10/12 Spectral Clustering (cont.) 4

Consistency of spectral clustering

ref (LBB'08)

Suppose $X_i \sim P$ where P is some distribution on $\Omega \subset \mathbb{R}^D$

W_{ij} is the affinity matrix
$= k(x_i, x_j)$ ex) $W_{ij} = e^{\frac{-|x_i - x_j|^2}{\Sigma}}$, $\Sigma > 0$

As $n \to \infty$, we want to show the convergence of graph Laplacian $\mathcal{L}$.

1. $\mathcal{L}_n = D - W$ where $D_{ij} = \sum_{j=1}^{n} W_{ij}$
   ($\mathcal{L}_n = \mathcal{L}_{un}$)
   $\longrightarrow U$
2. $\mathcal{L}_n' = D^{-1/2}(D-W)D^{-1/2}$
   $\longrightarrow U'$
   ($\mathcal{L}_n' = \mathcal{L}_{sym}$)
3. $\mathcal{L}_n'' = D^{-1}(D-W) = I - P$
   ($\mathcal{L}_n'' = \mathcal{L}_{rw}$)

$U : C(\Omega) \to C(\Omega)$.

$Uf(x) = f(x)d(x) - \int k(x,y) f(y) dP(y)$

where $dP(x) = p(x)dx$
$d(x) = \int k(x,y) dP(y)$
and x is any point in Ω.

$$U'f(x) = f(x) - \int \frac{k(x,y)}{\sqrt{d(x)}\sqrt{d(y)}} f(y) dP(y)$$

proof aside

$$(D^{-1/2} W D^{-1/2})_{ij} = \frac{1}{\sqrt{D_{ii}}} \cdot W_{ij} \cdot \frac{1}{\sqrt{D_{jj}}}$$

$$= \frac{\frac{1}{n} k(x_i, x_j)}{\sqrt{\frac{1}{n}\sum_{j'} k(x_i, x_{j'})}\sqrt{\frac{1}{n}\sum_{j'} k(x_j, x_{j'})}}$$

$$\approx \frac{\frac{1}{n} k(x_i, x_j)}{\sqrt{d(x_i)}\sqrt{d(x_j)}}$$

"first-r spectral convergence":

M_n $n \times n$, $T: C(\Omega) \to C(\Omega)$

M_n first-r spectral convergence to T if first r eigenvalues of M_n converge to those of T, and the associated eigenvectors converge to the eigenfunctions of T. the first smallest r eigenva

---

remark on 3. $\mathcal{L}_n''$

We won't consider ③ because it has a one-to-one relationship with case ② since same eigenvalues

Theorem

For fixed $r>0$ and $n\to\infty$,

Under mild conditions,

2. $\mathcal{L}_n'$ first r spectral converge to the operator U' (the normalized sym version)

1. $\mathcal{L}_n$ first r spectral converge to U if the first r eigenvalues of U lie outside of the range of the degree function $d(x)$. (the unnormalized)

need extra constraints for convergence since U might coincide with the range of $d(x)$

Proof case ②

goal: Have M_n converge to T

where $M_n : \mathbb{R}^n \to \mathbb{R}^n$

$T : C(\Omega) \to C(\Omega)$

$M_n = I - D^{-1/2} W D^{-1/2}$

$T = U'$

$Tf(x) = f(x) - \int h(x,y) f(y)\, dP(y)$

where $h(x,y) = \dfrac{k(x,y)}{\sqrt{d(x)}\sqrt{d(y)}}$

$= f(x) - \int h(x,y) f(y)\, dP_n(y)$

where $dP_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}(x)\, dx$

$= f(x) - \frac{1}{n}\sum_{i=1}^{n} h(x, x_i) f(x_i)$

as long as $M_n \underset{(\text{lemma1})}{\sim} T_n \underset{(\text{lemma2})}{\sim} T$

$M_n = I - D^{-1/2} W D^{-1/2}$

Lemma 1 (spectral equivalence btwn M_n and T_n)

(i) If $T_n \psi = \lambda\psi$, let $v \in \mathbb{R}^n$, $i=1,\dots,n$

$v_i = \psi(x_i)$, then $M_n v = \lambda v$

(ii) If $M_n v = \lambda v$ and $\lambda \neq 1$, then let

$\psi(x) = \dfrac{\frac{1}{n}\sum h(x,x_j) v_j}{1-\lambda}$ and so

then $T_n\psi = \lambda\psi$

Lemma 2 (replacing $dP_n(y)$ to be $dP(y)$)

T_n first r spectral converge to T.

Proof

$\forall f$, $T_n f \to Tf$ by Law of Large Number

$\|T_n f - Tf\|_\infty \to 0$ simultaneously for "sufficiently many" f such that for each eigenval of T ($\lambda \neq 1$), the associated eigvalue of T_n converge to λ and the associated eigen function of T_n converge

$T_n \psi_n = \lambda_n \psi_n$ so $\lambda_n \to \lambda$ asymptotically.

$T\psi = \lambda\psi$, $\lambda \neq 1$ and $\|\psi_n - \psi\|_\infty \to 0$

Remark (Bochner's Theorem)

W is a PSD where W_{ij} comes from Gaussian kernel $e^{-\frac{|x_i-x_j|^2}{\Sigma}}$

This implies $D^{-1/2}WD^{-1/2}$ is PSD.

$\Rightarrow \mathcal{L}_{sym} = I - D^{-1/2}WD^{-1/2}$,

with eigenvalues on $[0,1]$

eig of $D^{-1/2}WD^{-1/2} \in [0,1)$

Prop p is density in $\mathbb{R}^D$

$$k(x,y) = e^{-\frac{|x-y|^2}{\Sigma}}, \quad \Sigma > 0$$

(i) $k(x,y)$ is PSD kernel, ie $\forall f$ and $\int f(x)^2 p(x)dx < \infty$

$$\int_\Omega\int_\Omega k(x,y)f(x)f(y)\,p(x)dx\,p(y)dy \geq 0$$

(ii) $W_{ij} = (k(x_i,x_j))_{1\leq i,j\leq n}$,

W is PSD ie) $\forall v \in \mathbb{R}^n$,

$v^T W v \geq 0$.

Fourier Transform

Forward $\rightarrow f(x)$, $x \in \mathbb{R}$

$$\hat{f}(\xi) = \int f(x) e^{-i\xi x} dx$$

Inverse $\leftarrow$ $f(x) = \frac{1}{2\pi}\int \hat{f}(\xi)\, e^{i\xi^T x} d\xi$

Proof

$$k(x,y) = \underbrace{g(x-y)}_{} = \int_{\mathbb{R}^D} \hat{g}(\xi) \underbrace{e^{i\xi^T(x-y)}}_{e^{i\xi^T x}e^{-i\xi^T y}} d\xi$$

(Fourier transform)

$$\forall v \in \mathbb{R}^n, \quad v^T W v = \sum_{ij} v_i v_j W_{ij}$$

$$= \sum_{ij} v_i v_j k(x_i, x_j)$$

$$= \sum v_i v_j \int \hat{g}(\xi) e^{i\xi^T x_i} e^{-i\xi^T x_j} d\xi$$

$$= \int_{\mathbb{R}^D} \hat{g}(\xi) \left[\sum_{i=1}^{n}\sum_{j=1}^{n} v_i e^{i\xi^T x_i} v_j e^{-i\xi^T x_j}\right] d\xi$$

$$= \int_{\mathbb{R}^D} \hat{g}(\xi) \left(\sum v_i e^{i\xi^T x_i}\right) \overline{\left(\sum v_j e^{i\xi^T x_j}\right)}$$

$$= \int \hat{g}(\xi) v(\xi) \overline{v(\xi)}\, d\xi$$

$$= \int \hat{g}(\xi) \left| v(\xi) \right|^2 d\xi \quad \geq 0$$

Since $\hat{g}(\xi) = e^{-a|\xi|^2} \Rightarrow \hat{g}(\xi) > 0 \; \forall \xi$