Boundaries with Gaussian Width

Zhuoyan Xu, Jiaxin Hu

8.8 2019

1 Boundaries for unsupervised generalized model

Suppose $\mathbf{P}(Y=1)=f(\Theta)$ and $\Theta \in \mathcal{P}$,where

$$\mathcal{P}=\{\Theta : rank(\Theta)=R_T, \|\Theta\|_\infty \leq \alpha\}.$$

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. Then we have:

$$\|\hat{\Theta}-\Theta_{true}\|_F \leq \frac{2L_\alpha}{\gamma_\alpha} sup_{\mu\in\frac{\mathcal{P}-\mathcal{P}'}{|\mathcal{P}-\mathcal{P}'|}}\langle\mathcal{E},\mu\rangle$$

where every entry of $\mathcal{E}$ i.i.d follows $sG(1)$. L_α and γ_α are defined in the paper *Wang 2019*.
Proof: Apply Taylor Expansion of the log-likelihood:

$$\mathcal{L}(\hat{\Theta})=\mathcal{L}(\Theta_{true})+\langle\mathcal{S}_Y(\Theta_{true}),\Theta-\Theta_{true}\rangle+\frac{1}{2}vec(\Theta-\Theta_{true})^T\mathcal{H}_Y(\tilde{\Theta})vec(\Theta-\Theta_{true})$$

Then according to $\mathcal{L}_Y(\hat{\Theta})\geq\mathcal{L}_Y(\Theta_{true})$:

$$\begin{aligned}
0 &\leq \langle\mathcal{S}_Y(\Theta_{true}),\Theta-\Theta_{true}\rangle-\frac{\gamma_\alpha}{2}\|\hat{\Theta}-\Theta_{true}\|_F^2\\
\|\hat{\Theta}-\Theta_{true}\|_F &\leq \frac{2L_\alpha}{\gamma_\alpha}sup\langle L_\alpha^{-1}\mathcal{S}_Y(\Theta_{true}),\frac{\Theta-\Theta_{true}}{\|\hat{\Theta}-\Theta_{true}\|_F}\rangle\\
&\leq \frac{2L_\alpha}{\gamma_\alpha} sup_{\mu\in\frac{\mathcal{P}-\mathcal{P}'}{|\mathcal{P}-\mathcal{P}'|}}\langle\mathcal{E},\mu\rangle
\end{aligned}$$

2 Boundaries for semi-supervised generalized model

2.1 Using RIP property

Consider we have an extra covariate matrix $X^{d1\times p}$(accounting for features), which contains the information of countries. We want to connect the membership matrix(or factor matrix) A and B with the information in tensor X.

The general form is:

$$
\begin{aligned}
\text{logit}\left\{E\left[\mathcal{Y}^{d_1d_2\ldots d_K}\right]\right\} &= \Theta = \mathcal{G}^{r_1r_2\ldots r_K} \times_1 W^{d_1r_1} \times_2 N_2^{d_2r_2} \ldots \times_K N_K^{d_Kr_K} \\
W^{d_1r_1} &= X^{d_1p}N_1^{pr_1}
\end{aligned}
$$

where $\mathcal{G}$ is the low rank core tensor of factorization. $W, N_2, \ldots, N_K$ are factor matrices. N_1 is the regression coefficient matrix for X on W. Under this scenario, $p > d1$.

We can write down the model in another view, which helps to compute:

$$
\begin{aligned}
\Theta =& \mathcal{C} \times_1 X^{d_1p} \\
\mathcal{C} =& \mathcal{G}^{r_1r_2\ldots r_K} \times_1 N_1^{d_1r_1} \times_2 N_2^{d_2r_2} \ldots \times_K N_K^{d_Kr_K}
\end{aligned}
$$

where $\mathcal{C}$ is tensor with tucker rank $(r_1, \ldots, r_K)$.

Definition (Restricted Isometry Property). The isometry constant of X is the smallest number δ_R such as the following holds for all C with Tucker rank at most R = $\max\{r_1, \ldots, r_K\}$.

$$
(1-\delta_R)\|\mathcal{C}\|_F^2 \leq \|\mathcal{C} \times_1 X\|_F^2 \leq (1+\delta_R)\|\mathcal{C}\|_F^2
$$

Thus:

$$
\|\hat{\mathcal{C}} - \mathcal{C}\|_F^2 \leq \frac{1}{1-\delta_R}\|(\hat{\mathcal{C}} - \mathcal{C})_{\times 1}X\|_F^2 = \frac{1}{1-\delta_R}\|(\hat{\Theta} - \Theta)\|_F^2
$$

Define:

$$
\begin{aligned}
\hat{\Theta} =& \hat{\mathcal{C}} \times_1 X_1 \\
\Theta_{\text{true}} =& \mathcal{C}_{\text{true}} \times_1 X_1
\end{aligned}
$$

we have:

$$
\begin{aligned}
0 \leq & \mathcal{L}_{\mathcal{Y}}(\hat{\Theta})-\mathcal{L}_{\mathcal{Y}}\left(\Theta_{\text{true}}\right) \\
& =\left\langle S_{\mathcal{Y}}\left(\Theta_{\text{true}}\right), \hat{\Theta}-\Theta_{\text{true}}\right\rangle+\frac{1}{2} \operatorname{vec}\left(\hat{\Theta}-\Theta_{\text{true}}\right)^{T} \mathcal{H}_{\mathcal{Y}}(\check{\Theta}) \operatorname{vec}\left(\hat{\Theta}-\Theta_{\text{true}}\right) \\
& \leq\left\langle S_{\mathcal{Y}}\left(\Theta_{\text{true}}\right), \hat{\Theta}-\Theta_{\text{true}}\right\rangle-\frac{\gamma_{\alpha}}{2}\left\|\hat{\Theta}-\Theta_{\text{ true}}\right\|_{F}^{2}
\end{aligned}
$$

Apply our result of Gaussian width, we have:

$$
\|\hat{\mathcal{C}}-\mathcal{C}\|_{F}^{2} \leq \frac{1}{1-\delta_{R}} 2 C_{2} \frac{L_{\alpha}}{\gamma_{\alpha}} \sqrt{\prod_{k=1}^{K-1} r_{k} \sum_{k=1}^{K} d_{k}}
$$

2.2 Without using RIP

Suppose $\mathbf{P}(Y=1)=f(\Theta \times_{1} X)$, $X \in R^{d_{1} \times p}$ is the covariate matrix(n ¿ p) and rank(X) = p .$\Theta \in \mathcal{P}^{*}$,where

$$
\begin{aligned}
& \mathcal{P}=\left\{\Theta: rank(\Theta)=R_{T},\|\Theta\|_{\infty} \leq \alpha\right\} \\
& \Theta \times_{1} X \in \mathcal{P} \Leftrightarrow \Theta \in \mathcal{P}^{*}
\end{aligned}
$$

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. Then we have:

$$
\|\hat{\Theta}-\Theta_{true}\|_{F} \leq \frac{2 L_{\alpha} \kappa(X)}{\gamma_{\alpha}\|X\|_{2}} sup_{\mu \in \frac{\mathcal{P}-\mathcal{P}^{\prime}}{|\mathcal{P}-\mathcal{P}^{\prime}|}}\langle\mathcal{E}, \mu\rangle
$$

where every entry of $\mathcal{E}$ i.i.d follows $sG(1)$. L_{α} and γ_{α} are defined in the paper *Wang 2019*; $\kappa(X)$ is the condition number of X; $\|X\|_{2}$ is the spectral norm of X.

Proof: It is obvious that:

$$
\mathcal{L}_{Y}(\Theta)=\sum\left[I_{Y=1} log\left(f\left(\Theta \times_{1} X\right)\right)+I_{Y=0} log\left(1-f\left(\Theta \times_{1} X\right)\right)\right]
$$

Define:

$$
\tilde{\mathcal{L}}_{Y}\left(\Theta^{*}\right)=\sum\left[I_{Y=1} log\left(f\left(\Theta^{*}\right)\right)+I_{Y=0} log\left(1-f\left(\Theta^{*}\right)\right)\right]
$$

Due the MLE $\hat{\Theta} = arg\max_{\Theta\in\mathcal{P}^*} \mathcal{L}_Y(\Theta)$:

$$\mathcal{L}_Y(\hat{\Theta}) \geq \mathcal{L}_Y(\Theta_{true}) \Leftrightarrow \tilde{\mathcal{L}}_Y(\hat{\Theta} \times_1 X) \geq \tilde{\mathcal{L}}_Y(\Theta_{true} \times_1 X)$$

Define $\mathcal{S}_Y^* = \tilde{\mathcal{L}}'_Y, \mathcal{H}_Y^* = \tilde{\mathcal{L}}''_Y$. Apply Taylor Expansion on $\tilde{\mathcal{L}}_Y(\hat{\Theta} \times_1 X)$:

$$\begin{aligned}\tilde{\mathcal{L}}_Y(\hat{\Theta} \times_1 X) &= \tilde{\mathcal{L}}_Y(\Theta_{true} \times_1 X) + \langle \mathcal{S}_Y^*(\Theta_{true} \times_1 X), (\Theta - \Theta_{true}) \times_1 X\rangle \\ &+\frac{1}{2} vec((\Theta - \Theta_{true}) \times_1 X)^T \mathcal{H}_Y^*(\tilde{\Theta} \times_1 X) vec((\Theta - \Theta_{true}) \times_1 X)\end{aligned}$$

Because $\Theta \times_1 X \in \mathcal{P}$, using the notation in unsupervised case:

$$\mathcal{S}_Y^*(\Theta_{true} \times_1 X) \leq L_\alpha; \quad \mathcal{H}_Y^*(\tilde{\Theta} \times_1 X) \leq -\gamma_\alpha$$

Therefore:

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

According to the **Theorem 6** in *Bo Jiang, et al 2016*, we have:

$$c\|(\hat{\Theta} - \Theta_{true})\|_F \leq \|(\hat{\Theta} - \Theta_{true}) \times_1 X\|_F$$

where $c = \|X\|_2/\kappa(X)$ in our setting. That would lead to:

$$\|\hat{\Theta} - \Theta_{true}\|_F \leq \frac{2L_\alpha \kappa(X)}{\gamma_\alpha \|X\|_2} sup_{\mu \in \frac{\mathcal{P}-\mathcal{P}'}{|\mathcal{P}-\mathcal{P}'|}} \langle \mathcal{E}, \mu\rangle.$$

3 Algorithm Simulation

When we conduct the constrained optimization in unsupervised model, we implement the algorithm in the paper you assigned us *[1-Bit Matrix Completion under Exact Low-Rank Constraint]*. The object function is the cross entropy plus penalty. And we found it must be solved by some solvers in R instead of GLM. And the log-Likelihood is not non-decreasing. We're still working on it.