Notes on 'Adaptive Non-Parametric Regression With the K-NN Fused Lasso'

Alden Green

February 8, 2019

Let $\mathbf{X} = x_1, \ldots, x_n$ be sampled i.i.d from μ with density function $p(\cdot)$ over some subset $\mathcal{X}$ of Euclidean space, and suppose

$$y_i = f_0(x_i) + \epsilon_i, \quad \epsilon_i \overset{i.i.d}{\sim} SG(\sigma^2)$$

holds for some unknown f_0. Let $\widehat{\theta}$ be the solution to the *fused lasso*

$$\widehat{\theta} := \operatorname*{argmin}_{\theta \in \mathbb{R}^n} \left\{ \frac{1}{2} \|y - \theta\|_2^2 + \lambda \|\nabla_G \theta\|_1 \right\}$$

where $\lambda > 0$ is a tuning parameter, and ∇_G is an oriented incidence matrix of the graph G.

The K*-NN-FL* estimator computes the fused lasso over the K-NN graph G_K of $\mathbf{X}$. The ϵ*-FL* estimator computes the fused lasso over the ϵ graph G_ϵ.

The assumptions required for Theorems 1 and 2 are as follows.

(a) For all $x \in \mathcal{X}$

$$0 < p_{\min} \leq p(x) \leq p_{\max} < \infty$$

(b) The base measure μ in $\mathcal{X}$ satisfies

$$r^d c_{1,d} \leq \mu(B_r(x)) \leq c_{2,d} r^d \qquad (\forall x \in \mathcal{X})$$

(c) There exists a homeomorphism (continuous bijection with continuous inverse) $h : \mathcal{X} \to [0,1]^d$ such that

$$L_{\min} d_{\mathcal{X}}(x, x') \leq \|h(x) - h(x')\|_2 \leq L_{\max} d_{\mathcal{X}}(x, x') \qquad (\forall x, x' \in \mathcal{X})$$

(d) The function $g_0 := f_0 \circ h^{-1}$ has bounded variation, meaning

$$|g_0|_{BV(\Omega)} := \sup \left\{ \int_\Omega g_0(x) \mathrm{div}(g)(x) dx; g \in C_c^1(\Omega, \mathbb{R}^d), \|g\|_\infty \leq 1 \right\} < \infty$$

where $\mathfrak{B}$ is the Borel σ-algebra of $\Omega = [0,1]^d$, $C_c^1(\Omega, \mathbb{R}^d)$ is the set of $\mathfrak{B}$-measurable, continuously differentiable functions into $\mathbb{R}^d$, and

$$\|g\|_\infty := \left\| \left(\sum_{i=1}^{d} g_i^2 \right)^{1/2} \right\|_{L_\infty(\Omega)}.$$

(e) g_0 is piecewise Lipschitz[1], meaning there exists a set $\mathcal{S} \subset (0,1)^d$ such that

  (a) $\nu(\mathcal{S}) = 0$.

  (b) $\mu\Big(h^{-1}\big(S_\epsilon \cup ([0,1]^d \setminus \Omega_\epsilon)\big)\Big) \leq C_{\mathcal{S}}\epsilon$

  (c) There exists a positive constant L_0 such that if z and z' belong to the same connected component of $\Omega_\epsilon \setminus B_\epsilon(\mathcal{S})$, then

$$|g(z) - g(z')| \leq L_0 \|z - z'\|_2$$

where $\Omega_\epsilon = [0,1]^d \setminus B_\epsilon(\partial[0,1]^d)$.

Theorem 1. *Let $K \asymp \log^{1+2r} n$ for some $r > 0$, Then under Assumptions 1-3, with an appropriate choice of the tuning parameter λ, the K-NN-FL estimator $\widehat{\theta}$ satisfies*

$$\left\| \widehat{\theta} - \theta^\star \right\|_n^2 = O_{\mathbb{P}} \left(\frac{\log^{1+2r} n}{n} + \frac{\log^{1.5+r} n}{n} \|\nabla_{G_K} \theta^\star\|_1 \right)$$

This upper bound also holds for ϵ-NN-FL if we replace $\|\nabla_{G_K}\theta^\star\|_1$ with $\|\nabla_{G_\epsilon}\theta^\star\|_1$.

Theorem 2. *Under Assumptions 1-5, with an appropriate choice of the tuning parameter λ, the K-NN-FL estimator $\widehat{\theta}$ satisfies*

$$\left\| \widehat{\theta} - \theta^\star \right\|_n^2 = \widetilde{O}_{\mathbb{P}} \left(\frac{1}{n^{1/d}} \right).$$

1 Proof of Theorem 1

To ease proofs, we will assume $\mathcal{X} = [0,1]^d$.

Construct $G_{lat} = (V_{lat}, E_{lat})$ a lattice graph with equal side lengths in $[0,1]^d$, where

$$V_{lat} = P_{lat}(N) := \left\{ \left(\frac{i_1}{N} - \frac{1}{2N}, \cdots, \frac{i_d}{N} - \frac{1}{2N} \right) : i_1, \ldots, i_d \in \{1, \ldots, N\} \right\}$$

$(z, z') \in E_{lat}$ if and only if $\|z - z'\| \leq \dfrac{1}{N}$

[1]Technically, the requirement is slightly weaker than piecewise Lipschitz.

where z and $z' \in P_{lat}(N)$.

Denoting $I = P_{lat}$, we define

$$P_I(x) = \text{argmin}\left\{\|x - z'\|_\infty, z' \in P_{lat}(N)\right\}$$

Then, let $C(z) = \left\{x \in [0,1]^d : z = P_I(x)\right\}$ be the collection of cells associated with the mesh $P_{lat}(N)$, noting that $\{C(z) : z \in P_{lat}(N)\}$ defines a partition over $[0,1]^d$.

Quantization. For a given $\theta \in \mathbb{R}^n$, the *quantization* $\theta_I \in \mathbb{R}^n$

$$(\theta_I)_i := \theta_j, \quad \text{where} \quad x_j = \underset{x_l, l \in [n]}{\text{argmin}} \|P_I(x_i) - x_l\|_\infty$$

is constant over every cell $C(z)$. We now induce a signal in $\mathbb{R}^{N^d}$ corresponding to the elements in I. Let $\{z_1, \ldots, z_{N^d}\} = I$. Then we write

$$I_k = \{i \in [n] : P_I(x_i) = z_k\}$$

for $k = 1, \ldots, N^d$. Define $\theta^I \in \mathbb{R}^{N^d}$ by

$$(\theta^I)_k := \begin{cases} (\theta_I)_i, x_i \in I_k \\ 0, I_k = \emptyset \end{cases}$$

where we note that (θ^I) is well-defined since $(\theta_I)_i = (\theta_I)_j$ if x_i and x_j are both in I_k.

1.1 Controlling counts of mesh

Define the event Ω as: "If $x_i \in C(z_k)$ and $x_i \in C(z_l)$ for $z_k, z_l \in I$ with $\|z_k - z_l\|_2 \le \frac{1}{N}$, then x_i and x_j are connected in the K-NN graph." Then,

Lemma 1. *Take Assumptions 1-3, and additionally assume that N in the construction of $G_{lat}(N)$ is chosen as*

$$N \ge \left[\frac{3\sqrt{d}(2c_{2,d}p_{\max})^{1/d} n^{1/d}}{L_{\min} K^{1/d}}\right]. \tag{1}$$

Then,

$$\mathbb{P}(\Omega) \ge 1 - n\exp(-K/3).$$

1.2 Bounding Empirical Process

Lemma 2.

1.3 Mesh embedding for K-NN graph

Lemma 3. *Fix N to satisfy (1), and let us assume that the event Ω from Lemma 1 holds. Denote $I = P_{lat}(N)$ to be the mesh. Then, for all $e \in \mathbb{R}^n$, it holds that*

$$\left|e^T(\theta - \theta_I)\right| \le 2\,\|e\|_\infty\,\|\nabla_{G_K}\theta\|_1\,, \qquad (\forall \theta \in \mathbb{R}^n)$$

Moreover,

$$\left\|D\theta^I\right\|_1 \le \|\nabla_{G_K}\theta\|_1\,, \qquad (\forall \theta \in \mathbb{R}^n)$$

where D is the incidence matrix of G_{lat}.