L4: Kernelized affinity on graphs and manifolds

- graph Laplacian
- knn graph and self-tuned kernel
- heat kernel on manifold
- from graph Laplacian to manifold Laplacian

- Introduction to graph-based geometric data analysis
  - graph Laplacian
  - manifold learning, ISOMAP
  - spectral convergence demo

- kNN bandwidth (self-tuned kernel)

given $\{x_i\}_{i=1}^n$ in $\mathbb{R}^D$, $x_i \sim p$ iid.

affinity matrix

$$W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}}, \quad \sigma > 0$$

Problem: at x_i which is further away from other x_j's, the universal σ may make i isolate.

idea: adaptively set σ_i for each node x_i

k-nearest neighbor (kNN) of x_i in $X = \{x_j\}_{j=1}^n$.

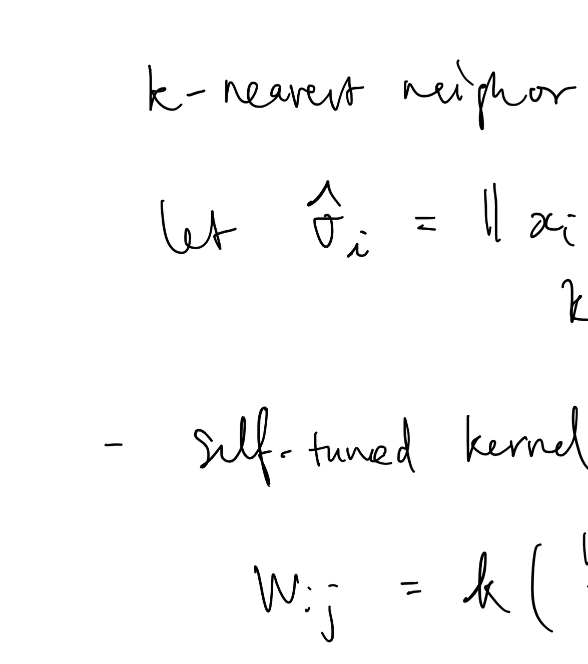
let $\hat{\sigma}_i = \| x_i - x_{j(i),k} \|$, where $x_{j(i),k}$ is the k-NN of x_i in X, $1 < k < n$

- self-tuned kernel [Zelnik-Manor and Perona 2005]

$$W_{ij} = k\left(\frac{\|x_i - x_j\|^2}{\hat{\sigma}_i \hat{\sigma}_j} \right), \quad k(\cdot) : \mathbb{R} \to \mathbb{R}$$

eg. $k(z) = e^{-z}$

$\hat{\sigma}_i = \hat{R}(x_i)$ $\quad$ $\hat{R}(\cdot)$ a function in $\mathbb{R}^D$, equivalent definition of kNN



$$\hat{R}(x) = \inf_{r>0} \left\{ \sum_{j=1}^n \mathbb{1}_{\{\|x_j - x\| < r\}} \geq k \right\}$$

- Convergence of kNN distance [NNDE, 60s-90s]

suppose $\hat{R}$ is small

$$\int_{B_{\hat{R}}(x)} p(u)\,du \approx p(x) \cdot v_d \hat{R}^d \approx \frac{k}{N}$$

estimate of $p(x)$ by kNN

$$\hat{R}(x) \sim \left(\frac{k}{N}\right)^{1/d} \to 0 \text{ if } \frac{k}{N} \to 0.$$

- For manifold data, p on M_d in $\mathbb{R}^D$,

$$\hat{\rho}(x) := \hat{R}(x) \left(\frac{1}{v_d} \frac{k}{N} \right)^{-1/d} \to \bar{\rho}(x) = p(x)^{-1/d}$$

Thm [C-Wu 2020] Manifold data assumption, as $n \uparrow$, $\log n \ll k \ll n$, then for large n, w.h.p

$$\frac{|\hat{\rho}(x) - \bar{\rho}(x)|}{\bar{\rho}(x)} = O_p\left(\left(\frac{k}{n}\right)^{2/d} \right) + O\left(\sqrt{\frac{\log n}{k}} \right)$$

for $x \in M$ uniformly.

This provides the C^0 convergence of (scaled) kNN adaptive bandwidth function to the limit $\bar{\rho} = p^{-1/d}$.

- Heat kernel on manifold

The spectral convergence shows the convergence of the graph Laplacian operator $L_{rw} = I - D^{-1}W$

to the manifold Laplacian Δ_M

M is smooth compact (no-boundary), $H_t(x,y)$ the heat kernel on M is the Green's function of the heat eqn on M: $u(t,x)$, $t > 0$, $x \in M$,

$$\begin{cases} \partial_t u = \Delta_M u \\ u(0,x) = f(x) \end{cases}$$

write Δ_M as Δ

The diffusion semi-group $Q_t := e^{t\Delta}$

$$u(t,x) = Q_t f(x) = \int_M H_t(x,y) f(y)\, dV(y)$$

(M, dV) the measure space

↑ Riemannian volume form

Eigen-functions of Δ: M smooth (connected) compact, $\{\mu_k, \psi_k\}_{k=1}^\infty$ eigen-pairs of $-\Delta$

$$-\Delta \psi_k = \mu_k \psi_k,$$

$$0 = \mu_1 < \mu_2 \leq \mu_3 \leq \cdots, \quad \psi_k \in C^\infty(M),$$

$$\langle \psi_k, \psi_l \rangle_M = \int_M \psi_k(x) \psi_l(x)\, dV(x) = \delta_{kl},$$

ψ_k is also the eigenfunction of the integral operator of the heat kernel

$$Q_t \psi_k = e^{-t\mu_k} \psi_k, \quad k = 1, 2, \cdots$$

Eg $\quad \psi_1(x) = $ constant.

bcz $\quad \int_M H_t(x,y)\, dV(y) = 1, \quad \forall x \in M.$

Under generic conditions,

$$H_t(x,y) = \sum_{k=1}^\infty e^{-t\mu_k} \psi_k(x) \psi_k(y), \quad \forall x, y \in M$$

Rk. Locally, $H_t(x,y)$ can be approximated by

$$G_t(x,y) = \frac{1}{(4\pi t)^{d/2}} e^{-\frac{d_M(x,y)^2}{4t}}, \quad 0 < t < t_0, \quad d_M(y,x) < \delta_0$$

Heat kernel parametrix on manifold.

Steven Rosenberg. The Laplacian on a Riemannian manifold: an introduction to analysis on manifolds. Number 31. Cambridge University Press, 1997

Globally, $H_t(x,y) \not\approx G_t(x,y)$ for $O(1)$ time, but $H_t(x,y)$ can have sub-gaussian decay w.r.t. $d_M(x,y)$.

- Eigen-convergence of graph Laplacian

L_n is $n \times n$ graph Laplacian. L_{un} or L_{rw}

$$L_n u_k = \lambda_k u_k, \quad k = 1, 2, \cdots$$

eigen-convergence: $\begin{cases} \lambda_k \approx \mu_k & \text{up to normalization} \\ u_k \approx \rho_X \psi_k \end{cases}$

For $f \in C(M)$, $\rho_X f = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} \in \mathbb{R}^n$

Under manifold data setting, for $k = 1, \cdots, K$,

$$|\lambda_k - \mu_k| = \tilde{O}\left(n^{-\frac{1}{d/2+2}} \right)$$

$$\|u_k - \rho_X \psi_k\|_2 = \tilde{O}\left(n^{-\frac{1}{d+4}} \right)$$

if $\quad \epsilon \sim \tilde{\Theta}\left(n^{-\frac{1}{d/2+2}} \right)$

using Gaussian kernelised affinity [C-Wu 2020].