Notes for Week 3/27/20 - 4/2/20

Alden Green

April 3, 2020

We observe data $X_1, \ldots, X_n \sim P$, where P is defined over a sample space $\mathcal{X} \subset \mathbb{R}^d$. For a given kernel $K : \mathbb{R}^+ \to \mathbb{R}^+$, and a radius r, define $K_r(z) = K(z/r)$. Construct the neighborhood graph $G_{n,r} = ([n], W)$, where $W_{ij} = K_r(\|X_i - X_j\|_2)$. The neighborhood graph Laplacian $\mathbf{L}_{n,r}$ is then defined to be $D - W$, where D is the diagonal degree matrix of $G_{n,r}$ with entries $D_{ii} = \sum_{j=1}^n W_{ij}$.

Let $f : \mathcal{X} \to \mathbb{R}$ possess s derivatives, which we will eventually assume are well-behaved in some sense (e.g. f is an the order-s Holder or Sobolev class). Our goal is to understand the behavior of three statistics: first, the pointwise evaluation $\mathbf{L}_{n,r}^s f(x)$ for some $x \in \mathcal{X}$ (which we will formally define momentarily); second the pointwise evaluation $\mathbf{L}_{n,r}^s f(X_i)$ for some $i = 1, \ldots, n$; and third the seminorm $f^T \mathbf{L}_{n,r}^s f$.

1 Pointwise Evaluation

For any $x \in \mathcal{X}$, define

$$\Bigl(\mathbf{L}_{n,r} f\Bigr)(x) := \sum_{i=1}^n \Bigl(f(x) - f(X_i)\Bigr) K_r(x, X_i). \tag{1}$$

When $x = X_i$ for $i = 1, \ldots, n$, we have $\Bigl(\mathbf{L}_{n,r} f\Bigr)(X_i) = \Bigl(\mathbf{L}_{n,r} f\Bigr)_i$, so (1) defines an extension of $\mathbf{L}_{n,r}$ from the data $X_1, \ldots, X_n$ to all of $\mathcal{X}$. For $s > 1$, recursively define

$$\Bigl(\mathbf{L}_{n,r}^s f\Bigr)(x) := \sum_{i=1}^n \Bigl\{\Bigl(\mathbf{L}_{n,r}^{s-1} f\Bigr)(x) - \Bigl(\mathbf{L}_{n,r}^{s-1} f\Bigr)(X_i)\Bigr\} K_r(x, X_i)$$

Our goal is to show (a) compute the expectation of $\bigl(\mathbf{L}_{n,r}^s f\bigr)(x)$, and (b) to show that $\bigl(\mathbf{L}_{n,r}^s f\bigr)(x)$ concentrates around its expectation. We will begin with the simplest non-trivial case of $s = 2$.

1.1 $s = 2$

The evaluation $\bigl(\mathbf{L}_{n,r}^2 f\bigr)(x)$ can be written as

$$\bigl(\mathbf{L}_{n,r}^2 f\bigr)(x) = \sum_{i=1}^n \sum_{j=1}^n \bigl(f(x) - f(X_j)\bigr) K_r(x, X_j) K_r(x, X_i) - \sum_{i=1}^n \sum_{j=1}^n \bigl(f(X_i) - f(X_j)\bigr) K_r(X_i, X_j) K_r(x, X_i) \tag{2}$$

1.1.1 Expectation at a fixed point.

From (2), we see

$$\mathbb{E}\Bigl[\bigl(\mathbf{L}_{n,r}^2 f\bigr)(x)\Bigr] = \sum_{i=1}^n \sum_{j=1}^n \Bigl\{\mathbb{E}\Bigl[\bigl(f(x) - f(X_j)\bigr) K_r(x, X_j) K_r(x, X_i)\Bigr] - \mathbb{E}\Bigl[\bigl(f(X_i) - f(X_j)\bigr) K_r(X_i, X_j) K_r(x, X_i)\Bigr]\Bigr\}$$

The expectation of the summand depends on whether $i = j$. If $i = j$, then

$$\mathbb{E}\Big[\big(f(x) - f(X_j)\big)K_r(x, X_j)K_r(x, X_i)\Big] = \int \big(f(x) - f(y)\big)\big(K_r(\|y - x\|)\big)^2\, dP(y) =: \big(I_{1,P}f\big)(x)$$

and clearly $\mathbb{E}\Big[\big(f(X_i) - f(X_j)\big)K_r(X_i, X_j)K_r(x, X_i)\Big] = 0$. Otherwise if $i \neq j$,

$$\mathbb{E}\Big[\big(f(x)-f(X_j)\big)K_r(x, X_j)K_r(x, X_i)\Big] = \mathbb{E}\Big[\big(f(x)-f(X)\big)K_r(x, X)\Big]\cdot\mathbb{E}\Big[K_r(x, X)\Big] =: \big(L_{P,r}f\big)(x)\cdot\mathbb{E}\Big[K_r(x, X)\Big],$$

by the law of conditional expectation

$$\mathbb{E}\Big[\big(f(X_i) - f(X_j)\big)K_r(X_i, X_j)K_r(x, X_i)\Big] = \mathbb{E}\Big[\big(L_{P,r}f\big)(X)\cdot K_r(x, X)\Big],$$

and therefore

$$\mathbb{E}\Big[\big(f(x) - f(X_j)\big)K_r(x, X_j)K_r(x, X_i)\Big] - \mathbb{E}\Big[\big(f(X_i) - f(X_j)\big)K_r(X_i, X_j)K_r(x, X_i)\Big] = \big(L_{P,r}^2 f\big)(x)$$

We conclude that

$$\mathbb{E}\Big[\big(\mathbf{L}_{n,r}^2 f\big)(x)\Big] = n(n-1)\big(L_{P,r}^2 f\big)(x) + n\big(I_{1,P}f\big)(x)$$

2 Semi-norm

2.1 $s = 3$

We use the notation $\mathbb{E}_{-i,j}[X] = \mathbb{E}[X|X_i, X_j]$. The third-order graph Sobolev seminorm $f^T L_{n,r}^3 f$ can be written as

$$\begin{aligned}
f^T L_{n,r}^3 f &= \sum_{i,j=1}^{n}\Big(\big(L_{n,r}f\big)(X_i) - \big(L_{n,r}f\big)(X_j)\Big)^2 K_r\Big(\|X_i - X_j\|_2\Big) \\
&\leq 3\sum_{i,j=1}^{n}\bigg(\big(L_{n,r}f\big)(X_i) - \mathbb{E}_{-i,j}\Big[\big(L_{n,r}f\big)(X_i)\Big]\bigg)^2 K_r\Big(\|X_i - X_j\|_2\Big) \;+ \\
&\quad 3\sum_{i,j=1}^{n}\bigg(\big(L_{n,r}f\big)(X_j) - \mathbb{E}_{-i,j}\Big[\big(L_{n,r}f\big)(X_j)\Big]\bigg)^2 K_r\Big(\|X_i - X_j\|_2\Big) \;+ \\
&\quad 3\sum_{i,j=1}^{n}\bigg(\mathbb{E}_{-i,j}\Big[\big(L_{n,r}f\big)(X_i)\Big] - \mathbb{E}_{-i,j}\Big[\big(L_{n,r}f\big)(X_j)\Big]\bigg)^2 K_r\Big(\|X_i - X_j\|_2\Big)
\end{aligned}$$

We upper bound the expectation of each of the three terms in the summand on the right hand side.

Term 1.

By the law of iterated expectation,

$$\begin{aligned}
&\mathbb{E}\Bigg[\bigg(\big(L_{n,r}f\big)(X_i) - \mathbb{E}_{-i,j}\Big[\big(L_{n,r}f\big)(X_i)\Big]\bigg)^2 K_r\Big(\|X_i - X_j\|_2\Big)\Bigg] = \\
&\quad \mathbb{E}\Bigg[\mathbb{E}_{-i,j}\bigg[\bigg(\big(L_{n,r}f\big)(X_i) - \mathbb{E}_{-i,j}\Big[\big(L_{n,r}f\big)(X_i)\Big]\bigg)^2\bigg] K_r\Big(\|X_i - X_j\|_2\Big)\Bigg] = \\
&\quad \mathbb{E}\Bigg[\mathrm{Var}_{-i,j}\Big[\big(L_{n,r}f\big)(X_i)\Big] K_r\Big(\|X_i - X_j\|_2\Big)\Bigg]
\end{aligned}$$

We express the conditional variance as a sum of conditional covariances,

$$\begin{aligned}\mathrm{Var}_{-i,j}\Big[\Big(\big(L_{n,r}f\big)(X_i)\Big)\Big] &= \sum_{k=1}^{n}\sum_{\ell=1}^{n}\mathrm{Cov}_{-i,j}\Big[\Big(f(X_i)-f(X_k)\Big)K_r\Big(\|X_i-X_k\|_2\Big),\Big(f(X_i)-f(X_\ell)\Big)K_r\Big(\|X_i-X_\ell\|_2\Big)\Big] \\ &= (n-3)\mathrm{Var}_{-i}\Big[\Big(f(X_i)-f(X)\Big)K_r\Big(\|X_i-X\|_2\Big)\Big].\end{aligned}$$

where in the second equality we have used that $X_1,\dots,X_n$ are i.i.d samples, and so conditional on X_i and X_j the covariance is equal to zero unless $k=\ell$. By Lemma 1,

$$\mathrm{Var}_{-i}\Big[\Big(f(X_i)-f(X)\Big)K_r\Big(\|X_i-X\|_2\Big)\Big] \leq \mathbb{E}_{-i}\Big[\Big(\Big(f(X_i)-f(X)\Big)K_r\Big(\|X_i-X\|_2\Big)\Big)^2\Big] \leq M^2r^{2+d}K_{\max}^2p_{\max}$$

and as a result the expectation of term 1 is upper bounded

$$\begin{aligned}\mathbb{E}\Big[\Big(\big(L_{n,r}f\big)(X_i)-\mathbb{E}_{-i,j}\Big[\big(L_{n,r}f\big)(X_i)\Big]\Big)^2K_r\Big(\|X_i-X_j\|_2\Big)\Big] &\leq M^2K_{\max}^2p_{\max}\cdot nr^{2+d}\mathbb{E}\Big[K_r\Big(\|X_i-X_j\|_2\Big)\Big] \\ &\leq M^2K_{\max}^3p_{\max}^2\cdot nr^{2+2d}\end{aligned}$$

Term 2. By symmetry, the same bound holds for term 2.

Term 3. Expressing $L_{n,r}f(X_i)=\sum_{k=1}^n\big(f(X_i)-f(X_k)\big)K_r(X_i,X_k)$, by the linearity of expectation

$$\begin{aligned}\mathbb{E}_{-i,j}\Big[L_{n,r}f(X_i)\big] &= (n-2)L_{P,r}f(X_i)+\big(f(X_i)-f(X_j)\big)K_r(X_i,X_j) \\ \mathbb{E}_{-i,j}\Big[L_{n,r}f(X_j)\big] &= (n-2)L_{P,r}f(X_j)+\big(f(X_j)-f(X_i)\big)K_r(X_j,X_i)\end{aligned}$$

Upper bounding the square of sums by twice the sum of squares, we have

$$\begin{aligned}&\Big(\Big(\mathbb{E}_{-i,j}\Big[\big(L_{n,r}f\big)(X_i)\Big]-\mathbb{E}_{-i,j}\Big[\big(L_{n,r}f\big)(X_j)\Big]\Big)^2K_r\Big(\|X_i-X_j\|_2\Big) \leq \\ &\quad 2(n-2)^2\Big(\big(L_{P,r}f\big)(X_i)-\big(L_{P,r}f\big)(X_j)\Big)^2K_r(X_i,X_j)+8\big(f(X_j)-f(X_i)\big)^2\big(K_r(X_j,X_i)\big)^3\end{aligned}$$

3 Additional Theory

We make some assumptions on the function f, the distribution P, and the kernel function K.

(A1) $f\in C^s(\mathcal{X};M)$ for some $s>0$. If $s>1$, then f is also compactly supported on a strict subset of $\mathcal{X}$.

(A2) P admits a density p with respect to the Lebesgue measure on $\mathbb{R}^d$. The density $p\in C^k(\mathcal{X};p_{\max})$, for some $k>0$.

(A3) K is supported on a subset of $[0,1]$, and $K(z)\leq K_{\max}<\infty$ for all $z\in[0,1]$.

Under these assumptions, we can bound various integrals.

Lemma 1. *Let f satisfy (A1) with $s=1$, P satisfy (A2) with $k=0$, and K satisfy (A3). Then for any $x\in\mathcal{X}$,*

$$\mathbb{E}\Big[\big(f(x)-f(X)\big)^2\big(K_r(\|x-X\|_2)\big)^2\Big] \leq M^2r^{2+d}K_{\max}^2p_{\max}$$

Lemma 2. *Let f satisfy (A1) with $s=2$, P satisfy (A2) with $k=1$, and K satisfy (A3). Then for any $x\in\mathcal{X}$,*

$$\Big|\mathbb{E}\Big[\big(f(x)-f(X)\big)\big(K_r(\|x-X\|_2)\big)\Big]\Big| \leq Mr^{2+d}K_{\max}p_{\max}$$

Lemma 3. *Let f satisfy (A1) with $s = 2$, P satisfy (A2) with $k = 1$, and K satisfy (A3). Then for any $x \in \mathcal{X}$,*

$$\left|\mathbb{E}\Big[\big(f(x) - f(X)\big)\big(f(x) - f(X')\big)\big(K_r(\|x - X\|_2)\big)\big(K_r(\|x - X'\|_2)\big)\Big]\right| \leq M^2 r^{4+2d} K_{\max}^2 p_{\max}^2$$

The following Lemma is more easily stated and proved using multi-index notation. For a function $f : \mathbb{R}^d \to \mathbb{R}$ which is k-times differentiable, and $\alpha \in [\mathbb{N}]^d$ satisfying $|\alpha| := \alpha_1 + \ldots \alpha_d = k$, we write

$$D^\alpha f(x) = \frac{\partial^k f}{\partial x_1^{\alpha_1} \cdots \partial x_1^{\alpha_d}},$$

for the (α)th-partial derivative of f at k. Additionally, let $(x)^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ for any $x \in \mathbb{R}^d$.

Lemma 4. *Let f satisfy (A1) with $s > 2$, P satisfy (A2) with $k = 2$, and K satisfy (A3). Then for any $x \in \mathcal{X}$,*

content...

Proof. Let $\gamma = \min\{s - 2, 1\}$. Since $f \in C^s(\mathcal{X}; M)$ for $s > 2$, we have that f is twice-differentiable for all $x \in \mathcal{X}$, and moreover for any $y \in \mathcal{X}$,

$$\left|f(y) - \Big(\sum_{|\alpha|=0}^{2} D^\alpha f(x)(x)^\alpha\Big)\right| \leq M\|y - x\|_2^\gamma$$

Therefore,

$$\begin{aligned} -L_{P,r} f(x) &= \int \Big(f(y) - f(x)\Big) K_r(\|y - x\|)\, dP(x) \\ &= \sum_{|\alpha|=1}^{2} \end{aligned}$$

□