Supplementary Materials for ”LambdaMF: Learning Nonsmooth Ranking Functions in Matrix Factorization Using Lambda”

Guang-He Lee
Department of Computer Science and Information Engineering
National Taiwan University
b00902055@csie.ntu.edu.tw

Shou-De Lin
Department of Computer Science and Information Engineering
National Taiwan University
sdlin@csie.ntu.edu.tw

Formally speaking, let us first denote the set of users observing $\hat{i}$ in the training set as $Rel(\hat{i})$, the rating of $\hat{i}$ for user k as $R_{k,\hat{i}}$, the latent factors in the iteration t as U^t/V^t.

Theorem 1 (Popular Item Theorem): If there exists an item $\hat{i}$, such that for all users $k \in Rel(\hat{i})$, $R_{k,\hat{i}} \geq R_{k,j}$ for all other observed item j of user k. Furthermore, if after certain iteration τ, latent factors of all users $k \in Rel(\hat{i})$ converge to certain extent. That is, there exists a vector $\bar{U}^t$ such that for all $k \in Rel(\hat{i})$ in all iteration $t > \tau$, inner-product$(U_k^t, \bar{U^\tau}) > 0$. Then the norm of $V_{\hat{i}}$ will eventually grow to infinity for any MF model satisfying the constraint that $\frac{\partial C^u_{\hat{i},j}}{\partial s_{\hat{i}}} > 0$ for all j with $R_{u,\hat{i}} > R_{u,j}$, as shown below:

$$\lim_{n\to\infty} \|V_{\hat{i}}^n\|^2 = \infty$$

Proof: Given latent factor space with D dimensions, there exists $D-1$ vectors $\overrightarrow{c_2}$,$\overrightarrow{c_3}$,...,$\overrightarrow{c_D}$ and $\overrightarrow{c_1} = \bar{U^\tau}$, such that they are mutually orthogonal.

Denote $\frac{\partial C^k_{\hat{i}}}{\partial s_{\hat{i}}}(t) = \sum_{j\in R_k} \frac{\partial C^k_{\hat{i},j}}{\partial s_{\hat{i}}}$ in iteration t, then for any iteration $n > \tau$,

$$V_{\hat{i}}^n = V_{\hat{i}}^\tau + \sum_{n\geq t>\tau} \sum_{k\in Rel(\hat{i})} \eta \frac{\partial C^k_{\hat{i}}}{\partial s_{\hat{i}}}(t) U_k^t$$

Then we perform coordinate axis transform on $V_{\hat{i}}^\tau$ and U_k^t to $c_1, ..., c_D$.

$$\Rightarrow V_{\hat{i}}^n = \alpha_{\hat{i}}^\tau(1)\overrightarrow{c_1} + ... + \alpha_{\hat{i}}^\tau(D)\overrightarrow{c_D} + \sum_{n\geq t>\tau} \sum_{k\in Rel(\hat{i})} \eta \frac{\partial C^k_{\hat{i}}}{\partial s_{\hat{i}}}(t)(\beta_k^t(1)\overrightarrow{c_1} + ... + \beta_k^t(D)\overrightarrow{c_D})$$

We have $V_{\hat{i}}^\tau = \alpha_{\hat{i}}^\tau(1)\overrightarrow{c_1} + ... + \alpha_{\hat{i}}^\tau(D)\overrightarrow{c_D}$ and $U_k^t = \beta_k^t(1)\overrightarrow{c_1} + ... + \beta_k^t(D)\overrightarrow{c_D}$; η, $\frac{\partial C^k_{\hat{i}}}{\partial s_{\hat{i}}}(t) > 0$, $\beta_k^t(1) > 0$ as inner-product$(U_k^t, \overrightarrow{c_1})$ = inner-product$(U_k^t, \bar{U^\tau}) > 0$, and all other variables $\in \mathbb{R}$.

$$\Rightarrow V_{\hat{i}}^n = \alpha_{\hat{i}}^\tau(1)\overrightarrow{c_1} + ... + \alpha_{\hat{i}}^\tau(D)\overrightarrow{c_D} + \sum_{n\geq t>\tau} (\gamma^t(1)\overrightarrow{c_1} + ... + \gamma^t(D)\overrightarrow{c_D})$$

$\gamma^t(d) = \sum_{k\in Rel(\hat{i})} \eta \frac{\partial C^k_{\hat{i}}}{\partial s_{\hat{i}}}(t)\beta_k^t(d)$ for integer $d \in [1, D]$; $\gamma^t(1) > 0$. Then since $\overrightarrow{c_1}$,...,$\overrightarrow{c_D}$ are mutually orthogonal,

$$\begin{aligned}\Rightarrow \lim_{n\to\infty} \|V_{\hat{i}}^n\|^2 &= \lim_{n\to\infty} (\alpha_{\hat{i}}^\tau(1) + \sum_{n\geq t>\tau} \gamma^t(1))^2\|\overrightarrow{c_1}\|^2 + ... \\ &\quad + (\alpha_{\hat{i}}^\tau(D) + \sum_{n\geq t>\tau} \gamma^t(D))^2\|\overrightarrow{c_D}\|^2 \\ &\geq \lim_{n\to\infty} (\alpha_{\hat{i}}^\tau(1) + \sum_{n\geq t>\tau} \gamma^t(1))^2\|\overrightarrow{c_1}\|^2\end{aligned}$$

And we have

$$\begin{aligned}&\lim_{n\to\infty} (\alpha_{\hat{i}}^\tau(1) + \sum_{n\geq t>\tau} \gamma^t(1)) \\ &\geq \lim_{n\to\infty} (\alpha_{\hat{i}}^\tau(1) + \sum_{n\geq t>\tau} min_{n\geq t>\tau}\gamma^t(1)) \\ &= \lim_{n\to\infty} (\alpha_{\hat{i}}^\tau(1) + (n-\tau)min_{n\geq t>\tau}\gamma^t(1)) \\ &= \infty\end{aligned}$$

Finally,

$$\Rightarrow \lim_{n\to\infty} \|V_{\hat{i}}^n\|^2 = \infty$$

■