Manifold Learning Small Talk

Po-Yi Tsai[1]

[1]EECS Undergraduate Honors Program
National Chiao Tung University
Hsinchu City, Taiwan

20 July 2020

Almost Completed

- Learn the basics of diffusion map
- Try on simple examples

Basic Idea

1. Define random walk on data based on local geometry.
2. Map points by eigendecomposing transition matrix.

Random Walk on Data

Define kernel $k(x, y)$. Usually

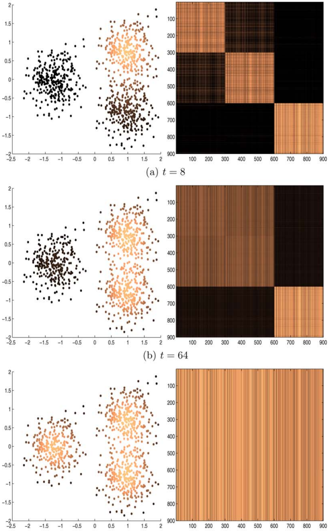
$$k(x, y) = e^{-\frac{\|x-y\|^2}{\alpha}}$$

Define $p(x, y)$ by

$$p(x, y) = \frac{k(x, y)}{\sum_x k(x, y)}$$

Then $P_{ij} = p(x_i, x_j)$ is a transition matrix of a Markov chain.

Random Walk on Data



(a) $t = 8$

(b) $t = 64$

(c) $t = 1024$

Diffusion Maps

Let

$$P = \sum_i \lambda_i \vec{v}_i \vec{e}_i^T$$

be the eigendecomposition of P. Define

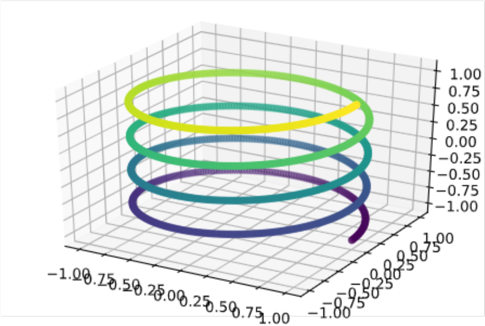
$$y_i = \begin{pmatrix} \lambda_1^t \vec{v}_1(i) \\ \lambda_2^t \vec{v}_2(i) \\ \vdots \\ \lambda_n^t \vec{v}_n(i) \end{pmatrix}$$

Diffusion Distance

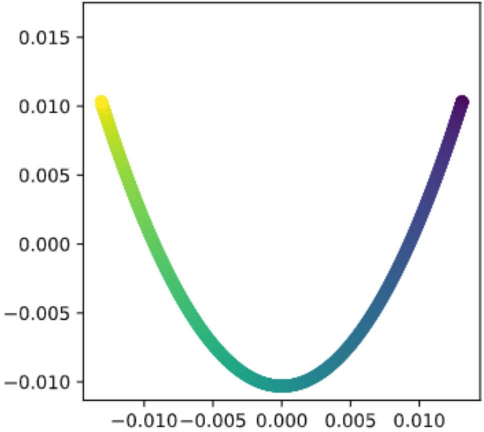
$$D_t(x, y) = \|p(x, -) - p(y, -)\|_{l^2(\mathbb{R}^n, D^{-1})}$$

The diffusion distance in the data space equals the Euclidean distance in the diffusion space.

Example: Helix



Example: Helix



Potential Future Directions

- Application
  - MNIST dataset (clustering / classification)
- Theoretical
  - Combine multi-scale geometry.
  - Find the actual lower-dimensional manifold in which the data are embedded.