

Math 690: Topics in Data Analysis and Computation

Lecture notes for October 12, 2017

Scribed by Dev Dabke and Andrew Cho

1 Introduction

The lecture covered the following on the **consistency of spectral clustering**:

- Setup of \mathcal{L}_{un} and \mathcal{L}_n and their limit operators U and U'
- First r spectral convergence
- Bochner's Theorem

2 Consistency of Spectral Clustering

2.1 The Problem Setup

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

W_{ij} is the affinity matrix where, as an example,

$$W_{ij} = e^{\frac{-|x_i - x_j|^2}{\varepsilon}}$$

where $\varepsilon > 0$.

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

1. $\mathcal{L}_n = D - W$ where $D_{ij} = \sum_{j=1}^n W_{ij}$ ($\rightarrow U$), i.e. unnormalized
2. $\mathcal{L}'_n = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$ where ($\rightarrow U'$), i.e. symmetric
3. $\mathcal{L}''_n = D^{-1}(D - W) = I - P$, i.e. random walk

2.2 Limit operators U and U'

We construct linear limit operators U and U' on $C(X)$ which are the limit of the discrete operators \mathcal{L}_n and \mathcal{L}'_n . We prove that the first "r" eigenvectors of the discrete operators converge to eigenfunction of the limit operators.

Here, U is defined as follows:

$$U : C(\Omega) \rightarrow C(\Omega)$$
$$Uf(x) = f(x)d(x) - \int k(x, y)f(y)dP(y)$$

where

$$\begin{aligned} dP(x) &= p(x)dx \\ dx &= \int k(x, y)dP(y) \\ x &\in \Omega \end{aligned}$$

U' is derived and proved as follows:

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

Proof.

$$\begin{aligned} (D^{-\frac{1}{2}}WD^{-\frac{1}{2}})_{ij} &= \frac{1}{\sqrt{D_{ii}}}W_{ij}\frac{1}{\sqrt{D_{jj}}} \\ &= \frac{\frac{1}{n}k(x_i, x_j)}{\sqrt{\frac{1}{n}\sum_{j'}k(x_i, x'_{j'})}\sqrt{\frac{1}{n}\sum_{j'}k(x_j, x'_{j'})}} \\ &\approx \frac{\frac{1}{n}k(x_i, x_j)}{\sqrt{d(x_i)}\sqrt{d(x_j)}} \end{aligned}$$

□

2.3 First r spectral convergence

Let's discuss what "first r spectral convergence means. 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 eigenvalues.

Theorem. For fixed $r > 0$, $n \rightarrow \infty$, and mild conditions,

1. (**unnormalized**) \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 $d(x)$. We need extra constraints for convergence since U might coincide with the range of $d(x)$.
2. (**normalized symmetric**) \mathcal{L}'_n first r spectral converge to the operator U' .

Proof. (Case 2 \mathcal{L}_{sym}) Have M_n converge to T where

$$\begin{aligned} M_n &: \mathbb{R}^n \rightarrow \mathbb{R}^n \\ T &: C(\Omega) \rightarrow C(\Omega) \\ M_n &: I - D^{-\frac{1}{2}}WD^{-\frac{1}{2}} \\ T &= U' \end{aligned}$$

$$\begin{aligned} Tf(x) &= f(x) - \int h(x, y)f(y)dP(y) \\ &= f(x) - \int h(x, y)f(y)dP_n(y) \\ &= f(x) - \frac{1}{n} \sum_{i=1}^n h(x, x_i)f(x_i) \end{aligned}$$

where

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

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

□

Lemma 1 (spectral equivalence between M_n and T_n)

1. If $T_n \varphi = \lambda \varphi$, let $v \in \mathbb{R}^n, i = 1, \dots, n, v_i = \varphi(x_i)$, then $M_n v = \lambda v$.
2. If $M_n v = \lambda v$ and $\lambda \neq 1$, then let $\varphi = \frac{\frac{1}{n} \sum h(x, x_j) v_j}{1 - \lambda}$ and so then $T_n \varphi = \lambda \varphi$.

Lemma 2 (replacing $dP_n(y)$ to be $dP(y)$, T_n first spectral converge to T)

Proof. For all f , $T_n \rightarrow T f$ by Law of Large Numbers. $\|T_n f - T f\|_{\inf} \rightarrow 0$ simultaneously for "sufficiently many" f such that for each eigenvalue of T ($\lambda \neq 1$), the associated eigenvalue of T_n converge to λ and the associated eigenfunction of T_n converges. In other words, $T_n \varphi_n = \lambda_n \varphi_n$ so $\lambda_n \rightarrow$ asymptotically $T \varphi = \lambda \varphi$ and $\|\varphi_n - \varphi\|_{\infty} \rightarrow 0$ □

Remark (Bochner's Theorem)

W is a PSD where W_{ij} comes from Gaussian kernel $\frac{e^{-|x_i - x_j|^2}}{\epsilon}$. This implies $D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$ is PSD. Thus, we have $\mathcal{L} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$, with eigenvalues on $[0, 1]$ since eigenvalues of $D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \in [0, 1]$

Proposition. P is a density in \mathbb{R}^D . We have the usual Gaussian kernel $\frac{e^{-|x_i - x_j|^2}}{\epsilon}$ where $\epsilon > 0$.

1. $k(x, y)$ is PSD kernel, ie. $\forall f$ and $\int f(x)^2 p(x) dx < \infty$, $\int_{\Omega} \int_{\Omega} k(x, y) f(x) f(y) p(x) dx p(y) dy \geq 0$.
2. $W_{ij} = (k(x_i, x_j))_{1 \leq i, j \leq n}$. We know W is PSD, so $\forall v \in \mathbb{R}^n, v^T W v \geq 0$.

Proof. By Fourier Transform,

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

$$\begin{aligned} \forall v \in \mathbb{R}^n, v^T W v &= \sum_{ij} v_i v_j W_{ij} \\ &= \sum_{ij} v_i v_j k(x_i, x_j) \\ &= \sum_{ij} v_i v_j \int \hat{g}(\xi) e^{i\xi^T x_i} e^{-i\xi^T x_j} d\xi \\ &= \int_{\mathbb{R}} \hat{g}(\xi) \left(\sum_i v_i e^{i\xi^T x_i} \right) \left(\sum_j v_j e^{-i\xi^T x_j} \right) d\xi \\ &= \int_{\mathbb{R}} \hat{g}(\xi) \left(\sum_i v_i e^{i\xi^T x_i} \right) \overline{\left(\sum_j v_j e^{i\xi^T x_j} \right)} d\xi \\ &= \int \hat{g}(\xi) V(\xi) \overline{V(\xi)} d\xi \\ &= \int \hat{g}(\xi) |V(\xi)|^2 d\xi \geq 0 \end{aligned}$$

since $\hat{g}(\xi) = e^{-a|\xi|^2} \implies \hat{g} > 0, \forall \xi$ □