blog-post_29.html
LLE

Find a set of $z^i$ minimizing

$$\sum_i \left\| z^i - \sum_j w_{ij} z^j \right\|_2$$

Keep $w_{ij}$ unchanged

Original Space

New (Low-dim) Space
Laplacian Eigenmaps

- Graph-based approach

Distance defined by graph approximate the distance on manifold

Construct the data points as a graph
Laplacian Eigenmaps

- **Review in semi-supervised learning**: If $x^1$ and $x^2$ are close in a high density region, $\hat{y}^1$ and $\hat{y}^2$ are probably the same.

\[
L = \sum_{x^r} C(y^r, \hat{y}^r) + \lambda S
\]

As a regularization term

\[
S = \frac{1}{2} \sum_{i,j} w_{i,j} (y^i - y^j)^2 = y^T Ly
\]

$L$: $(R+U) \times (R+U)$ matrix

Graph Laplacian

\[
L = D - W
\]
Laplacian Eigenmaps

- *Dimension Reduction*: If $x^1$ and $x^2$ are close in a high density region, $z^1$ and $z^2$ are close to each other.

$$S = \frac{1}{2} \sum_{i,j} w_{i,j} (z^i - z^j)^2$$

Any problem? How about $z^i = z^j = 0$?

Giving some constraints to $z$:

- If the dim of $z$ is $M$, $\text{Span}\{z^1, z^2, \ldots, z^N\} = \mathbb{R}^M$

*Spectral clustering*: clustering on $z$

T-distributed Stochastic Neighbor Embedding (t-SNE)

• Problem of the previous approaches
  • Similar data are close, but different data may collapse

LLE on MNIST  
LLE on COIL-20
t-SNE

Compute similarity between all pairs of \( x \): \( S(x^i, x^j) \)

\[
P(x^j | x^i) = \frac{S(x^i, x^j)}{\sum_{k \neq i} S(x^i, x^k)}
\]

Compute similarity between all pairs of \( z \): \( S'(z^i, z^j) \)

\[
Q(z^j | z^i) = \frac{S'(z^i, z^j)}{\sum_{k \neq i} S'(z^i, z^k)}
\]

Find a set of \( z \) making the two distributions as close as possible

\[
L = \sum_i KL \left( P(\ast | x^i) \| Q(\ast | z^i) \right)
\]

\[
= \sum_i \sum_j P(x^j | x^i) \log \frac{P(x^j | x^i)}{Q(z^j | z^i)}
\]
t-SNE – Similarity Measure

\[ S(x^i, x^j) \]
\[ = \exp \left( -\|x^i - x^j\|_2^2 \right) \]

SNE:
\[ S'(z^i, z^j) = \exp \left( -\|z^i - z^j\|_2 \right) \]

t-SNE:
\[ S'(z^i, z^j) = 1/1 + \|z^i - z^j\|_2 \]

Ignore \( \sigma \) for simplicity
t-SNE

- Good at visualization

- t-SNE on MNIST
- t-SNE on COIL-20
To learn more ...

• Locally Linear Embedding (LLE): [Alpaydin, Chapter 6.11]
• Laplacian Eigenmaps: [Alpaydin, Chapter 6.12]
• t-SNE
  • Excellent tutorial: https://github.com/oreilлимedia/t-SNE-tutorial