Boundaries for different Prediction Error metrics

Jiaxin Hu

8.9 2019

1 Forbenius Norm

Consider the model $\mathbf{P}(Y=1) = f(\Theta \times_1 X)$, where f is the (inverse) link function. When X has the RIP property and $rank_T(\Theta) = \{r_1, \ldots, r_d\} \leq R$, we have the conclusion that:

$$\|\hat{\Theta} - \Theta_{true}\|_F \leq \frac{2L_\alpha}{\gamma_\alpha(1-\delta_R(X))} sup_{\mu \in \frac{\mathcal{P}-\mathcal{P}'}{|\mathcal{P}-\mathcal{P}'|}} \langle \mathcal{E}, \mu \rangle;$$
$$\|\hat{\Theta} \times_1 X - \Theta_{true} \times_1 X\|_F \leq \frac{2L_\alpha}{\gamma_\alpha} sup_{\mu \in \frac{\mathcal{P}-\mathcal{P}'}{|\mathcal{P}-\mathcal{P}'|}} \langle \mathcal{E}, \mu \rangle$$

where $\delta_R(X)$ is the isometry constant of X under the tucker rank assumption of Θ.

Now consider the prediction error $\|\mathbb{E}[\hat{Y}] - \mathbb{E}[Y]\|_F^2$. Due to the Bernoulli assumption on Y, it is natural that:

$$\|\mathbb{E}[\hat{Y}] - \mathbb{E}[Y]\|_F^2 = \|f(\Theta_{true} \times_1 X) - f(\hat{\Theta} \times_1 X)\|_F^2$$

Due to the subtraction is the entry-wise subtraction between tensor, consider the entry-wise Taylor Expansion of $f(\hat{\Theta} \times_1 X)$. Take π as any entry of tensor $\hat{\Theta} \times_1 X$ and π_t as the corresponding entry in $\Theta_{true} \times_1 X$:

$$f(\pi) = f(\pi_t) + f'(\tilde{\pi})(\pi - \pi_t); \quad \tilde{\pi} = a\pi + (1-a)\pi_t, \ a \in (0,1)$$

For $\Theta_{true} \times_1 X, \hat{\Theta} \times_1 X \in \mathcal{P}$, which means $|\pi_t|, |\pi| \leq \alpha$, we can define M_α as:

$$sup_{|\tilde{\pi}| \leq \alpha} |f'(\tilde{\pi})| \leq M_\alpha$$

Therefore,

$$
\begin{aligned}
&f(\pi_t) - f(\pi) \leq M_\alpha(\pi_t - \pi) \\
\Rightarrow\ &\|f(\Theta_{true} \times_1 X) - f(\hat{\Theta} \times_1 X)\|_F^2 \leq M_\alpha^2 \|\hat{\Theta} \times_1 X - \Theta_{true} \times_1 X\|_F^2 \\
\Rightarrow\ &\|\mathbb{E}[\hat{Y}] - \mathbb{E}[Y]\|_F \leq \frac{2L_\alpha M_\alpha}{\gamma_\alpha} sup_{\mu \in \frac{\mathcal{P}-\mathcal{P}'}{|\mathcal{P}-\mathcal{P}'|}} \langle \mathcal{E}, \mu \rangle
\end{aligned}
$$

2 K-L loss

Now consider the prediction error in K-L loss term: $KL(\mathbb{E}[Y] \| \mathbb{E}[\hat{Y}]) = KL(\mathbb{P}_{\Theta_{true}} \| \mathbb{P}_{\hat{\Theta}})$. Combine **Lemma 5** in *Wang 2019*:

$$
KL(X, Y) \leq \frac{(p-q)^2}{q(1-q)},
$$

where p, q are the parameter of Bernoulli random variable X, Y. Similarly with **Lemma 6** in *Wang 2019*, we can find the relationship between K-L loss and frobenius norm:

$$
\begin{aligned}
\prod_{k=1}^{K} d_k KL(\mathbb{P}_{\Theta_{true}} \| \mathbb{P}_{\hat{\Theta}}) &= \sum_{i_1,\dots,i_K} KL(Y_{i_1,\dots,i_K} | \pi_{t,i_1,\dots,i_K}, Y_{i_1,\dots,i_K} | \pi_{i_1,\dots,i_K}) \\
&\leq \sum_{i_1,\dots,i_K} \frac{(f(\pi_{t,i_1,\dots,i_K}) - f(\pi_{i_1,\dots,i_K}))^2}{f(\pi_{i_1,\dots,i_K}))(1 - f(\pi_{i_1,\dots,i_K})))}
\end{aligned}
$$

where $\pi_{t,i_1,\dots,i_K}$ refers to the entry of true parameter tensor $\Theta_{true} \times_1 X$ and $\pi_{i_1,\dots,i_K}$ refers to the entry of $\hat{\Theta} \times_1 X$.

According to the Taylor Expansion result and the notation M_α, we know that:

$$
f(\pi_{t,i_1,\dots,i_K}) - f(\pi_{i_1,\dots,i_K})^2 \leq M_\alpha^2 (\pi_{t,i_1,\dots,i_K} - \pi_{i_1,\dots,i_K})^2
$$

Therefore we get:

$$
\begin{aligned}
\prod_{k=1}^{K} d_k KL(\mathbb{P}_{\Theta_{true}} \| \mathbb{P}_{\hat{\Theta}}) &\leq \sum_{i_1,\dots,i_K} \frac{M_\alpha^2}{f(\pi_{i_1,\dots,i_K}))(1 - f(\pi_{i_1,\dots,i_K})))} (\pi_{t,i_1,\dots,i_K} - \pi_{i_1,\dots,i_K})^2 \\
&\leq 4M_\alpha^2 \|\hat{\Theta} \times_1 X - \Theta_{true} \times_1 X\|_F^2 \\
&\leq \frac{16 L_\alpha^2 M_\alpha^2}{\gamma_\alpha^2} sup^2_{\mu \in \frac{\mathcal{P}-\mathcal{P}'}{|\mathcal{P}-\mathcal{P}'|}} \langle \mathcal{E}, \mu \rangle
\end{aligned}
$$

3 Hellinger loss

Consider the prediction error in Hellinger loss: $d_H^2(\mathbb{E}[Y], \mathbb{E}[\hat{Y}]) = d_H^2(\mathbb{P}_{\Theta_{true}}, \mathbb{P}_{\hat{\Theta}})$, where the definition of Hellinger loss is:

$$\begin{aligned}
&d_H^2(p,q) = (\sqrt{p} - \sqrt{q})^2 + (\sqrt{1-p} - \sqrt{1-q})^2 \\
&1 - \frac{1}{2} d_H^2(p,q) = \sqrt{pq} + \sqrt{(1-p)(1-q)}
\end{aligned}$$

Consider the relationship between K-L divergence and Hellinger distance:

$$\begin{aligned}
-\frac{1}{2}KL(p\|q) &= -\frac{1}{2}p\log(\frac{p}{q}) - \frac{1}{2}(1-p)\log(\frac{1-p}{1-q}) \\
&= p\log(\sqrt{\frac{q}{p}}) + (1-p)\log(\sqrt{\frac{1-q}{1-p}}) \\
for\ \ \log(x) \leq x-1,\ \ x>0 \quad &\leq p(\sqrt{\frac{q}{p}} - 1) + (1-p)(\sqrt{\frac{1-q}{1-p}} - 1) \\
&= \sqrt{pq} + \sqrt{(1-p)(1-q)} - 1 \\
&= -\frac{1}{2}d_H^2(p,q)
\end{aligned}$$

Therefore,

$$d_H^2(\mathbb{P}_{\Theta_{true}}, \mathbb{P}_{\hat{\Theta}}) \leq KL(\mathbb{P}_{\Theta_{true}} \| \mathbb{P}_{\hat{\Theta}}) \leq \frac{16L_\alpha^2 M_\alpha^2}{\gamma_\alpha^2 \prod_{k=1}^K d_k} sup^2_{\mu \in \frac{\mathcal{P}-\mathcal{P}'}{|\mathcal{P}-\mathcal{P}'|}} \langle \mathcal{E}, \mu \rangle$$