Probabilistic Graphical Model

Final Project.

—

Ali Izadi

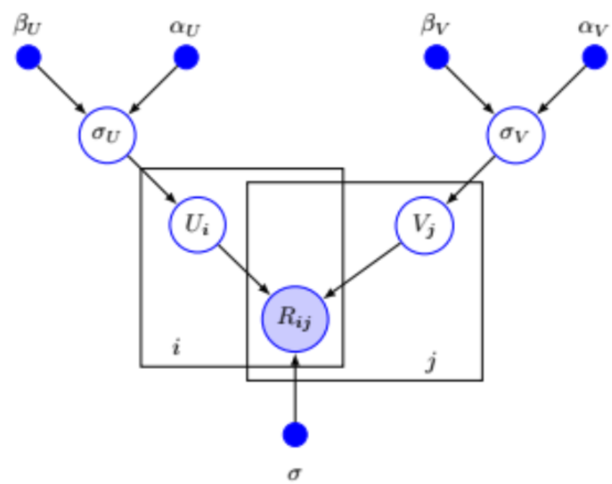
810199102

فهرست مطالب

بخش اول: رویکرد Bayesian

هدف از این بخش یادگیری دو ماتریس U و V است. بنابراین نیاز به محاسبه توزیع posterior زیر داریم که مطابق با مدل گرافیگی احتمالی آن به صورت زیر factorize میشود:

$$P(U, V \mid R) = P(R \mid U, V, \sigma) . P(U \mid \sigma_U) . P(V \mid \sigma_V) . P(\sigma_U \mid \alpha_U, \beta_U) . P(\sigma_V \mid \alpha_V, \beta_V)$$



دو ماتریس U و V را با استفاده از gibbs sampling تخمین میزنیم.

برای این کار نیاز به محاسبه احتمالات complete conditional متغیرهای نهان داریم. یعنی محاسبه دو ماتریس U و V و σهای U و V.

- ابتدا به محاسبه complete conditional ماتریس U میپردازیم. چون U_iها از هم مستقل اند احتمال شرطی U_i به شرط بقیه متغییر ها را مطابق با زیر محاسبه میکنیم

$$P(U_i \mid R, V, \sigma, \sigma_U, \sigma_V, \alpha_U, \beta_U, \alpha_V, \beta_V) = \frac{P(U_i, R, V, \sigma, \sigma_U, \sigma_V, \alpha_U, \beta_U, \alpha_V, \beta_V)}{\sum_{U_i} P(U_i, R, V, \sigma, \sigma_U, \sigma_V, \alpha_U, \beta_U, \alpha_V, \beta_V)}$$

- توزیع joint را مطابق با مدل احتمالاتی factorize میکنیم و آن متغیرهایی که به U_i وابسته نیستند را از صورت و مخرج حذف میکنیم.

$$
\begin{aligned}
P(U_i \,|\, R, V, \sigma, \sigma_U, \sigma_V, \alpha_U, \beta_U, \alpha_V, \beta_V) &= \frac{P(U_i, R, V, \sigma, \sigma_U, \sigma_V, \alpha_U, \beta_U, \alpha_V, \beta_V)}{\sum_{U_i} P(U_i, R, V, \sigma, \sigma_U, \sigma_V, \alpha_U, \beta_U, \alpha_V, \beta_V)} \\
&= \frac{P(R \,|\, U_i, V, \sigma) \,.\, P(U_i \,|\, \sigma_U)}{\sum_{U_i} P(R \,|\, U_i, V, \sigma) \,.\, P(U_i \,|\, \sigma_U)} \\
&\propto P(R \,|\, U_i, V, \sigma) \,.\, P(U_i \,|\, \sigma_U)
\end{aligned}
$$

- در نتیجه complete conditional متناسب با توزیع بالا به دست می آید.

حال توزیع به دست آمده را بسط میدهیم(روابط توزیع نرمال را مینویسم و توان دوی به دست آمده را نیز باز میکنیم.) I_ij نیز به این معناست که کاربر i به jامین item امتیاز داده است.

$$
\begin{aligned}
\log P(R \,|\, U_i, V, \sigma) \,.\, P(U_i \,|\, \sigma_U) &= \log\left(\left(\prod_{j=1}^{M} N\big(R_{ij} \,\big|\, U_i^T V_j, \sigma\big)^{I_{ij}}\right).\; N(U_i \,|\, 0, \sigma_U I)\right) \\
&\propto \left(\sum_{j=1}^{M} -\frac{I_{ij}}{2\sigma}\big(R_{ij} - U_i^T V_j\big)^2\right) - \frac{1}{2} U_i^T (\sigma_U I)^{-1} U_i \\
&= \left(\sum_{j=1}^{M} -\frac{I_{ij}}{2\sigma}\Big(R_{ij}^2 + \big(U_i^T V_j\big)^2 - 2R_{ij} U_i^T V_j\Big)\right) - \frac{1}{2} U_i^T (\sigma_U I)^{-1} U_i
\end{aligned}
$$

- حال فرم quadratic را نوشته و تنها متغیرهای وابسته به U_i را نگه میداریم زیرا complete conditional بر حسب U_i است.

$$
\begin{aligned}
\log P(R \,|\, U_i, V, \sigma) \,.\, P(U_i \,|\, \sigma_U) &\propto \left(\sum_{j=1}^{M} -\frac{I_{ij}}{2\sigma}\Big(\cancel{R_{ij}^2} + \big(U_i^T V_j\big)^2 - 2R_{ij} U_i^T V_j\Big)\right) - \frac{1}{2} U_i^T (\sigma_U I)^{-1} U_i \\
&= \left(\sum_{j=1}^{M} -\frac{I_{ij}}{2\sigma}\big(U_i^T V_j V_j^T U_i\big)\right) - \frac{1}{2} U_i^T (\sigma_U I)^{-1} U_i + \left(\sum_{j=1}^{M} -\frac{I_{ij}}{2\sigma}\big(-2R_{ij} U_i^T V_j\big)\right)
\end{aligned}
$$

- چون فرم به دست آمده quadratic است پس از یک توزیع نرمال پیروی میکند و بنابراین توزیع نرمال complete conditional برای U_i مطابق با زیر به دست می آید.

$$U_i \sim N(\mu, \Sigma)$$

$$\Sigma^{-1} = \left((\sigma_U I)^{-1} + \frac{1}{\sigma} \sum_{j=1}^{M} I_{ij} V_j V_j^T \right)$$

$$\mu = \Sigma \,.\, \left(\frac{1}{\sigma} \sum_{j=1}^{M} I_{ij} \big(R_{ij} V_j^T \big) \right)$$

- مشابه با آن چه برای محاسبه U_i به دست آمد روابط زیر برای V_j نیز به دست می آید که از یک توزیع نرمال مطابق با زیر پیروی میکند.

$$V_j \sim N(\mu, \Sigma)$$

$$\Sigma^{-1} = \left((\sigma_V I)^{-1} + \frac{1}{\sigma} \sum_{i=1}^{N} I_{ij} U_i U_i^T \right)$$

$$\mu = \Sigma \,.\, \left(\frac{1}{\sigma} \sum_{i=1}^{N} I_{ij} \big(R_{ij} U_i^T \big) \right)$$

- حال از متغیر های نهان باقیمانده به سراغ محاسبه complete conditional سیگمای U میرویم. در رابطه های پایین ~ به معنای بقیه متغیر هاست.

مطابق با قبل توزیع بر اساس مدل احتمالی factorize میشود و هم چنین توزیع هایی که به سیگمای U وابسته نیستند از صورت و مخرج حذف میشوند.

$$
\begin{aligned}
P(\sigma_U \mid \sim) &= \frac{P(\sigma_U,\ \sim)}{\sum_{\sigma_U} P(\sigma_U,\ \sim)} \\
&= \frac{P(U \mid \sigma_U)\,.\,P(\sigma_U \mid \alpha_U,\, \beta_U)}{\sum_{\sigma_U} P(U \mid \sigma_U)\,.\,P(\sigma_U \mid \alpha_U,\, \beta_U)} \\
&\propto P(U \mid \sigma_U)\,.\,P(\sigma_U \mid \alpha_U,\, \beta_U)
\end{aligned}
$$

- در نتیجه complete conditional متناسب با توزیع بالا به دست می آید.

حال توزیع به دست آمده را بسط میدهیم(روابط توزیع نرمال و گاما را مینویسم)

توزیع نرمال زیر چون دارای سیگمای ثابت در ابعاد Dتایی آن است پس میتوان آن را به صورت ضرب D توزیع نرمال نوشت تا بتوان نتیجه را با توزیع gamma inverse ترکیب کرد.

$$
\begin{aligned}
P(U \mid \sigma_U)\,.\,P(\sigma_U \mid \alpha_U,\, \beta_U) &= \left(\prod_{i=1}^{N} N(0,\, \sigma_U I)\right).\ invgamma(\sigma_U \mid \alpha_U,\, \beta_U) \\
&= \left(\prod_{i=1}^{N} \frac{1}{|\sigma_U I|^{0.5}} \exp\left(\frac{-1}{2} U_i^T (\sigma_U I)^{-1} U_i\right)\right).\ \left(\frac{1}{\sigma_U}\right)^{\alpha_U + 1}.\exp\left(\frac{-\beta}{\sigma_U}\right) \\
&= \left(\prod_{i=1}^{N} \prod_{d=1}^{D} \frac{1}{\sigma_{Ud}^{0.5}} \exp\left(\frac{-1}{2\sigma_{Ud}} U_{id}^2\right)\right).\ \left(\frac{1}{\sigma_U}\right)^{\alpha_U + 1}.\exp\left(\frac{-\beta}{\sigma_U}\right) \\
&= \exp\left(\left(\sum_{i=1}^{N} \sum_{d=1}^{D} \frac{-1}{2\sigma_U} U_{id}^2\right) - \frac{\beta}{\sigma_U}\right).\left(\frac{1}{\sigma_U}\right)^{\frac{N \times D}{2} + \alpha_U + 1}
\end{aligned}
$$

- در نتیجه توزیع complete conditional برای سیگمای U از فرم توزیع gamma inverse پیروی میکند با alpha و betaی جدید که مطابق با زیر به دست می آید.

$$\sigma_U \sim Invgamma(\alpha_U,\, \beta_U)$$

$$\alpha_U = \alpha_U + \frac{N \times D}{2}$$

$$\beta_U = \beta_U + \sum_{i=1}^{N}\sum_{d=1}^{D}\frac{1}{2}U_{id}^2$$

- مشابه با آن چه برای محاسبه سیگمای U به دست آمد روابط زیر برای سیگمای V نیز به دست می آید که از یک توزیع gamma inverse مطابق با زیر پیروی میکند.

$$\sigma_V \sim Invgamma(\alpha_V,\, \beta_V)$$

$$\alpha_V = \alpha_V + \frac{M \times D}{2}$$

$$\beta_V = \beta_V + \sum_{j=1}^{M}\sum_{d=1}^{D}\frac{1}{2}V_{jd}^2$$

بخش دوم: استفاده از داده‌های شبکه‌های اجتماعی

همان طور که در صورت سوال ذکر شده میانگین توزیع امتیازات علاوه بر وابستگی به ماتریس‌های U و V به جمع وزن دار امتیازات دوستان فرد نیز وابسته است. وزن ها تاثیر هر فرد روی فرد دیگر را نشان میدهد و در این مسئله یک متغیر نهان بوده و مشابه با دیگر متغیرهای نهان نیاز به یادگیری دارد.

همان طور که در رابطه‌ی زیر مشاهده میشود امتیاز فرد i به زامین item j علاوه بر ضرب داخلی U_i در V_j به علاوه‌ی مقدار زیر نیز میشود:

برای هر همسایه کاربرi ام:

(ضریب W_ik که مربوط به تاثیر کاربر i و همسایه آن یعنی k است)

ضربدر

(امتیازی که همسایه آن به زامین jitem)

و برای همه همسایه های کاربر i این مقدار جمع میشود.

$$R_{ij} \sim N\left(\left(U_i^T V_j + \sum_{k \in \, neigbor(i)} W_{ik} R_{kj}^{I_{kj}}\right),\, \sigma\right)$$

- مشابه با قبل نیاز به محاسبه توزیع احتمال posterior متغیرهای نهان داریم که در این بخش علاوه بر متغیرهای قبلی W نیز اضافه شده است.

$$P(U,\, V,\, W,\, \sigma_U,\, \sigma_V \,|\, R) \;=\; \frac{P(R\,|\,U,\, V,\, W,\, \sigma).\, P(U\,|\,\sigma_U).\, P(V\,|\,\sigma_V).\, P(\sigma_U\,|\,\alpha_U,\, \beta_U).\, P(\sigma_V\,|\,\alpha_V,\, \beta_V\,)}{\int P(R\,|\,U,\, V,\, W,\, \sigma).\, P(U\,|\,\sigma_U).\, P(V\,|\,\sigma_V).\, P(\sigma_U\,|\,\alpha_U,\, \beta_U).\, P(\sigma_V\,|\,\alpha_V,\, \beta_V\,)}$$

- برای محاسبه عبارت بالا از روش variationl mean field مطابق با روش خواسته شده در صورت سوال استفاده میکنیم.

- در نتیجه از توزیع تقریبی Q استفاده میکنیم که روی متغیرهای نهان مطابق با زیر factorize میشود.

$$Q(U, V, W, \sigma_U, \sigma_V) = Q(U).\, Q(V)\,.\, Q(W)\,.\, Q\,(\sigma_U).\, Q(\sigma_V)$$

- در نتیجه میتوانیم توزیع Q روی هر متغیرنهان را با استفاده از variational optimization به دست آوریم.

- ابتدا به محاسبه توزیع ماتریس U با استفاده از این روش خواهیم پرداخت.

مطابق با varitional optimzation توزیع Q از رابطه زیر به دست می آید.

$$\ln Q(U) = E_{\sim U}[\ln P(U, V, W, \sigma_U, \sigma_V, R)]$$

- حال سمت راست و توزیع p را مطابق با مدل احتمالاتی factorize میکنیم و آن termهایی که به U وابسته نیستند را نیز حذف کرده و به عنوان constant در نظر میگیریم. هم چنین توزیع W را 0 در نظر میگیریم تا در عبارات به constant تبدیل شود.

$$\begin{aligned}\ln Q(U) &= E_{\sim U}[\ln P(U, V, W, \sigma_U, \sigma_V, R)]\\ &= E_{\sim U}\left[\left(\sum_{i=1}^{N}\sum_{j=1}^{M} \ln N\left(R_{ij} \middle| U_i^T V_j + \sum_{k \in neighbor(i)} W_{ik} R_{kj}^{I_{kj}}, \sigma\right)\right) + \sum_{i=1}^{N} \ln N(U_i \,|\, 0, \sigma_U)\right] + const\end{aligned}$$

- همان طور که مشاهده میشود عبارت به دست آمده مطابق با زیر روی U_i ها نیز factorize میشود و بنابراین میتوان در ادامه توزیع U_i را محاسبه کرد.

$$\ln Q(U) = \sum_{i=1}^{N} \ln Q(U_i) \implies Q(U) = \prod_{i=1}^{N} Q(U_i)$$

- برای محاسبه توزیع U_i ها عبارات به دست آمده از مرحله قبل که همه شامل U_i هستند را بسط میدهیم(توزیع نرمال را مینویسیم).

توان دوی به دست آمده در زیر را نیز باز کرده و تنها عباراتی که به U_i وابسته هستند را نگه میداریم.

سپس امیدریاضی را محاسبه میکنیم که در عباراتی که متغیرهای نهان غیر از U_i وجود دارد از آن ها امید ریاضی گرفته میشود.

$$\begin{aligned}
\ln Q(U_i) &= E_{\sim U_i}\left[\left(\sum_{j=1}^{M} \frac{-I_{ij}}{2\sigma}\left(R_{ij} - U_i^T V_j - \sum_{k\in neighbor(i)} W_{ik} R_{kj}^{I_{kj}}\right)^2\right)\right] \\
&\quad + E_{\sim U_i}\left[-\frac{1}{2} U_i^T (\sigma_U I)^{-1} U_i\right] \\
&= E_{\sim U_i}\left[\sum_{j=1}^{M} \frac{-I_{ij}}{2\sigma}\left(\left(U_i^T V_j\right)^2 - 2R_{ij} V_j^T U_i - 2\left(\sum_{k\in neighbor(i)} W_{ik} R_{kj}^{I_{kj}}\right).\left(V_j^T U_i\right)\right)\right] \\
&\quad + E_{\sim U_i}\left[-\frac{1}{2} U_i^T (\sigma_U I)^{-1} U_i\right] \\
&= \sum_{j=1}^{M} \frac{-I_{ij}}{2\sigma}\left(U_i^T E\left[V_j V_j^T\right] U_i - 2R_{ij} E\left[V_j^T\right] U_i - 2 \sum_{k\in neighbor(i)} E[W_{ik}] R_{kj}^{I_{kj}}.E\left[V_j^T\right] U_i\right) \\
&\quad - \frac{1}{2} U_i^T (E[\sigma_U] I)^{-1} U_i \\
&\quad \implies E\left[V_j V_j^T\right] = \Sigma_{V(j)}^{new} + \left(\mu_{V(j)}^{new}\right)^2,\ E\left[V_j^T\right] = \mu_{V(j)}^{new},\ E[\sigma_U] = \frac{\beta_U^{new}}{\alpha_U^{new} - 1},\ E[W_{ik}] = \mu_{W(ik)}^{new} \\
&= \sum_{j=1}^{M} \frac{-I_{ij}}{2\sigma}\left(U_i^T\left(\Sigma_{V(j)}^{new} + \left(\mu_{V(j)}^{new}\right)^2\right) U_i\right) + \sum_{j=1}^{M} \frac{I_{ij}}{\sigma}\left(R_{ij} \mu_{V(j)}^{new} U_i\right) \\
&\quad + \sum_{j=1}^{M} \sum_{k\in neighbor(i)} \frac{I_{ij}}{\sigma} \mu_{W(ik)}^{new} R_{kj}^{I_{kj}} \mu_{V(j)}^{new} U_i - \frac{1}{2} U_i^T\left(\left(\frac{\beta_U^{new}}{\alpha_U^{new} - 1}\right) I\right)^{-1} U_i
\end{aligned}$$

- هم چنین در عبارات بالا امیدریاضی ها به دست آمده از متغیرهای نهان بر اساس توزیع های آن ها(قسمت نتیجه) جایگذاری شده اند.
- چون فرم به دست آمده quadratic است پس از یک توزیع نرمال پیروی میکند و بنابراین توزیع نرمال variational برای U_i مطابق با زیر به دست می آید.

$$U_i \sim N\left(\mu_{U(i)}^{new},\ \Sigma_{U(i)}^{new}\right)$$

$$\left(\Sigma_{U(i)}^{new}\right)^{-1} = \left(\left(\left(\frac{\beta_U^{new}}{\alpha_U^{new}-1}\right)I\right)^{-1} + \frac{1}{\sigma}\sum_{j=1}^{M} I_{ij}\left(\Sigma_{V(j)}^{new} + \left(\mu_{V(j)}^{new}\right)^2\right)\right)$$

$$\mu_{U(i)}^{new} = \Sigma_{U(i)}^{new}\left(\sum_{j=1}^{M}\frac{I_{ij}}{\sigma}\left(R_{ij}\mu_{V(j)}^{new} + \sum_{k\in neighbor(i)} \mu_{W(ik)}^{new} R_{kj}^{I_{kj}} \mu_{V(j)}^{new}\right)\right)$$

- مشابه با آن چه برای محاسبه U_i به دست آمد روابط زیر برای V_i نیز به دست می آید که از یک توزیع نرمال مطابق با زیر پیروی میکند.

$$V_j \sim N\left(\mu_{V(j)}^{new},\ \Sigma_{V(j)}^{new}\right)$$

$$\left(\Sigma_{V(j)}^{new}\right)^{-1} = \Bigg(\bigg(\Big(\frac{\beta_V^{new}}{\alpha_V^{new}-1}\Big)I\bigg)^{-1} + \frac{1}{\sigma}\sum_{i=1}^{N} I_{ij}\Big(\Sigma_{U(i)}^{new} + \left(\mu_{U(i)}^{new}\right)^2$$

$$\mu_{V(j)}^{new} = \Sigma_{V(j)}^{new}\left(\sum_{i=1}^{N}\frac{I_{ij}}{\sigma}\left(R_{ij}\mu_{U(i)}^{new} + \sum_{k\in neighbor(i)} \mu_{W(ik)}^{new} R_{kj}^{I_{kj}} \mu_{U(i)}^{new}\right)\right)$$

- حال محاسبه توزیع سیگمای U با استفاده ازروش variational مطابق با قبل خواهیم پرداخت.

مطابق با varitional optimzation توزیع Q از رابطه زیر به دست می آید.

برای محاسبه توزیع سیگمای U عباراتی که همه شامل سیگمای U هستند را بسط میدهیم(توزیع نرمال و گاما را مینویسیم).

توزیع نرمال زیر را به صورت ضرب D توزیع نرمال نوشت تا بتوان نتیجه را با توزیع gamma inverse ترکیب کرد.

سپس امیدریاضی را محاسبه میکنیم که در عباراتی که متغیرهای نهان غیر از سیگمای U وجود دارد از آن ها امید ریاضی گرفته میشود.

$$\begin{aligned}
\ln Q(\sigma_u) \;&=\; E_{\sim\sigma_u}[\ln P(U,\, V,\, W,\, \sigma_U,\, \sigma_V,\, R)] \\
&= E_{\sim\sigma_u}\left[\ln\left(\left(\prod_{i=1}^{N} N(U_i \,|\, 0,\, \sigma_u I)\right) .\; invgamma(\sigma_u \,|\, \alpha_U,\, \beta_u)\right)\right] \\
&= E_{\sim\sigma_u}\left[\ln\left(\left(\prod_{i=1}^{N} \frac{1}{|\sigma_u I|^{0.5}} \exp\left(\frac{-1}{2} U_i^T (\sigma_u I)^{-1} U_i\right)\right) . \left(\frac{1}{\sigma_u}\right)^{\alpha_U+1} . \exp\left(\frac{-\beta}{\sigma_u}\right)\right)\right] \\
&= E_{\sim\sigma_u}\left[\ln\left(\left(\prod_{i=1}^{N}\prod_{d=1}^{D} \frac{1}{\sigma_{ud}^{0.5}} \exp\left(\frac{-1}{2\sigma_{ud}} U_{id}^2\right)\right) . \left(\frac{1}{\sigma_u}\right)^{\alpha_U+1} . \exp\left(\frac{-\beta}{\sigma_u}\right)\right)\right] \\
&= \ln\left(\left(\prod_{i=1}^{N}\prod_{d=1}^{D} \frac{1}{\sigma_{ud}^{0.5}} \exp\left(\frac{-1}{2\sigma_{ud}} E[U_{id}^2]\right)\right) . \left(\frac{1}{\sigma_u}\right)^{\alpha_U+1} . \exp\left(\frac{-\beta}{\sigma_u}\right)\right)
\end{aligned}$$

$$\Longrightarrow\;\; E[U_{id}^2] \;=\; \left(\mu_{U(i,\,d)}^{new}\right)^2 \;+\; \Sigma_{U(idd)}^{new}$$

$$\begin{aligned}
&= \ln\left(\left(\prod_{i=1}^{N}\prod_{d=1}^{D} \frac{1}{\sigma_{ud}^{0.5}} \exp\left(\sum_{i=1}^{N}\sum_{d=1}^{D} \frac{-1}{2\sigma_{ud}}\left(\left(\mu_{U(i,\,d)}^{new}\right)^2 \;+\; \Sigma_{U(idd)}^{new}\right)\right)\right) . \left(\frac{1}{\sigma_u}\right)^{\alpha_U+1} . \exp\left(\frac{-\beta}{\sigma_u}\right)\right) \\
&= \ln\left(\left(\frac{1}{\sigma_u}\right)^{\frac{N\times D}{2}\;+\;\alpha_U+1} . \exp\left(\sum_{i=1}^{N}\sum_{d=1}^{D} \frac{-1}{2\sigma_{ud}}\left(\left(\mu_{U(i,\,d)}^{new}\right)^2 \;+\; \Sigma_{U(idd)}^{new}\right) \;+\; \frac{-\beta}{\sigma_u}\right)\right)
\end{aligned}$$

$$\Longrightarrow$$

$$Q(\sigma_u) = \left(\frac{1}{\sigma_u}\right)^{\frac{N\times D}{2}\;+\;\alpha_U+1} . \exp\left(\sum_{i=1}^{N}\sum_{d=1}^{D} \frac{-1}{2\sigma_u}\left(\left(\mu_{U(i,\,d)}^{new}\right)^2 \;+\; \Sigma_{U(idd)}^{new}\right) \;+\; \frac{-\beta}{\sigma_u}\right)$$

- هم چنین در عبارات بالا امیدریاضی ها به دست آمده از متغیرهای نهان بر اساس توزیع های آن ها(قسمت نتیجه) جایگذاری شده اند.

- در نتیجه توزیع variational برای سیگمای U از فرم توزیع gamma inverse پیروی میکند با alpha و betaی جدید که مطابق با زیر به دست می آید.

پس

$$\sigma_U \sim Invgamma(\alpha_U^{new},\ \beta_U^{new})$$

$$\alpha_U^{new} = \alpha_U + \frac{N \times D}{2}$$

$$\beta_U^{new} = \beta_U + \sum_{i=1}^{N}\sum_{d=1}^{D}\frac{1}{2}\left(\left(\mu_{U(id)}^{new}\right)^2 + \Sigma_{U(idd)}^{new}\right)$$

- مشابه با آن چه برای محاسبه سیگمای U به دست آمد روابط زیر برای سیگمای V نیز به دست می آید که از یک توزیع gamma inverse مطابق با زیر پیروی میکند.
-

$$\sigma_V \sim Invgamma(\alpha_V^{new},\ \beta_V^{new})$$

$$\alpha_V^{new} = \alpha_V + \frac{N \times D}{2}$$

$$\beta_V^{new} = \beta_V + \sum_{j=1}^{M}\sum_{d=1}^{D}\frac{1}{2}\left(\left(\mu_{V(jd)}^{new}\right)^2 + \Sigma_{V(jdd)}^{new}\right)$$

- حال محاسبه توزیع W با استفاده ازروش variational مطابق با قبل خواهیم پرداخت.

سمت راست و توزیع p را مطابق با مدل احتمالاتی factorize میکنیم و آن termهایی که به W وابسته نیستند را نیز حذف کرده و به عنوان constant در نظر میگیریم.

$$\ln Q(W) = E_{\sim W}[\ln P(U, V, W, \sigma_U, \sigma_V, R)]$$

$$\begin{aligned}\ln Q(W) &= E_{\sim W}[\ln P(U, V, W, \sigma_U, \sigma_V, R)] \\ &= E_{\sim W}\left[\sum_{i=1}^{N}\sum_{j=1}^{M} \ln N\left(R_{ij} \middle| U_i^T V_j + \sum_{k \in neighbor(i)} W_{ik} R_{kj}^{I_{kj}}, \sigma\right)\right] + const\end{aligned}$$

- همان طور که مشاهده میشود عبارت به دست آمده مطابق با زیر روی W_ik ها نیز factorize میشود و بنابراین میتوان در ادامه توزیع W_ik را محاسبه کرد.

$$\ln Q(W) = \sum_{i=1}^{N} \sum_{k \in neighbor(i)} \ln Q(W_{ik}) \implies Q(W) = \prod_{i=1}^{N} \prod_{k \in neighbor(i)} Q(W_{ik})$$

- برای محاسبه توزیع W_ik ها عبارات به دست آمده از مرحله قبل که همه شامل W_ik هستند را بسط میدهیم(توزیع نرمال را مینویسیم).

  توان دوی به دست آمده در زیر را نیز باز کرده و تنها عباراتی که به W_ik وابسته هستند را نگه میداریم.

  سپس امیدریاضی را محاسبه میکنیم که در عباراتی که متغیرهای نهان غیر از W_ik وجود دارد از آن ها امید ریاضی گرفته میشود.

$$\begin{aligned}
\ln Q(W_{ik}) \;&=\; E_{\sim W_{ik}}\left[\left(\sum_{j=1}^{M}\frac{-I_{ij}}{2\sigma}\left(R_{ij} \;-\; U_i^T V_j \;-\; W_{ik}R_{kj}^{I_{kj}}\right)^2\right)\right] \\
&= E_{\sim W_{ik}}\left[\sum_{j=1}^{M}\frac{-I_{ij}}{2\sigma}\left(\left(W_{ik}R_{kj}^{I_{kj}}\right)^2 + 2U_i^T V_j W_{ik}R_{kj}^{I_{kj}} - 2R_{ij}W_{ik}R_{kj}^{I_{kj}}\right)\right] \\
&= \sum_{j=1}^{M}\frac{-I_{ij}}{2\sigma}\left(\left(R_{kj}^{I_{kj}}\right)^2 . W_{ik}^2 + 2E\left[U_i^T\right]E[V_j]W_{ik}R_{kj}^{I_{kj}} \;-\; 2R_{ij}R_{kj}^{I_{kj}}W_{ik}\right)
\end{aligned}$$

$$\Longrightarrow \quad E\left[U_i^T\right] \;=\; \mu_{U(i)}^{new},\; E\left[V_j^T\right] \;=\; \mu_{V(j)}^{new}$$

$$= \sum_{j=1}^{M}\frac{-I_{ij}}{2\sigma}\left(\left(R_{kj}^{I_{kj}}\right)^2 . W_{ik}^2 \;+\; 2\left(\mu_{U(i)}^{new}\right)^T \mu_{V(j)}^{new}R_{kj}^{I_{kj}}W_{ik} \;-\; 2R_{ij}R_{kj}^{I_{kj}}W_{ik}\right)$$

- هم چنین در عبارات بالا امیدریاضی ها به دست آمده از متغیرهای نهان بر اساس توزیع های آن ها(قسمت نتیجه) جایگذاری شده اند.

- چون فرم به دست آمده quadratic است پس از یک توزیع نرمال پیروی میکند و بنابراین توزیع نرمال variational برای W_ik مطابق با زیر به دست می آید.

$$W_{ik} \sim N\left(\mu_{W(ik)}^{new}, \sigma_{W(ik)}^{new}\right)$$

$$\left(\sigma_{W(ik)}^{new}\right)^{-1} = \left(\frac{1}{\sigma} \sum_{j=1}^{M} I_{ij} \left(R_{kj}^{I_{kj}}\right)^{2}\right)$$

$$\mu_{W(ik)}^{new} = \sigma_{ik} \left(\sum_{j=1}^{M} I_{ij} \left(R_{ij} R_{kj}^{I_{kj}} - \left(\mu_{U(i)}^{new}\right)^{T} \mu_{V(j)}^{new} R_{kj}^{I_{kj}} \right) \right)$$

در نتیجه توزیع هر کدام از متغیرهای نهان به دست آمد.

همان طور که مشاهده میشود توزیع ها به هم وابستگی دارند (به عنوان مثال توزیع U به میانگین توزیع V وابسته است.)بنابراین با شروع از U به ترتیب V ، سیگمای U، سیگمای V و در نهایت W تخمین زده میشود و تا convergence این کار ادامه میابد.

- در ادامه پیاده سازی انجام شده و نتایج به دست آمده را بررسی خواهیم کرد.

بخش سوم: پیاده سازی

پیاده سازی قسمت اول در فایل gibbs_sampling.py و قسمت دوم در فایل variational_mean_field.py انجام شده

Gibbs sampling:

در این بخش در هر مرحله مطابق با روابط به دست آمده ماتریس‌های U و V و سپس سیگمای U و سیگمای V بر اساس توزیع های هر کدام نمونه برداری میشوند.

- مقادیر اولیه ماتریس‌های U و V و سیگمای U و سیگمای V نیز بر اساس توزیع های اولیه آن ها و با مقادیر اولیه فراپارامترهای sigma, alpha_U, alpha_V, beta_U, beta_v نمونه برداری شده اند.
- برای هر user و item چون مستقل از هم در نظر گرفته شده اند بنابراین نمونه برداری ماتریس‌های U و V نیز به صورت مستقل بر روی iامین user و jامین item انجام شده است.

در انتها بر اساس ماتریس های U و V نمونه برداری شده از ضربر داخلی این دو بردار R_hat به دست می آید. سپس نتیجه با استفاده از MinMaxScaler به بازه بین صفر تا پنج scale میشود.

برای محاسبه RMSE مراحل آموزش در هر مرحله تفاوت بین R_hat و جاهایی که در ماتریس R امتیاز وجود دارد(غیر صفرها) محاسبه میشود.

در نهایت با استفاده از ماتریس های U و V نهایی مقدار خطا مشابه با قبل برای داده های validation محاسبه شده است.

:Variational Mean Field

در این بخش ابتدا همسایگی‌های هر فرد در یک لیست مجاورت نگه داری شده است.

در ادامه مطابق با روش gibbs sampling در هر مرحله طابق با روابط به دست آمده ماتریس‌های U و V و سپس سیگمای U و سیگمای V بر اساس توزیع های هر کدام نمونه برداری میشوند.

تفاوت روابط در این مرحله شامل نتایج به دست آمده از روش variational mean field و هم چنین اضافه شدن ماتریس W است که بیانگر تاثیرگذاری افراد و همسایه‌های آن هاست.

- مقادیر اولیه ماتریس‌های U و V و سیگمای U و سیگمای V نیز بر اساس توزیع های اولیه آن ها و با مقادیر اولیه فراپارامترهای sigma, alpha_U, alpha_V, beta_U, beta_v نمونه برداری شده اند.

- مقدار اولیه ماتریس W نیز در ابتدا صفر فرض شده است. به این معنا که در iteration اول کاربران بر روی هم تاثیری ندارند.

- برای هر user و item چون مستقل از هم در نظر گرفته شده اند بنابراین نمونه برداری ماتریس‌های U و V نیز به صورت مستقل بر روی iامین user و jامین item انجام شده است.

- توزیع ها به هم وابستگی دارند (به عنوان مثال توزیع U به میانگین توزیع V وابسته است.)بنابراین با شروع از U به ترتیب V ، سیگمای U، سیگمای V و در نهایت W تخمین زده میشود و تا convergence این کار ادامه میابد. بنابراین نیاز است پارامترهای توزیع ها در هر مرحله شامل mean , cov ها و غیره نگه داری شوند.

محاسبه RMSE مرحله آموزش و هم چنین مقدار خطا بر روی داده vaildation مطابق با بخش اول انجام شده است.

بخش چهارم: نتایج

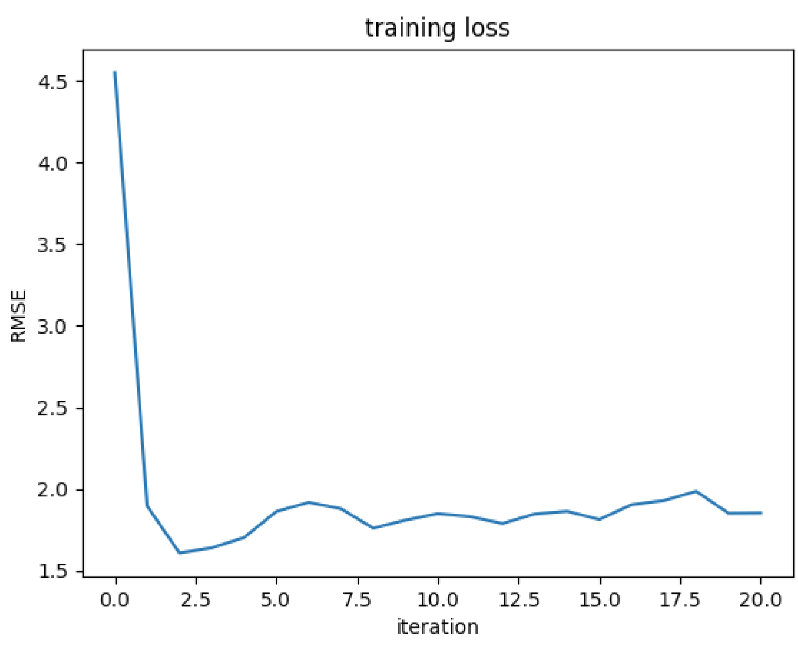
فراپارامترها:

- تعداد ابعاد latent دو ماتریس U و V مقدار **2** انتخاب شد.
- بر اساس چند تست انجام شده و مقایسه میزان خطا مقادیر زیر برای دیگر پارامترها انتخاب شد.

- sigma=1, alpha_U=1, beta_U=1, alpha_V=1, beta_V=1

نتایج بخش اول - رویکرد bayesian :

نمودار خطای RMSE در iterationهای مرحله آموزش مشاهده میشود.



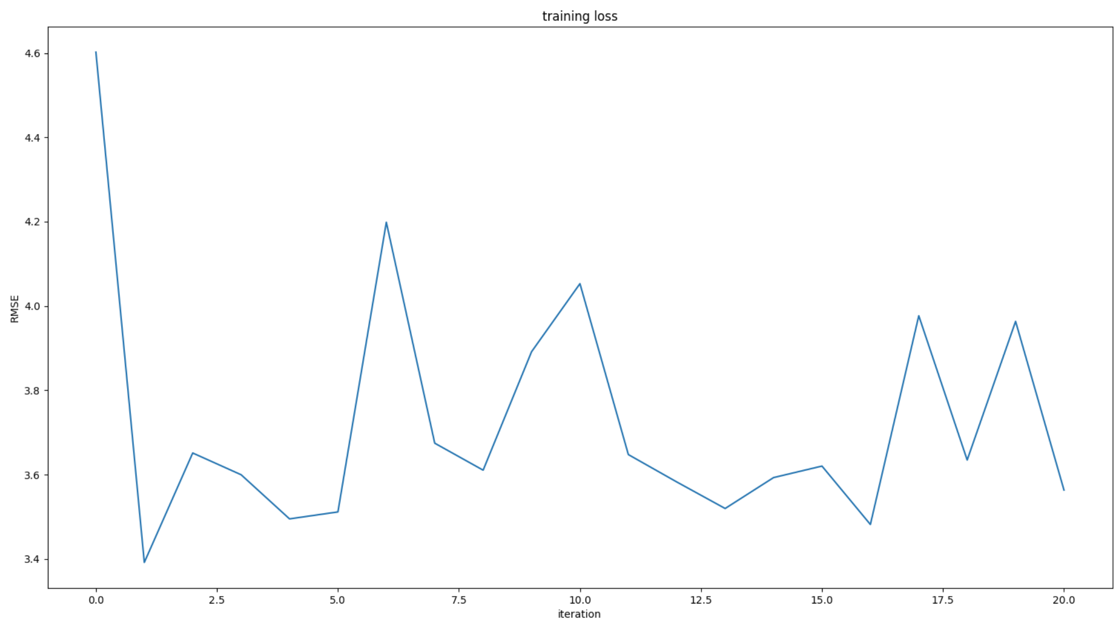
هم چنین مقدار خطای rmse بر روی دادههای validation مطابق با زیر به دست آمده است.

RMSE_VALIDATION = 4.805378367441022

نتایج بخش دوم - استفاده از داده‌های شبکه‌های اجتماعی :

مطابق با روابط به دست آمده توزیع های ماتریس های U و V و سیگمای U و سیگمای V به هم وابستگی داشتند. وابستگی نیز در محاسبه Epectation بر اساس متغیرهای نهان به وجود می آمد که یک راه همان طور که در روابط به دست آمد جایگذاری آن با expectationهای توزیع های جدید به دست آمده بود و یک راه جایگذاری با توزیع های اصلی U و V که راه دوم نتایج بهتری ایجاد میکرد.

نمودار خطای RMSE در iterationهای مرحله آموزش مشاهده میشود.



هم چنین مقدار خطای rmse بر روی داده‌های validation مطابق با زیر به دست آمده است.

RMSE_VALIDATION = 4.255453720455846

- RMSE آموزش بخش اول از این بخش بهتر است که میتواند به خاطر برتری روش gibss sampling نسبت به variational mean field در این مسئله باشد.

- نوع مدل سازی و اضافه شدن تاثیر دوستان در امتیازات هر فرد به نظر میرسد که میتواند در بهتر شدن نتیجه تاثیر گذار باشد. ولی در دادهی موجود بر این مسئله تعداد حالاتی که یک فرد به itemای امتیاز داده باشد که دوست او نیز به آن item داده باشد در مقایسه با تعداد userها و itemها بسیار کم است.