30.11.2018 FBDA

4.3. Diffusion Maps

→ Non-linear dimensionality Reduction

→ Coifman-Lafon 2006

Goal: represent data in a lower dimensional space while preserving the geometry

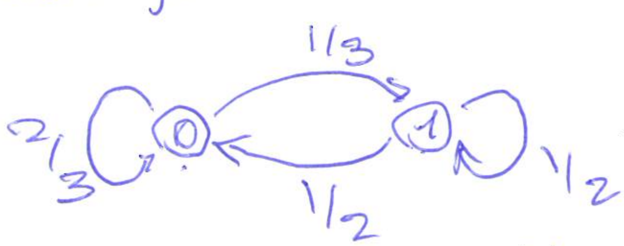
Main steps:

1. Construct a weighted graph $G(V, E, W)$ on the data

2) Define a homogeneous random walk on the graph determined by a transition matrix.

X_t random process

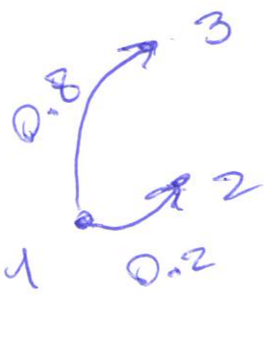
homogen.



→ transition matrix

3) Perform a non-linear embedding of the points

example:



$x_1, \ldots, x_n \in \mathbb{R}^p$, n samples

$G(V, E, W)$

1) $\underline{v_i} \longrightarrow x_i$

The weight of an edge : between x_i and x_j

using the weigh function or Kernel $K(x_i, x_j)$

Important: different from the notion of Kernel in SVM.

properties of $K(x_i, x_j)$

- $K(x_i, x_j) = K(x_j, x_i)$ symmetry
- Non-negativity $K(x_i, x_j) \geq 0$
- Locality if $\|x_i - x_j\| \ll \varepsilon$ then $K(x_i, x_j) \longrightarrow 1$

 and if $\|x_i - x_j\| \gg \varepsilon$ then $K(x_i, x_j) \longrightarrow 0$

* Pay attention to ε

examples: Gaussian Kernel

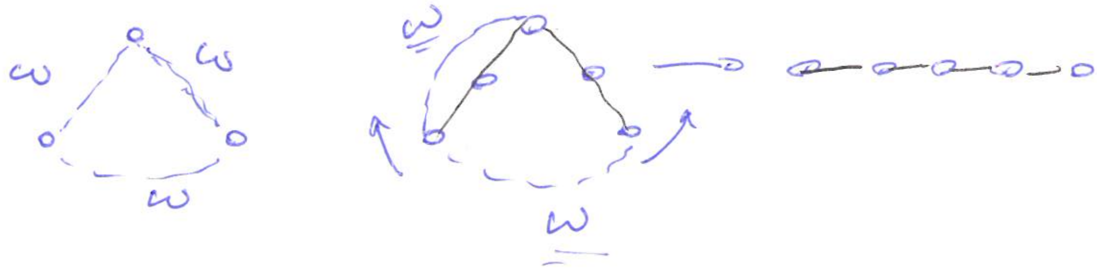
$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\varepsilon^2}\right)$$

(Ex. verify the above properties)

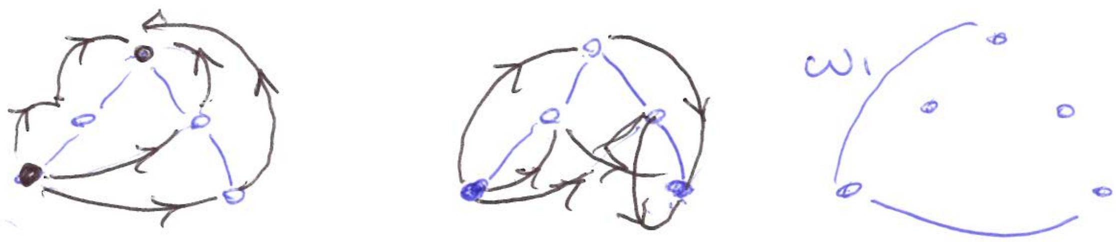
Example: $K(x_i, x_j) = \begin{cases} 1 & \text{if } \|x_i - x_j\| \le \varepsilon \\ 0 & \text{otherwise} \end{cases}$

(how to choose ε? ISOMAP)

$\underline{v_i} \longrightarrow \underline{x_i} \qquad w_{ij} = K(x_i, x_j)$



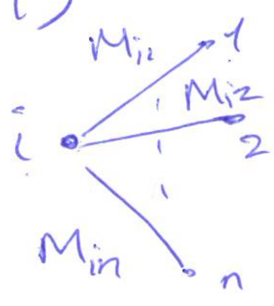
Two steps



$w_1 > w_2$

$M = (M_{ij})_{i,j=1,\dots,n}$ transition matrix.

$$M_{ij} = P(X_{t+1} = j \mid X_t = i)$$



$$M_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$$

$$\sum_j M_{ij} = 1, \quad \sum_j w_{ij} \neq 1$$

$$\boxed{M_{ij} = \frac{w_{ij}}{\deg(i)}} \qquad \deg(i) = \sum_j w_{ij}$$

$$\Rightarrow \quad M = D^{-1} W \qquad W = (w_{ij})_{1 \le i,j \le n}$$

$$D = \operatorname{diag}(\deg(1), \deg(2), \dots, \deg(n))$$

$$D = \begin{pmatrix} \deg(1) & & & 0 \\ & \deg(2) & & \\ & & \ddots & \\ 0 & & & \deg(n) \end{pmatrix}$$

$*$ $\mathbb{P}(X_t = j \mid X_0 = i) = (M^t)_{ij}$ $\quad j = 1, \dots, n$

$t \in \mathbb{N}$ (Ex.-verify)

$M = (M_{ij})$ $\qquad M_{ij} = \mathbb{P}(X_1 = j \mid X_0 = i)$

$M = D^{-1} W$

distances between $\underline{i}$ and other points

$((M^t)_{i1}, M^t_{i2}, \dots, M^t_{in}) \longrightarrow$ ith row of M^t

$\Rightarrow e_i^T M^t$ $\qquad e_i$ canonical basis vectors

$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \rightarrow$ position $\underline{i}$

$\cdot\ v_i \longrightarrow e_i^T M^t$

$\rightarrow M^t$ is not symmetric $\Rightarrow$ MDS does not work

Spectral decomposition:

$M = D^{-1} W$ $\qquad W$ symm. D^{-1} sym.

$S = D^{1/2} M D^{-1/2} = D^{-1/2} W D^{-1/2} \Rightarrow S$ $\underline{\text{sym.}}$

$M = D^{-1} W \qquad S = D^{-1/2} W D^{-1/2} = \underline{V \Lambda V^T}$

$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots)$ eigenvalue

$V := (v_1 \dots v_n)$ eigenvectors.

$S = D^{1/2} M D^{-1/2} \Rightarrow M = D^{-1/2} S D^{1/2}$

$\Rightarrow \quad M = D^{-1/2} V \Lambda V^T D^{1/2}$

$= \Phi \Lambda \Psi^T$ — $\Phi = D^{-1/2} V$, $\Psi = D^{1/2} V$

Φ and Ψ are bi-orthogonal:

$\Rightarrow \Phi^T \Psi = I$ (because V is an orth. matrix.)

$$\left.\begin{aligned} M\phi_k &= \lambda_k \phi_k \\ \psi_k^T M &= \lambda_k \psi_k^T \\ \phi_i^T \psi_j &= \delta_{ij} \end{aligned}\right\} \longrightarrow \text{Ex. verify!}$$

$\Phi = (\phi_1 \dots \phi_n)$

$\Psi = (\psi_1 \dots \psi_n)$

$\Rightarrow M$ can be written as: $\boxed{M = \sum_{k=1}^{n} \lambda_k \phi_k \psi_k^T}$

$M^t = ? \Rightarrow \boxed{M^t = \sum_{k=1}^{n} \lambda_k^t \phi_k \psi_k^T}$

$v_i \longrightarrow e_i^T M^t = \sum_{k=1}^{n} \lambda_k^t \underbrace{e_i^T \phi_k}\, \psi_k^T \Rightarrow$

$$e_i^T M^t = \sum_{k=1}^{n} \underbrace{\lambda_k^t \phi_{k,i}}_{} \psi_k^T$$

$$\phi_k = (\phi_{k,1}, \phi_{k,2}, \dots, \phi_{k,n})^T$$

$$\phi_{k,i} \in \mathbb{R}$$

Definition 4.5. The <u>diffusion</u> map at step time t is defined as

$$\phi_t(v_i) = \begin{pmatrix} \lambda_1^t \phi_{1,i} \\ \vdots \\ \lambda_n^t \phi_{n,i} \end{pmatrix} \quad i = 1, \dots, n$$

Thm. 4.6. The eigenvalues $\lambda_1, \dots, \lambda_n$ of M satisfy $|\lambda_k| \leq 1$. It also holds that $M \mathbb{1}_n = \mathbb{1}_n$ and $\underline{1}$ is an eigenvalue of M.

$$\phi_t(v_i) = \begin{pmatrix} 1 \\ \lambda_2^t \phi_{2,i} \\ \lambda_n^t \phi_{n,i} \end{pmatrix}$$

$$* \sum_j M_{ij} = 1 \Rightarrow M \mathbb{1}_n = \mathbb{1}_n$$

- $$\Phi_t(v_i) = \begin{pmatrix} \lambda_2^t \phi_{2i} \\ \vdots \\ \lambda_n^t \phi_{n,i} \end{pmatrix}$$

- It is possible to have more than eigenvalues equal to one. $\Rightarrow$ graph disconnected or bipartite

Definition 4.7. The diffusion map truncated to d dimensions is defined as.

$$\Rightarrow \quad \Phi_t^{(d)}(v_i) = \begin{pmatrix} \lambda_2^t \phi_{2,i} \\ \vdots \\ \lambda_{d+1}^t \phi_{d+1,i} \end{pmatrix} \quad *$$

$\Phi_t^{(d)}(v_i)$ an approximate embedding of $v_1, \dots, v_n$ in a d-dimensial space.

Theorem. 4.8 For any pair of nodes v_i and v_j we have

$$\| \Phi_t(v_i) - \Phi_t(v_j) \|^2 = \sum_{l=1}^{n} \frac{1}{\deg(l)} \left(P(X_t = l \mid X_0 = i) - P(X_t = l \mid X_0 = j) \right)^2$$

Proof. <u>Exercise</u>