A unified statement for smoothness

Miaoyan Wang, 03/31/2021

Let $\Theta \in [-1,1]^{d_1\times\cdots\times d_K}$ be the signal tensor, and $d_t = \prod_{k=1}^K d_k$ the total dimension. We quantify the distribution of entries in tensor Θ using a histogram with bin width $\Delta s = 1/d_t$. Specifically, we divide the range $[-1,1]$ into d_t equally sized bins, and let $(p_i)_{i=1,\ldots,d_t}$ denote the frequency of tensor entries in each bin; that is, $p_i \stackrel{\text{def}}{=} \mathbb{P}_{\omega\sim\Pi}(-1+(i-1)\Delta s < \Theta(\omega) \leq -1+i\Delta s)$. Let $\mathcal{N} = \{i\colon p_i \gg \Delta s\} \subset [d_t]$ to collect the indices for heavy bins whose probability mass is asymptotically larger than the bin size Δs as $d\to\infty$,

Definition 1 (α-smoothness)**.** Fix $i \notin \mathcal{N}$. Assume there exist constant $\alpha = \alpha(\pi) > 0, c = c(\pi) > 0$, independent of tensor dimension, such that

$$\max_{j=1,\ldots,\rho(i,\mathcal{N})} \frac{p_i + \ldots + p_{i+j}}{(j\Delta s)^\alpha} \leq c, \tag{1}$$

where $\rho(i,\mathcal{N}) = \min_{j\in\mathcal{N}} |i-j|$ denotes the distance from index i to the nearest index in $\mathcal{N}$. We make the conversion that $\alpha = \infty$ if the numerator in (1) is zero, implying almost no entries of which $\Theta(\omega)$ is around the i-th bin.

Let $\mathbb{P}_{\boldsymbol{X}}$ denote either continuous or discrete distribution over feature space $\mathcal{X}$. Define a reference mass $\Delta s = 0$ if $\mathcal{X}$ is uncountable, or $\Delta s = \Omega\left(\frac{1}{|\mathcal{X}|}\right)$ if $\mathcal{X}$ is countable. We use $\mathcal{N} \subset [-1,1]$ to collect levels whose probability mass in a Δs-neighborhood is heavier than the uniform measure; i.e,

$$\mathcal{N} = [-1,1]/\left\{\pi\colon \mathbb{P}_{\boldsymbol{X}}(|f(\boldsymbol{X})-\pi| \leq \Delta s) \leq C\Delta s\right\}. \tag{2}$$

Here $C > 0$ is a constant independent of feature dimension d.

Definition 2 (α-smoothness)**.** For a given $\pi \notin \mathcal{N}$, we say f is (α,π)-smooth if there exists $\alpha = \alpha(\pi) \geq 1, c = c(\pi) > 0$, independent of feature dimension d, such that

$$\sup_{\Delta s \leq t < \rho(\pi,\mathcal{N})} \frac{\mathbb{P}_{\boldsymbol{X}}(\Delta s \leq |f(\boldsymbol{X})-\pi| \leq t)}{t^\alpha} \leq c, \tag{3}$$

where $\rho(\pi,\mathcal{N}) = \inf_{\pi'\in\mathcal{N}} |\pi-\pi'|$ denotes the distance from π to the nearest point in $\mathcal{N}$. We make the conversion that $\alpha = \infty$ if $\rho(\pi,\mathcal{N}) \leq \Delta s$.

Remark 1 (Smoothness index)**.** The largest possible $\alpha = \alpha(\pi)$ in (3) is called the smoothness index at level π. By definition (2) and (3), we always have $\alpha(\pi) \geq 1$ at levels $\pi \notin \mathcal{N}$.

Theorem 0.1 (Nonparametric regression via sign series)**.** *Assume f is globally α-smooth and r-sign rank representable. Denote $t_n = \frac{dr}{|\Omega|}$ for tensor completion problem, or $t_n = \frac{r(s_1+s_2)\log d}{n}$ for sparse matrix regression problem.*

1. *(Sign estimation) For all* $\pi \notin \mathcal{N}$,

$$\|\text{sign } \hat{f} - \text{sign } (f - \pi)\|_1 \leq t_n^{\alpha/(2+\alpha)} + \frac{1}{\rho^2(\pi, \mathcal{N})} t_n.$$

2. *(Signal estimation) Assume* $\mathcal{N}$ *is countable. Then, we have*

$$\|\hat{f} - f\|_1 \lesssim t_n^{\alpha/(2+\alpha)} + \frac{|\mathcal{N}| \vee 1}{H} + H t_n.$$

Setting $H = \sqrt{\frac{|\mathcal{N}| \vee 1}{t_n}}$ *yields the optimal rate*

$$\|\hat{f} - f\|_1 = \max\left(t_n^{\alpha/(2+\alpha)},\ t_n^{1/2}\sqrt{|\mathcal{N}|},\ t_n^{1/2}\right).$$