Estimation and Prediction Error in Supervised Setting

Jiaxin Hu Zhuoyan Xu

8.10 2019

1 Estimation Error

Considering the covariate X is already known, we can sharpen the boundary of estimate error.

Suppose $\mathbf{P}(Y=1)=f(\Theta\times_1 X)$, $X\in\mathbb{R}^{d_1\times p}, d_1>p, rank(X)=p$ is the given covariate matrix. $\Theta\in\mathcal{P}^*$,where

$$\mathcal{P}=\{\Theta: rank(\Theta)=R_T, \|\Theta\|_\infty\leq\alpha\}$$
$$\Theta\times_1 X\in\mathcal{P}\Leftrightarrow\Theta\in\mathcal{P}^*$$

Let $\hat{\Theta}=arg\max_{\Theta\in\mathcal{P}^*}\mathcal{L}_Y(\Theta)$ where $\mathcal{L}_Y(\Theta)$ is the log-likelihood of parameter Θ. Assume X satisfies the RIP property, i.e. there exists an isometry constant of X which is the smallest number $\delta_R(X)$ guarantees the following inequality holds for all Θ with Tucker rank at most R $=\max\{r_1,\ldots,r_K\}$:

$$(1-\delta_R(X))\|\Theta\|_F^2\leq\|\Theta\times_1 X\|_F^2\leq(1+\delta_R(X))\|\Theta\|_F^2.$$

Continue the discussion in the second line from the bottom on page 3 of *Boundaries with Gaussian Width*:

$$0\leq\langle\mathcal{S}_Y^*(\Theta_{true}\times_1 X),(\Theta-\Theta_{true})\times_1 X\rangle-\frac{\gamma_\alpha}{2}\|(\hat{\Theta}-\Theta_{true})\times_1 X\|_F^2$$
$$\|(\hat{\Theta}-\Theta_{true})\times_1 X\|_F^2\leq\frac{2L_\alpha}{\gamma_\alpha}\langle L_\alpha^{-1}\mathcal{S}_Y^*(\Theta_{true}\times_1 X),(\Theta-\Theta_{true})\times_1 X\rangle$$

Use $\mathcal{S}$ to denote $L_\alpha^{-1}\mathcal{S}_Y^*(\Theta_{true} \times_1 X)$, then we have:

$$\begin{aligned}
\langle L_\alpha^{-1}\mathcal{S}_Y^*(\Theta_{true} \times_1 X), (\Theta - \Theta_{true}) \times_1 X\rangle = \langle \mathcal{S}, (\Theta - \Theta_{true}) \times_1 X\rangle \\
= &\langle \mathcal{S} \times_1 X^T, (\Theta - \Theta_{true})\rangle \\
= &\|X\|_\infty \langle \mathcal{S} \times_1 \frac{X^T}{\|X\|_\infty}, (\Theta - \Theta_{true})\rangle \\
= &\|X\|_\infty \langle \mathcal{E}, (\Theta - \Theta_{true})\rangle
\end{aligned}$$

Since $\forall s \in \mathcal{S}$, where s denote any entry in $\mathcal{S}$, we have

$$\mathbb{E}(s) = 0, \quad |s| \le 1$$

Then $\forall \epsilon \in \mathcal{E}$, where ϵ denote any entry in $\mathcal{E}$, we have

$$\mathbb{E}(\epsilon) = 0, \quad |\epsilon| \le 1$$

According to our bounds on Gaussian width, we have:

$$\langle \mathcal{E}, (\Theta - \Theta_{true})\rangle \le C_2 \sqrt{\sum_{k=2}^{K} r_k (\sum_{k=2}^{K} d_k + p)} \|\hat{\Theta} - \Theta_{\text{true}}\|_F$$

Thus, we have:

$$\begin{aligned}
\left\| \left(\hat{\Theta} - \Theta_{\text{true}}\right) \times_1 X \right\|_F^2 \le &\frac{2L_\alpha}{\gamma_\alpha} \|X\|_\infty \langle \mathcal{E}, (\Theta - \Theta_{\text{true}})\rangle && (1) \\
\le &\frac{2L_\alpha C_2}{\gamma_\alpha} \|X\|_\infty \sqrt{\sum_{k=2}^{K} r_k (\sum_{k=2}^{K} d_k + p)} \|\hat{\Theta} - \Theta_{\text{true}}\|_F && (2)
\end{aligned}$$

Then using the RIP property, we can conclude the boundary of estimate error is:

$$\begin{aligned}
\|(\hat{\Theta} - \Theta_{true})\|_F^2 &\le \frac{1}{1 - \delta_R(X)} \|(\hat{\Theta} - \Theta_{true}) \times_1 X\|_F^2 \\
&\le \frac{2L_\alpha C_2 \|X\|_\infty}{\gamma_\alpha (1 - \delta_R(X))} \sqrt{\sum_{k=2}^{K} r_k (\sum_{k=2}^{K} d_k + p)} \|\hat{\Theta} - \Theta_{\text{true}}\|_F \\
\Rightarrow \quad \|(\hat{\Theta} - \Theta_{true})\|_F &\le \frac{2L_\alpha C_2 \|X\|_\infty}{\gamma_\alpha (1 - \delta_R(X))} \sqrt{\sum_{k=2}^{K} r_k (\sum_{k=2}^{K} d_k + p)}
\end{aligned}$$

2 Prediction Error

According to RIP, we have:

$$\|(\hat{\Theta} - \Theta_{true})\|_F \leq \frac{1}{\sqrt{1 - \delta_R(X)}} \|(\hat{\Theta} - \Theta_{true}) \times_1 X\|_F$$

According to (2),

$$\begin{aligned}
\left\| \left(\hat{\Theta} - \Theta_{\text{true}} \right) \times_1 X \right\|_F^2 &\leq \frac{2L_\alpha C_2}{\gamma_\alpha} \|X\|_\infty \sqrt{\sum_{k=2}^{K} r_k (\sum_{k=2}^{K} d_k + p)} \|\hat{\Theta} - \Theta_{\text{true}}\|_F \\
&\leq \frac{2L_\alpha C_2}{\gamma_\alpha} \|X\|_\infty \sqrt{\sum_{k=2}^{K} r_k (\sum_{k=2}^{K} d_k + p)} \frac{1}{\sqrt{1 - \delta_R(X)}} \|(\hat{\Theta} - \Theta_{true}) \times_1 X\|_F \\
\left\| \left(\hat{\Theta} - \Theta_{\text{true}} \right) \times_1 X \right\|_F &\leq \frac{2L_\alpha C_2}{\gamma_\alpha \sqrt{1 - \delta_R(X)}} \|X\|_\infty \sqrt{\sum_{k=2}^{K} r_k (\sum_{k=2}^{K} d_k + p)}
\end{aligned}$$

According to the Taylor Expansion, we can conclude the prediction error in Frobenius term is:

$$\begin{aligned}
\|\mathbb{E}[\hat{Y}] - \mathbb{E}[Y]\|_F &= \|f(\Theta_{true} \times_1 X) - f(\hat{\Theta} \times_1 X)\|_F \\
&\leq \frac{2L_\alpha C_2 M_\alpha \|X\|_\infty}{\gamma_\alpha \sqrt{1 - \delta_R(X)}} \sqrt{\sum_{k=2}^{K} r_k (\sum_{k=2}^{K} d_k + p)}
\end{aligned}$$

Similarly, we can get the loss in K-L loss and Hellinger distance.