Notes on 'The Geometry of Kernelized Spectral Clustering'

Alden Green

March 25, 2019

1 Analysis of normalized Laplacian embedding.

For a given set of distributions $\mathbb{P}_1, \ldots, \mathbb{P}_m$ and weights $w_1, \ldots, w_K$ in the probability simplex, define the **mixture distribution**

$$\bar{\mathbb{P}} := \sum_{m=1}^{K} w_m \mathbb{P}_m.$$

Given a non-negative, continuous, symmetric kernel function $k(x, y)$ and a distribution $\mathbb{P}$ we introduce the **square root kernelized density** as the function $q \in L^2(\mathbb{P})$ given by

$$q(x) := \sqrt{\int k(x, y) d\mathbb{P}(y)}.$$

In particular, we denote the square root kernelized density of the mixture distribution $\bar{P}$ by $\bar{q}$ and those of the mixture components $\{\mathbb{P}_m\}_{m=1}^K$ by $\{q_m\}_{m=1}^K$.

Because we typically deal with the matrix $L = D^{-1/2} A D^{-1/2}$ when performing spectral embedding, it is useful to define analogous continuum operators. In particular, we define the **normalized kernel function** $\bar{k}$ to be

$$\bar{k}(x, y) = \frac{1}{\bar{q}(x)} k(x, y) \frac{1}{\bar{q}(y)}$$

and the normalized kernel for mixture component k_m to be

$$k_m(x, y) := \frac{k(x, y)}{q_m(x) q_m(y)} \quad \text{for } m = 1, \ldots, K.$$

Now, we introduce the **coupling parameter**

$$\mathcal{C}(\bar{P}) := \max_{m=1,\ldots,K} \left\| k_m - w_m \bar{k} \right\|_{\mathbb{P}_m \bigotimes \mathbb{P}_m}^2$$

This controls the maximum average connection between points generated by any one mixture component $\mathbb{P}_i$ and those generated by a different mixture component $\mathbb{P}_j$.

1.1 Similarity parameter

For any two distributions $\mathbb{P}_\ell$ and $\mathbb{P}_m$, the similarity between the two is given by

$$\mathcal{S}(\mathbb{P}_\ell, \mathbb{P}_m) = \frac{\int_\mathcal{X} \int_\mathcal{X} k(x,y) d\mathbb{P}_\ell(x) d\mathbb{P}_m(y)}{\int_\mathcal{X} \int_\mathcal{X} k(x,y) d\mathbb{P}_\ell(x) d\overline{\mathbb{P}}(y)}$$

2 Population-level analysis

The continuum normalized Laplacian operator $\overline{T} : L^2(\mathbb{P}) \to L^2(\mathbb{P})$ is given by

$$(\overline{T} f)(\cdot) = \int \bar{k}(\cdot, y) f(y) d\overline{\mathbb{P}} y$$

Let

- $\mathcal{R} \subset L^2(\overline{\mathbb{P}})$ be the span of the top K eigenfunctions of the normalized Laplacian operator $\overline{T}$, and
- $\mathcal{Q} = \operatorname{span}\{q_1, \ldots, q_K\} \subset L^2(\overline{\mathbb{P}})$ be the span of the square root kernelized densities.

We would like to show that these subspaces are close, measured according to

$$\rho(\mathcal{Q}, \mathcal{R}) := \|\Pi_\mathcal{Q} - \Pi_\mathcal{R}\|_{HS}$$

where $\Pi_\mathcal{Q}$ and $\Pi_\mathcal{R}$ are the orthogonal projection operators onto $\mathcal{Q}$ and $\mathcal{R}$, respectively. This will happen, so long as the difficulty function φ

$$\varphi(\overline{\mathbb{P}}; K) := \frac{\sqrt{K} \left[\mathcal{S}_{\max}(\overline{\mathbb{P}}) + \mathcal{C}(\overline{\mathbb{P}})\right]^{1/2}}{w_{\min} \Gamma^2_{\min}(\overline{\mathbb{P}})}$$

is small.

Theorem 1. *For any finite mixture $\bar{P}$ with difficulty function bounded as*

$$\varphi(\overline{\mathbb{P}}; K) \lesssim \Gamma^2_{\min}(\overline{\mathbb{P}})$$

the distance between subspaces $\mathcal{Q}$ and $\mathcal{R}$ is bounded as

$$\rho(\mathcal{Q}, \mathcal{R}) \lesssim \varphi(\overline{\mathbb{P}}; K)$$