Infinite noise case

Chanwoo Lee
February 6, 2021

Assumption 1 (sub-Gaussian noise)**.**

1. There exists a constant $\beta > 0$, independent of tensor dimension such that $\|\Theta\|_\infty \leq \beta$.
2. The noise entries $\mathcal{E}(\omega)$ are independent and sub-Gaussian, i.e, $\mathbb{P}(|\mathcal{E}(\omega)| \geq B) \leq Ce^{-B^2/\sigma^2}$ for $B > 0$.

There are several equations that we will use.

$$\begin{aligned}
L(\mathcal{Z}, \mathcal{Y}_\Omega - \pi) &= \frac{1}{|\Omega|}\sum_{\omega\in\omega} |\mathcal{Y}(\omega) - \pi| \times |\mathrm{sgn}(\mathcal{Z}(\omega)) - \mathrm{sgn}(Y(\omega) - \pi)| \\
&:= \frac{1}{|\Omega|}\sum_{\omega\in\Omega} \ell_\omega(\mathcal{Z}, \mathcal{Y} - \pi), \\
\mathrm{Risk}(\mathcal{Z}) - \mathrm{Risk}(\bar{\Theta}) &= \mathbb{E}\left[L(\mathcal{Z}(\omega), \bar{\mathcal{Y}}(\omega)) - L(\bar{\Theta}(\omega), \bar{\mathcal{Y}}(\omega))\right] \\
&= \mathbb{E}_{\omega\in\Pi}|\bar{\Theta}(\omega)||\mathrm{sgn}(\bar{\Theta}(\omega)) - \mathrm{sgn}(\mathcal{Z}(\omega))|, \qquad \text{I'd conjecture an additional range term in this relationship.} \\
\mathrm{MAE}(\mathrm{sgn}(\mathcal{Z}), \mathrm{sgn}(\Theta - \pi)) &\lesssim [\mathrm{Risk}(\mathcal{Z}) - \mathrm{Risk}(\bar{\Theta})]^{\frac{\alpha}{\alpha+1}} + \frac{1}{\rho(\pi, \mathcal{N})}[\mathrm{Risk}(\mathcal{Z}) - \mathrm{Risk}(\bar{\Theta})].
\end{aligned}$$

scale free — scale dependent

In the case with infinite noise, previous variance term is not true because $|\mathcal{Y}(\omega) - \pi|$ is not bounded anymore. Therefore, we modify the variance term as,

$$\begin{aligned}
\mathrm{Var}\left[\ell_\omega\left(\mathcal{Z}(\omega), \bar{\mathcal{Y}}(\omega)\right) - \ell_\omega\left(\bar{\Theta}(\omega), \bar{\mathcal{Y}}(\omega)\right)\right] &\leq \mathbb{E}|\ell_\omega(\mathcal{Z}(\omega), \bar{\mathcal{Y}}(\omega)) - \ell_\omega(\bar{\Theta}(\omega), \bar{\mathcal{Y}}(\omega))|^2 \\
&= \mathbb{E}|\bar{\mathcal{Y}}(\omega)|^2|\mathrm{sgn}(\bar{\Theta}(\omega)) - \mathrm{sgn}(\mathcal{Z}(\omega))|^2 \\
&\leq 2\mathbb{E}(\bar{Y}(\omega) - \bar{\Theta}(\omega) + \bar{\Theta}(\omega))^2|\mathrm{sgn}(\bar{\Theta}(\omega)) - \mathrm{sgn}(\mathcal{Z}(\omega)| \\
&\leq 2(\|\mathcal{Y} - \Theta\|_\infty^2 + \|\Theta - \pi\|_\infty^2)\mathbb{E}|\mathrm{sgn}(\bar{\Theta}(\omega)) - \mathrm{sgn}(\mathcal{Z}(\omega)| \\
&\lesssim (\|\mathcal{E}\|_\infty^2 + \beta^2)\mathbb{E}|\mathrm{sgn}(\bar{\Theta}(\omega)) - \mathrm{sgn}(\mathcal{Z}(\omega)|.
\end{aligned}$$

Therefore, we have

$$\mathrm{Var}\left[\ell_\omega\left(\mathcal{Z}(\omega), \bar{\mathcal{Y}}(\omega)\right) - \ell_\omega\left(\bar{\Theta}(\omega), \bar{\mathcal{Y}}(\omega)\right)\right] \lesssim (\|\mathcal{E}\|_\infty^2 + \beta^2)\left\{[\mathrm{Risk}(\mathcal{Z}) - \mathrm{Risk}(\bar{\Theta})]^{\frac{\alpha}{\alpha+1}} + \frac{1}{\rho(\pi, \mathcal{N})}[\mathrm{Risk}(\mathcal{Z}) - \mathrm{Risk}(\bar{\Theta})]\right\}.$$

From the variance to mean relationship, we have

$$\mathbb{P}\left(\mathrm{Risk}(\mathcal{Z}) - \mathrm{Risk}(\bar{\Theta}) \leq L_n\right) \lesssim e^{-nL_n},$$

where the convergence rate L_n is a solution of the following inequality,

$$\frac{1}{L_n}\int_{L_n}^{\sqrt{(\|\mathcal{E}\|_\infty^2 + \alpha^2)(L_n^{\frac{\alpha}{\alpha+1}} + \frac{L_n}{\rho})}} \sqrt{\mathcal{H}_{[]}(\epsilon, \mathcal{F}_\mathcal{T}, \|\cdot\|_2)}d\epsilon \leq C\sqrt{n},$$

for some constant $C > 0$. By the similar way in our paper, we can show that

$$L_n \approx t_n^{\frac{\alpha}{\alpha+2}} + \frac{1}{\rho}t_n, \text{ where } t_n = \frac{(\|\mathcal{E}\|_\infty^2 + \beta^2)d_{\max}rK\log K}{n}.$$

For any given $B > 0$, we can guarantee the following inequality with at least probability $1 - e^{-B}$,

$$\|\mathcal{E}\|_\infty \leq \sqrt{2\sigma^2(K\log d_{\max} + B)}$$

Therefore, with the probability at most $1-\exp\left(-(\sigma^2\log d_{\max}+\beta^2)rd_{\max}\right)-\exp(-B)$,

$$\mathrm{MAE}(\mathrm{sgn}\hat{\mathcal{Z}},\mathrm{sgn}\bar{\Theta}) \lesssim t_n^{\frac{\alpha}{\alpha+2}} + \frac{1}{\rho^2}t_n,$$

where $t_n = \left(\frac{rd_{\max}[\sigma^2(\log d_{\max}+B)+\beta^2]}{|\Omega|}\right)$.

By balancing $B \approx \log d_{\max}$, with the probability at most $1-\exp\left(-(\sigma^2\log d_{\max}+\beta^2)rd_{\max}\right)-\frac{1}{d_{\max}}$,

$$\mathrm{MAE}(\mathrm{sgn}(\hat{\mathcal{Z}}),\mathrm{sgn}(\bar{\Theta})) \lesssim \left(\frac{rd_{\max}[\sigma^2\log d_{\max}+\beta^2]}{|\Omega|}\right)^{\frac{\alpha}{\alpha+2}} + \frac{1}{\rho^2(\pi,\mathcal{N})}\frac{rd_{\max}[\sigma^2\log d_{\max}+\beta^2]}{|\Omega|}.$$

Does it mean the final bound is quadratic in sigma?

Remark 1. If we set $\pi \in [-\beta-\sigma g(d_{\max}), \beta+\sigma g(d_{\max})]$ where $g\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{x\to\infty} g(x) = \infty$, we can not control bias term,

$$\mathbb{E}\left|\frac{1}{2H+1}\sum_{\pi\in\Pi}\mathrm{sgn}(\Theta-\pi)-\Theta\right|.$$

This is because the number of $\pi \in [-\beta,\beta]$ is finite while the number of entries such that $\pi \in [-\beta,\beta]^c$ converges to infinity. This makes our estimation highly biased. One direct solution for this unbounded error case is we assume that we know the range of signal $[-\beta,\beta]$ as we did in bounded noise case.

Consider when β is positive integer without loss of generality, then

I like these three remarks!

$$\begin{aligned}\mathrm{MAE}(\hat{\Theta},\Theta) &= \mathbb{E}\left|\frac{1}{2\beta H+1}\sum_{\pi\in\Pi}\mathrm{sgn}\hat{\mathcal{Z}}_\pi-\Theta\right| \\ &\le \frac{1}{2\beta H+1}\sum_{\pi\in\Pi}\mathrm{MAE}(\mathrm{sgn}\hat{\mathcal{Z}}_\pi,\mathrm{sgn}(\Theta-\pi)) + \mathbb{E}\left|\frac{1}{2\beta H+1}\sum_{\pi\in\Pi}\mathrm{sgn}(\Theta-\pi)-\Theta\right|\end{aligned}$$

Therefore, everything goes through including the algorithm and theory. In real application, we can set $\beta = \max_{\omega\in\Omega}|\mathcal{Y}(\omega)|$.

How does the final MAE look like?

Remark 2. We have three terms related to signal and noise: σ, α, β. I explain how the sign estimation is affected by these three factors.

- **The noise level** σ: The large noise level makes the estimation difficult as we can see σ in the numerater of t_n. The main reason the noise level makes the estimation harder is it increases the variance of risk difference of weighted classification.
- **The smooth parameter** α: The signal and noise ratio concept arise when we estimate $\mathrm{sgn}(\Theta-\pi)$ from $\mathrm{sgn}(\mathcal{Y}-\pi)$. The signal concept in this weighted classification is absorbed in the smooth parameter α, i.e. we quantify how strong the signal is by finding α such that $\mathbb{P}(|\Theta(\omega)-\pi|\le t) \le t^\alpha$ for $t \in (0,\rho(\pi,\mathcal{N})]$.

I like this part!

- **The range of the signal tensor** β: When we estimate $\mathrm{sgn}(\Theta-\pi)$ from $\mathrm{sgn}(\mathcal{Y}-\pi)$, the magnitude β such that $|\Theta(\omega)| \le \beta$ is location concept, not the signal concept. We have already consdered the signal level through the behavior $\Theta-\pi$ with α. Then how this range affects the sign estimation? My answer is the large range makes the variance of the risk difference large because of the weight in weighted classification loss. In the end, the large β makes the estimation hard.

Remark 3. When $\sigma = 0$, why the current bound is not zero? My answer is that it should not be 0 because there is still randomness from sampling distribution Π. That is why the variance of the risk difference is not zero even when $\Theta = \mathcal{Y}$. We assume that we have missing values in observed tensors through sampling distribution, which makes estimation not perfect even when $\sigma = 0$. In addition, the variance is highly influenced by β (which is the range of Θ) under the sampling distribution Π. This is why we still have the β term in the convergence rate t_n.

What if we observe all entries. Would the bound still be nonzero?

Lemma 1. Let $X_1, \ldots, X_n$ be independent sub-Gaussian random variables with σ^2. Then, for any $t > 0$,

$$\mathbb{P}\left\{ \max_{1 \le i \le n} |X_i| \ge \sqrt{2\sigma^2(\log n + t)} \right\} \le 2e^{-t}.$$

Proof. Notice that

$$\mathbb{P}[\max_{1 \le i \le n} X_i \ge u] \le \sum_{i=1}^{n} \mathbb{P}[X_i \le u] \le n e^{-\frac{u^2}{2\sigma^2}} = e^{-t},$$

where we set $u = \sqrt{2\sigma^2(\log n + t)}$. □

Q: should MAE depend on sigma in a linear way, or in a quadratic way?

Earlier calculation shows a quadratic effect in sigma.
However, if we apply the lemma to Y, we may conclude a linear effect.
1) first scale Y to Y/Y_\max ~ Y/(\beta+\sigma\sqrt(log n))
2) Y/Y_\max is roughly bounded, so original conclusion applies. This argument lead to a linear sigma effect.

Anything wrong with the second argument? or first argument?