Notes for Week 2/8/19 - 2/15/19

Alden Green

February 16, 2019

Consider distributions $\mathbb{P}$ and $\mathbb{Q}$ supported on $\mathcal{D} \subset \mathbb{R}^d$ which are absolutely continuous with density functions f and g, respectively. For fixed $t \geq 0$, Let $\mathbf{Z} = (z_1, \ldots, z_t)$, where for $i = 1, \ldots, t$, $z_i \sim \frac{\mathbb{P}+\mathbb{Q}}{2}$ are independent. Given $\mathbf{Z}$, for $i = 1, ..., t$ let

$$\ell_i = \begin{cases} 1 \text{ with probability } \frac{f(z_i)}{f(z_i)+g(z_i)} \\ -1 \text{ with probability } \frac{g(z_i)}{f(z_i)+g(z_i)} \end{cases}$$

be conditional independent labels, and write

$$1_X = \begin{cases} 1, & l_i = 1 \\ 0, & \text{otherwise} \end{cases} \qquad 1_Y = \begin{cases} 1, & l_i = -1 \\ 0, & \text{otherwise.} \end{cases}$$

We will write $\mathbf{X} = \{x_1, \ldots, x_{N_1}\} := \{z_i : \ell_i = 1\}$ and similarly $\mathbf{Y} = \{y_1, \ldots, y_{N_2}\} := \{y_i : \ell_i = -1\}$, where N_1 and N_2 are random.

Our statistical goal is hypothesis testing: that is, we wish to construct a test function ϕ which differentiates between

$$\mathbb{H}_0 : f = g \text{ and } \mathbb{H}_1 : f \neq g.$$

For a given function class $\mathcal{H}$, some $\epsilon > 0$, and test function ϕ a Borel measurable function of the data with range $\{0, 1\}$, we evaluate the quality of the test using *worst-case risk*

$$R_\epsilon^{(t)}(\phi; \mathcal{H}) = \sup_{f \in \mathcal{H}} \mathbb{E}_{f,f}^{(t)}(\phi) + \sup_{\substack{f,g \in \mathcal{H} \\ \delta(f,g) \geq \epsilon}} \mathbb{E}_{f,g}^{(t)}(1 - \phi)$$

Total variation test. As in [1], define the *K-NN graph* $G_K = (V, E_K)$ to have vertex set $V = \{1, \ldots, t\}$ and edge set E_K which contains the pair (i, j) if and only if x_i is among the K-nearest neighbors (with respect to Euclidean distance) of x_j, or vice versa. Let D_K denote the incidence matrix of G_K.

Define the *kNN-total variation* test statistic to be

$$T_{TV} = \sup_{\substack{\theta \in \mathbb{R}^t: \\ \|D_k \theta\|_1 \leq C_{n,k}}} \left(\sum_{i:(1_X)_i = 1} \theta_i - \sum_{j:(1_Y)_j = 1} \theta_j \right) \tag{1}$$

Hereafter, take $\mathcal{D} = [0,1]^d$, and consider

$$\mathcal{H}_{lip}(L) = \left\{ f : [0,1]^d \to \mathbb{R}^+ : \int_{\mathcal{D}} f = 1, f \text{ } L\text{-piecewise Lipschitz, bounded above and below} \right\}$$

where

Definition 0.1 (Piecewise Lipschitz)**.** A function f is *L-piecewise lipschitz* over $[0,1]^d$ if there exists a set $\mathcal{S} \subset [0,1]^d$ such that

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

(b) There exist $C_{\mathcal{S}}, \epsilon_0$ such that $\mu\Big(\big(\mathcal{S}_\epsilon \cup (\partial \mathcal{D})_\epsilon\big) \cap [0,1]^d \Big) \leq C_{\mathcal{S}} \epsilon$ for all $0 < \epsilon \leq \epsilon_0$.

(c) For any z, z' in the same connected component of $[0,1]^d \setminus \big(\mathcal{S}_\epsilon \cup (\partial \mathcal{D})_\epsilon\big)$,

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

and

Definition 0.2 (Bounded above and below)**.** A function $f : \mathcal{D} \to \mathbb{R}$ is *bounded above and below* if there exists $p_{\min}, p_{\max}$ such that

$$0 < p_{\min} < f(x) < p_{\max} < \infty \qquad (\forall x \in \mathcal{D})$$

Conjecture 1. *For τ =??? and $K \asymp \log^{1+2r}(n)$ for some $r \geq 0$, the test $\phi_{TV} = \{T_{TV} \geq \tau\}$ has worst-case risk*

$$R_\epsilon^{(t)}(\mathcal{H}_{lip}(L)) \leq c_1 / a^2$$

whenever $\epsilon \geq c_2 a \log^\alpha m m^{-1/d}$ where $\alpha = 3r + 5/2 + (2r+1)/d$ and c_1 and c_2 are constants which depend only on (d, L).

Proof. Write

$$\left(\sum_{i:(1_X)_i = 1} \theta_i - \sum_{j:(1_Y)_j = 1} \theta_j \right) = \langle \theta, 1_X - 1_Y \rangle$$

and let

$$\widehat{\theta} \in \operatorname*{argmax}_{\theta \in \mathbb{R}^{2m}} \{ \langle \theta, 1_X - 1_Y \rangle : \|D_K \theta\|_1 \leq C_{n,k} \}$$

satisfy $T_{TV} = \langle \widehat{\theta}, 1_X - 1_Y \rangle$. For $i = 1, \ldots, 2m$, introduce $\theta^\star$ defined by

$$(\theta^\star)_i = \left(\frac{f(z_i) - g(z_i)}{f(z_i) + g(z_i)} \right)$$

We have

$$T_{TV} = \langle \widehat{\theta} - \theta^\star, 1_X - 1_Y \rangle + \langle \theta^\star, 1_X - 1_Y \rangle$$

Let

$$\widetilde{\theta} = \frac{\theta^\star}{\|D\theta^\star\|_1}.$$

We bound the first term, dividing into two cases.

Case 1: First, suppose $\left|\langle \widehat{\theta} - \theta^\star, 1_X - 1_Y \rangle\right| \geq \left\|\theta - \widetilde{\theta}\right\|_2^2$.

We begin by employing a basic inequality argument. Since $\widetilde{\theta}$ is feasible and $\widehat{\theta}$ optimal for (1), we have

$$\langle \widehat{\theta}, 1_X - 1_Y \rangle \geq \langle \widetilde{\theta}, 1_X - 1_Y \rangle$$

which, letting $w = (1_X - 1_Y) - \theta^\star$, means

$$\langle \widehat{\theta}, \theta^\star \rangle \geq \langle \widetilde{\theta}, \theta^\star \rangle - \langle \widehat{\theta} - \widetilde{\theta}, w \rangle \tag{2}$$

As a result, we have

$$\begin{aligned}
\left|\langle \widehat{\theta} - \theta^\star, 1_X - 1_Y \rangle\right| &\leq \left|\langle \widehat{\theta} - \theta^\star, \theta^\star \rangle\right| + \left|\langle \widehat{\theta} - \theta^\star, w \rangle\right| \\
&\stackrel{(i)}{\leq} \left|\langle \widetilde{\theta} - \theta^\star, \theta^\star \rangle\right| + \left|\langle \widehat{\theta} - \widetilde{\theta}, w \rangle\right| + \left|\langle \widehat{\theta} - \theta^\star, w \rangle\right| \\
&\leq \left|\langle \widetilde{\theta} - \theta^\star, \theta^\star \rangle\right| + 2\left|\langle \widehat{\theta} - \widetilde{\theta}, w \rangle\right| + \left|\langle \widetilde{\theta} - \theta^\star, w \rangle\right|
\end{aligned}$$

where (i) follows from (2).

w are mean-zero, bounded random variables, and are therefore subgaussian with parameter 2. Then, from [1], we have

Lemma 1 (Approximate restatement of [1])**.**

$$\left|\langle \widehat{\theta} - \theta^\star, w \rangle\right| \lesssim \sqrt{K \log n}\left[\left\|D_K \widetilde{\theta}\right\|_1 + \left\|D_K \widehat{\theta}\right\|\right] + \sqrt{K} \cdot \left(\left\|\widehat{\theta} - \widetilde{\theta}\right\|_2\right) \tag{3}$$

with probability at least

$$\Omega\left(1 - \frac{n}{K}\exp(-K) - n\exp(-K)\right)$$

Condition on the event of (3) holding. As a result, we have

$$\begin{aligned}
\left|\langle \widehat{\theta} - \theta^\star, 1_X - 1_Y \rangle\right| &\lesssim \left|\langle \widetilde{\theta} - \theta^\star, \theta^\star \rangle\right| + \sqrt{K \log n}\left[\left\|D_K \widetilde{\theta}\right\|_1 + \left\|D_K \widehat{\theta}\right\|\right] + \sqrt{K} \cdot \left(\left\|\widehat{\theta} - \widetilde{\theta}\right\|_2\right) \\
&\leq \left|\langle \widetilde{\theta} - \theta^\star, \theta^\star \rangle\right| + \sqrt{K \log n}\left[\left\|D_K \widetilde{\theta}\right\|_1 + \left\|D_K \widehat{\theta}\right\|\right] + \sqrt{K} \cdot \left(\sqrt{\left|\langle \widehat{\theta} - \theta^\star, 1_X - 1_Y \rangle\right|}\right)
\end{aligned}$$

which, along with the inequality $ab - b^2/4 \le a^2$, implies

$$\left|\langle \widehat{\theta} - \theta^\star, 1_X - 1_Y \rangle\right| \le \left|\langle \widetilde{\theta} - \theta^\star, \theta^\star \rangle\right| + \sqrt{K \log n}\left[\left\|D_K \widetilde{\theta}\right\|_1 + \left\|D_K \widehat{\theta}\right\|\right] + K$$

Continue from here.

Case 2: Otherwise, we may assume $\left|\langle \widehat{\theta} - \theta^\star, 1_X - 1_Y \rangle\right| \le \left\|\theta - \widetilde{\theta}\right\|_2^2$.

To upper bound $\left\|\theta - \widetilde{\theta}\right\|_2^2$, we use the relation

$$\langle \widehat{\theta}, 1_X - 1_Y \rangle \le \langle \widetilde{\theta}, 1_X - 1_Y \rangle$$

and therefore

$$\left\|\widehat{\theta} - (1_X - 1_Y)\right\|_2^2 \le \left\|\widetilde{\theta} - (1_X - 1_Y)\right\|_2^2 + \left(\left\|\widehat{\theta}\right\|_2^2 - \left\|\widetilde{\theta}\right\|_2^2\right)$$

Continue from here using standard basic inequality + Poincare bound on the second term.

$\square$

REFERENCES

[1] Oscar Hernan Madrid Padilla, James Sharpnack, Yanzhen Chen, and Daniela M Witten. Adaptive non-parametric regression with the k-nn fused lasso. *arXiv preprint arXiv:1807.11641*, 2018.