

Assumption modification

Chanwoo Lee, November 25, 2020

1 New assumption

(A) [Boundary noise] There exist constants $\alpha \in (0, 1]$ and $C > 0$, such that

$$\max_{\pi \in \Pi'} \mathbb{P}_{\mathbf{X}}(|p(\mathbf{X}) - \pi| \leq t/H) \leq C \left(\frac{t}{H} \right)^{\frac{\alpha}{1-\alpha}}, \quad \text{for all } t \in [0, 1] \text{ and all } H \in \mathbb{N}_+, \quad (1)$$

where $\Pi' \subset \Pi$ with $|\Pi| - |\Pi'| \leq m$ and m is a finite number independent on H . When $\alpha = 1$, the inequality (1) reads $\mathbb{P}(|p(\mathbf{X}) - \pi| \leq t/H) = 0$.

Theorem 1.1. The condition (1) is equivalent to

$$\mathbb{P}_{\mathbf{X}}[S \Delta S_{\text{bayes}}] \leq \begin{cases} C_1 [R_{\pi}(S) - R_{\pi}(S_{\text{bayes}}(\pi))]^{\alpha}, & \text{if } \alpha \in (0, 1), \\ C_2 H [R_{\pi}(S) - R_{\pi}(S_{\text{bayes}}(\pi))] & \text{if } \alpha = 1, \end{cases} \quad (2)$$

for all sets $S \in \mathbb{R}^{d_1 \times d_2}$, levels $\pi \in \Pi'$, and $H = |\Pi| \in \mathbb{N}_+$.

Remark 1. What are changed?

I excluded the worst case scenario of π using new set Π' . For example, consider the constant probability, ability case $p(\mathbf{X}) = p$. Suppose that $\pi_i = \frac{i}{H} \leq p(\mathbf{x}) < \frac{i+1}{H} = \pi_{i+1}$ for some $i \in [H]$. In previous assumption, we cannot guarantee that $\mathbb{P}_{\mathbf{X}}(|p(\mathbf{X}) - \pi_i| \leq t/H) = 0$ and $\mathbb{P}_{\mathbf{X}}(|p(\mathbf{X}) - \pi_{i+1}| \leq t/H) = 0$ for all $t \in [0, 1]$ because we can set p i/H or arbitrarily close to π_i or π_{i+1} . Simply excluding this two π_i and π_{i+1} from Π setting $\Pi' = \Pi \setminus \{\pi_i, \pi_{i+1}\}$ can solve the problem without changing convergence rate because finite number of exclusion is negligible.

If we exclude the worst case $\pi \in \Pi$ by adopting Π' , then I start to believe that (1) is enough to characterize the most of probability function we are interested in. Theorem 1.1 shows that only when $\alpha = 1$, we need extra H to bound $\mathbb{P}(S \Delta S_{\text{bayes}})$. I think this happens because (1) has the different trend when $\alpha = 1$ by the term $\alpha/(1 - \alpha)$.

Remark 2. Can this assumption cover logistic link case?

Yes, consider the case $p(\mathbf{X}) = e^x/(1 + e^x)$ where $\mathbf{X} \sim N(0, 1)$. Let $\Pi = \{\frac{i}{H}\}_{i=1}^{H-1}$ and $\Pi' = \{\frac{i}{H}\}_{i=2}^{H-2}$. The worst case in Π' is when $\pi = \frac{H-2}{H}$, In this case where $t = 1$, we have

$$\begin{aligned} \mathbb{P}\left(\frac{H-3}{H} \leq p(\mathbf{X}) \leq \frac{H-1}{H}\right) &= \mathbb{P}\left(\log\left(\frac{H-3}{2}\right) \leq \mathbf{X} \leq \log(H-1)\right) \\ &= \int_{\log(\frac{H-3}{2})}^{\log(H-1)} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &\leq \frac{1}{\sqrt{2\pi}} e^{-\frac{[\log(\frac{H-3}{2})]^2}{2}} \underbrace{\left(\log\left(\frac{H-3}{2}\right) - \log(H-1)\right)}_{(*)} \\ &\leq \frac{1}{\sqrt{2\pi}} e^{-\frac{\log(\frac{H-3}{2})}{2}} \log\left(\frac{2(H-1)}{H-3}\right) \\ &\leq \frac{\log 3}{\sqrt{2\pi}} \sqrt{\frac{2}{H-3}} \\ &\leq C \frac{1}{\sqrt{H}}. \end{aligned}$$

Therefore, we have $\alpha = 1/3$. If we include $\pi = \frac{H-1}{H}$, the $(*)$ part cannot be calculated well. This might leads to use extra H terms but haven't checked yet. Therefore, introduction of new set Π' makes things easier.

Remark 3. Why do we need t/H instead of simple t ?

If we do not use t/H for bounding $|p(\mathbf{X}) - \pi| \leq t/H$, we cannot make it clear that $|\Pi'|$ is finitely different from $|\Pi|$ independently on H . By adopting t/H instead of t , we can successfully exclude the finite number of the worst $\pi \in \Pi$ which is closest to concentrated mass of $p(\mathbf{X})$. For example, consider $p(\mathbf{X}) = \sum_{i=1}^M \frac{i}{M} \mathbb{1}_{\mathbf{X} \in G_i}$, where $\{G_i\}_{i=1}^M$ are partition of $\mathbb{R}^{d_1 \times d_2}$. Then the the number of π 's, that are excluded from Π' is at most $2M$, which is negligible because $\frac{2M}{H} \rightarrow 0$ as $H \rightarrow \infty$.

Proof. For ease of notation, we drop the argument π from $S_{\text{bayes}}(\pi)$ and $R_\pi(\cdot)$, and simply write S_{bayes} and $R(\cdot)$, respectively. The following identity is useful to relate the excess risk and set difference in classifiers,

$$\begin{aligned} R(S) - R(S_{\text{bayes}}) &= \mathbb{E}_{\mathbf{X}, y} [w(y) \mathbb{1}(y \neq I(S))] - \mathbb{E}_{\mathbf{X}, y} [w(y) \mathbb{1}(y \neq I(S_{\text{bayes}}))] \\ &= \mathbb{E}_{\mathbf{X}} [(\pi - p(\mathbf{X})) (I(S) - I(S_{\text{bayes}}))] \\ &= 2 \int_{\mathbf{X} \in S \Delta S_{\text{bayes}}} |p(\mathbf{X}) - \pi| d\mathbb{P}_{\mathbf{X}}. \end{aligned} \quad (3)$$

Now we will use the identity (3) to show the equivalence between (1) and (2). We divide the proof into two cases: $\alpha \in (0, 1)$ and $\alpha = 1$.

Case 1: $\alpha \in (0, 1)$.

(1) \Rightarrow (2). Consider an arbitrary set $S \subset \mathbb{R}^{d_1 \times d_2}$. Let t be an arbitrary number in the interval $[0, 1]$, and define the set $A = \{\mathbf{X} : |p(\mathbf{X}) - \pi| > t/H\}$. Based on the inequality (A2),

$$\begin{aligned} \int_{\mathbf{X} \in S \Delta S_{\text{bayes}}} |p(\mathbf{X}) - \pi| d\mathbb{P}_{\mathbf{X}} &\geq \frac{t}{H} [\mathbb{P}_{\mathbf{X}}((S \Delta S_{\text{bayes}}) \cap A)] \\ &\geq \frac{t}{H} (\mathbb{P}_{\mathbf{X}}(S \Delta S_{\text{bayes}}) - \mathbb{P}_{\mathbf{X}}(A^c)) \\ &\geq \frac{t}{H} \left(\mathbb{P}_{\mathbf{X}}(S \Delta S_{\text{bayes}}) - C \left(\frac{t}{H} \right)^{\frac{\alpha}{1-\alpha}} \right), \quad \text{for all } t \in [0, 1]. \end{aligned}$$

Combining the above inequality with the identity (3) yields

$$R(S) - R(S_{\text{bayes}}) \geq \frac{2t}{H} \left(\mathbb{P}_{\mathbf{X}}(S \Delta S_{\text{bayes}}) - C \left(\frac{t}{H} \right)^{\frac{\alpha}{1-\alpha}} \right), \quad \text{for all } t \in [0, 1]. \quad (4)$$

We maximize the lower bound of (4) with respect to t and obtain the optimal $t_{\text{opt}} \in [0, 1]$,

$$t_{\text{opt}} = \begin{cases} 1, & \text{if } \mathbb{P}_{\mathbf{X}}(S \Delta S_{\text{bayes}}) \geq \frac{C}{1-\alpha} H^{\frac{\alpha}{1-\alpha}}, \\ \left[\frac{1-\alpha}{CH^{\frac{\alpha}{1-\alpha}}} \mathbb{P}_{\mathbf{X}}(S \Delta S_{\text{bayes}}) \right]^{\frac{1-\alpha}{\alpha}}, & \text{if } \mathbb{P}_{\mathbf{X}}(S \Delta S_{\text{bayes}}) < \frac{C}{1-\alpha} H^{\frac{\alpha}{1-\alpha}}. \end{cases}$$

Notice that for sufficiently large n and H , we always have $\mathbb{P}_{\mathbf{X}}(S \Delta S_{\text{bayes}}) < \frac{C}{1-\alpha} H^{\frac{\alpha}{1-\alpha}}$. When N and H are not large enough, we can rescale C to satisfy $\mathbb{P}_{\mathbf{X}}(S \Delta S_{\text{bayes}}) < \frac{C}{1-\alpha} H^{\frac{\alpha}{1-\alpha}}$. Therefore, we have

$$R(S) - R(S_{\text{bayes}}) \geq 2\alpha \left(\frac{1-\alpha}{C} \right)^{\frac{1-\alpha}{\alpha}} \mathbb{P}_{\mathbf{X}}^{\frac{1}{\alpha}}(S \Delta S_{\text{bayes}}).$$

Finally, for all $\pi \in \Pi'$ and $H \in \mathbb{N}_+$

$$\mathbb{P}_{\mathbf{X}}(S \Delta S_{\text{bayes}}) \leq C_1 (R(S) - R(S_{\text{bayes}}))^{\alpha},$$

where we take $C_1 = \left(\frac{C}{1-\alpha}\right)^{1-\alpha} \left(\frac{1}{2\alpha}\right)^\alpha > 0$.

(2) \Rightarrow (1). Let t be an arbitrary number in the interval $[0, 1]$, and define the set $S = \{\mathbf{X} : p(\mathbf{X}) \in [\pi - t/H, \pi] \cup (\pi + t/H, 1]\}$. The set S satisfies

$$S\Delta S_{\text{bayes}} = \{\mathbf{X} : |p(\mathbf{X}) - \pi| \leq t/H\}.$$

Based on (3) and the definition of S ,

$$R(S) - R(S_{\text{bayes}}) \leq \left(\frac{2t}{H}\right) \mathbb{P}_{\mathbf{X}}(S\Delta S_{\text{bayes}}).$$

Combining the above inequality with (2) gives

$$\mathbb{P}_{\mathbf{X}}(|p(\mathbf{X}) - \pi| \leq t/H) = \mathbb{P}_{\mathbf{X}}(S\Delta S_{\text{bayes}}) \leq C \left(\frac{t}{H}\right)^{\frac{\alpha}{1-\alpha}}, \quad (5)$$

assumed that the left hand side of (5) is non-zero (otherwise, the result is trivial), where we set $C = (C_1 2^\alpha)^{\frac{1}{1-\alpha}}$. Because the above inequality holds for all $t \leq 1$, $\pi \in \Pi'$, and $H \in \mathbb{N}_+$, (1) holds

Case 2: $\alpha = 1$.

(1) \Rightarrow (2). The inequality (4) now becomes

$$R(S) - R(S_{\text{bayes}}) \geq \frac{2t}{H} \mathbb{P}_{\mathbf{X}}(S\Delta S_{\text{bayes}}), \quad \text{for all } t \in [0, 1], \pi \in \Pi'.$$

Therefore the inequality (2) holds with $C_2 = \frac{1}{2}$ and $t = 1$.

(2) \Rightarrow (1). We replace C_2 by $\max(C_2, 1)$ in (2). The inequality (5) now becomes

$$\mathbb{P}_{\mathbf{X}}(S\Delta S_{\text{bayes}}) \leq \max(2C_2, 2)t\mathbb{P}_{\mathbf{X}}(S\Delta S_{\text{bayes}}), \quad \text{for all } t \in [0, 1].$$

In particular, the inequality holds for all $H \in \mathbb{N}_+$ and all $\pi \in \Pi'$ which implies

$$\max_{\pi \in \Pi} \mathbb{P}_{\mathbf{X}}(|p(\mathbf{X}) - \pi| \leq t/H) = 0, \quad \text{for all } t \in [0, 1].$$

Therefore the inequality (1) holds. □