Purpose: 期望最大化算法

$$
\begin{aligned}
\max \quad \ln p(\boldsymbol{X}|\boldsymbol{\theta}) &= \left\{\sum_{\boldsymbol{Z}} q(\boldsymbol{Z})\right\} \ln p(\boldsymbol{X}|\boldsymbol{\theta}) \\
&= \left\{\sum_{\boldsymbol{Z}} q(\boldsymbol{Z})\right\} \ln \frac{p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})}{p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta})} \\
&= \mathcal{L}(q,\boldsymbol{\theta}) + \mathrm{KL}(q\|p)
\end{aligned}
$$

where X is **constant**, parameter $q(Z), \theta$, **random variable** Z;

X, Z are **discrete matrix**

$$
\begin{aligned}
\mathcal{L}(q,\boldsymbol{\theta}) &= \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln\left\{\frac{p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})}{q(\boldsymbol{Z})}\right\} \\
\mathrm{KL}(q\|p) &= -\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln\left\{\frac{p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta})}{q(\boldsymbol{Z})}\right\} \\
&= -\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln\{q(\boldsymbol{Z})\} + \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln\{p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta})\} \geq 0
\end{aligned}
$$

Where

- $q(\boldsymbol{Z})$ is $\forall$ arbitrary distribution for $\boldsymbol{Z}$
- $\mathcal{L}(q,\boldsymbol{\theta})$ 概率分布$q(\boldsymbol{Z})$的一个泛函, $\boldsymbol{\theta}$ 函数
- $\mathrm{KL}(q\|p)$ 概率分布$q(\boldsymbol{Z})$的一个泛函, $\boldsymbol{\theta}$ 函数，KL散度 of $p(Z|X,\theta)$, $q(Z)$, ≥ 0

EM法

$$
\ln p(\boldsymbol{X}|\boldsymbol{\theta}) = \mathcal{L}(q,\boldsymbol{\theta}) + \mathrm{KL}(q|p)
$$

[M步骤]

fix $q \equiv q^k = p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^k\right) = p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)$

$$
\begin{aligned}
\theta^{k+1} &\equiv \underset{\theta}{\operatorname{argmax}} \mathcal{L}(q^k,\boldsymbol{\theta}) \\
&= \underset{\theta}{\operatorname{argmax}} \sum_{\boldsymbol{Z}} q^k \ln\left\{\frac{p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})}{q^k}\right\} \\
&= \underset{\theta}{\operatorname{argmax}} \sum_{\boldsymbol{Z}} q^k \ln p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta}) - \sum_{\boldsymbol{Z}} q^k \ln(q^k) \\
&= \underset{\theta}{\operatorname{argmax}} \sum_{\boldsymbol{Z}} q^k \ln p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})
\end{aligned}
\qquad \mathcal{L}(q^k,\boldsymbol{\theta}^{k+1}) > \mathcal{L}(q^k,\boldsymbol{\theta}^k) = \ln p(\boldsymbol{X}|\boldsymbol{\theta}^k)
$$

where $\boldsymbol{\theta}^{\text{ old}} \equiv \boldsymbol{\theta}^k$

$$\boldsymbol{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}\right) \equiv \sum_{\boldsymbol{Z}} q^k \ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta})$$

[E步骤]

fix $\boldsymbol{\theta} \equiv \boldsymbol{\theta}^{k+1}$

$$\begin{aligned}
q^{k+1} &\equiv \underset{q}{\operatorname{argmax}} \mathcal{L}(q, \boldsymbol{\theta}^{k+1}) \\
&\equiv \underset{q}{\operatorname{argmin}} \mathrm{KL}(q|p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1})) = 0 \\
&= p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1})
\end{aligned}$$

$$\ln p(\boldsymbol{X}|\boldsymbol{\theta}^{k+1}) = \mathcal{L}(q^{k+1}, \boldsymbol{\theta}^{k+1}) > \mathcal{L}(q^k,$$

$$\begin{aligned}
\ln p(\boldsymbol{X}|\boldsymbol{\theta}^{k+1}) &\equiv \mathcal{L}(q, \boldsymbol{\theta}^{k+1}) + \mathrm{KL}(q|p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1})) \\
&= \mathcal{L}(q^k, \boldsymbol{\theta}^{k+1}) + \mathrm{KL}(q^k|p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1})) \\
&= \mathcal{L}(q^{k+1}, \boldsymbol{\theta}^{k+1}) + \mathrm{KL}(q^{k+1}|p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{k+1})) \\
&= \mathcal{L}(q^{k+1}, \boldsymbol{\theta}^{k+1})
\end{aligned}$$

in all

$$\ln p(\boldsymbol{X}|\boldsymbol{\theta}^{k+1}) = \mathcal{L}(q^{k+1}, \boldsymbol{\theta}^{k+1}) > \mathcal{L}(q^k, \boldsymbol{\theta}^{k+1}) > \ln p(\boldsymbol{X}|\boldsymbol{\theta}^k) = \mathcal{L}(q^k, \boldsymbol{\theta}^k)$$

So

$$\lim_{k\to\infty} \ln p(\boldsymbol{X}|\boldsymbol{\theta}^k) = \lim_{k\to\infty} \mathcal{L}(q^k, \boldsymbol{\theta}^{k+1}) = \max_{\theta} \ln p(\boldsymbol{X}|\boldsymbol{\theta})$$

$$\begin{aligned}
\boldsymbol{\theta}_{max} &\equiv \max_{\theta} \ln p(\boldsymbol{X}|\boldsymbol{\theta}) \\
&= \lim_{k\to\infty} \theta^{k+1} \\
&= \lim_{k\to\infty} \underset{\theta}{\operatorname{argmax}} \sum_{\boldsymbol{Z}} q^k \ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta}) \\
&= \lim_{k\to\infty} \underset{\theta}{\operatorname{argmax}} \boldsymbol{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^k\right)
\end{aligned}$$

formula

To sum up

$$\begin{aligned}
\boldsymbol{\alpha_1} &= [\boldsymbol{\phi}^T \cdot \boldsymbol{x_1}] \odot \boldsymbol{\pi} \\
\boldsymbol{\alpha_n} &= [\boldsymbol{\phi}^T \cdot \boldsymbol{x_n}] \odot [\boldsymbol{A} \cdot \boldsymbol{\alpha_{n-1}}] \\
\boldsymbol{\beta_N} &= 1. \\
\boldsymbol{\beta_n} &= \boldsymbol{A}^T \cdot \left([\boldsymbol{\phi}^T \cdot \boldsymbol{x_{n+1}}] \odot \boldsymbol{\beta_{n+1}}\right)
\end{aligned}$$

So if $\boldsymbol{\alpha_n}, \boldsymbol{\beta_n}$ is obtained, we could get $\boldsymbol{\gamma_n}, \boldsymbol{\xi_n}$

$$\begin{aligned}
\boldsymbol{\alpha_n} &\equiv \vec{p}(\boldsymbol{z_n}, \boldsymbol{x_1}, \ldots, \boldsymbol{x_n}|\boldsymbol{\theta}^{\text{old}}) \\
\boldsymbol{\beta_n} &\equiv \vec{p}(\boldsymbol{x_{n+1}}, \ldots, \boldsymbol{x_N}|\boldsymbol{z_n}, \boldsymbol{\theta}^{\text{old}})
\end{aligned}$$

Here

$\boldsymbol{\gamma_n} \equiv \left[\sum_{\boldsymbol{z_n}} p\left(\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)\boldsymbol{z_n}\right]$ 为常数，是期望$E[\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}] = \vec{p}(\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}})$, 大小$K \times 1$

$\boldsymbol{\xi_n} \equiv [\sum_{\boldsymbol{z_{n-1}},\boldsymbol{z_n}} \boldsymbol{z_n} p\left(\boldsymbol{z_{n-1}},\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)\boldsymbol{z}_{\boldsymbol{n-1}}^T]$ 为常数，是期望$E[\boldsymbol{z_n}\boldsymbol{z_{n-1}^T}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}] = \vec{p}(\boldsymbol{z_n}\boldsymbol{z_{n-1}^T}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}})$, 大小$K \times K$

$$
\begin{aligned}
\boldsymbol{\gamma_n} &= \boldsymbol{\alpha_n} \odot \boldsymbol{\beta_n} \odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}})} \\
&= \frac{1}{1.^T \cdot \boldsymbol{\alpha_N}}[\boldsymbol{\alpha_n} \odot \boldsymbol{\beta_n}] \\
\boldsymbol{\xi_n} &= \frac{1}{p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}})}\left\{\left[[\boldsymbol{\phi}^T \cdot \boldsymbol{x_n}] \odot \boldsymbol{\beta_n}\right] \cdot \boldsymbol{\alpha}_{\boldsymbol{n-1}}^T\right\} \odot \boldsymbol{A} \\
&= \frac{1}{1.^T \cdot \boldsymbol{\alpha_N}}\left\{\left[[\boldsymbol{\phi}^T \cdot \boldsymbol{x_n}] \odot \boldsymbol{\beta_n}\right] \cdot \boldsymbol{\alpha}_{\boldsymbol{n-1}}^T\right\} \odot \boldsymbol{A}
\end{aligned}
$$

其中$p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}}) = \sum_{Z_N} p(\boldsymbol{z_N},\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}}) = \sum \boldsymbol{\alpha_N} = 1.^T \cdot \boldsymbol{\alpha_N}$ is **constant**

$$
\begin{aligned}
\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}}) &= p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}}) \cdot 1. \\
p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}}) &= 1.^T \cdot p(\boldsymbol{z_N},\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}}) \\
&= 1.^T \cdot \boldsymbol{\alpha_N}
\end{aligned}
$$

after fixation

we divide $p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}})$ into 2 part

$$
\begin{aligned}
p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}}) &= p(\boldsymbol{x_1},\cdots,\boldsymbol{x_n}|\boldsymbol{\theta}^{\text{ old}})p(\boldsymbol{x_{n+1}},\cdots,\boldsymbol{x_N}|\boldsymbol{x_1},\cdots,\boldsymbol{x_n},\boldsymbol{\theta}^{\text{ old}}) \\
&= \left\{\prod_{k=1}^{n} p(\boldsymbol{x_k}|\boldsymbol{x_1},\cdots,\boldsymbol{x_{k-1}},\boldsymbol{\theta}^{\text{ old}})\right\}\left\{\prod_{k=n+1}^{N} p(\boldsymbol{x_k}|\boldsymbol{x_1},\cdots,\boldsymbol{x_{k-1}},\boldsymbol{\theta}^{\text{ old}})\right\} \\
&= \left\{\prod_{k=1}^{n} c_k\right\}\left\{\prod_{k=n+1}^{N} c_k\right\}
\end{aligned}
$$

define

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}_{\boldsymbol{n}} &\equiv \frac{1}{p(\boldsymbol{x_1}, \cdots, \boldsymbol{x_n} | \boldsymbol{\theta}^{\text{ old }})} \boldsymbol{\alpha_n} = \vec{p}(\boldsymbol{z_n} | \boldsymbol{x_1}, \cdots, \boldsymbol{x_n}, \boldsymbol{\theta}^{\text{ old }}) \\
&= \left\{ \frac{1}{\prod_{k=1}^{n} p(\boldsymbol{x_k} | \boldsymbol{x_1}, \cdots, \boldsymbol{x_{k-1}}, \boldsymbol{\theta}^{\text{ old }})} \right\} \boldsymbol{\alpha_n} \\
&= \left\{ \frac{1}{\prod_{k=1}^{n} c_k} \right\} \boldsymbol{\alpha_n} \quad \text{notice} \quad 1.^T \hat{\boldsymbol{\alpha}}_{\boldsymbol{n}} = \sum_{\boldsymbol{z_n}} p(\boldsymbol{z_n} | \boldsymbol{x_1}, \cdots, \boldsymbol{x_n}, \boldsymbol{\theta}^{\text{ old }}) = 1 \\
\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} &\equiv \frac{1}{p(\boldsymbol{x_{n+1}}, \cdots, \boldsymbol{x_N} | \boldsymbol{x_1}, \cdots, \boldsymbol{x_n}, \boldsymbol{\theta}^{\text{ old }})} \boldsymbol{\beta_n} \\
&= \left\{ \frac{1}{\prod_{k=n+1}^{N} p(\boldsymbol{x_k} | \boldsymbol{x_1}, \cdots, \boldsymbol{x_{k-1}}, \boldsymbol{\theta}^{\text{ old }})} \right\} \boldsymbol{\beta_n} \\
&= \left\{ \frac{1}{\prod_{k=n+1}^{N} c_k} \right\} \boldsymbol{\beta_n}
\end{aligned}
$$

where we get c_n by

$$
\begin{aligned}
c_n \equiv 1.^T c_n \cdot \boldsymbol{\alpha_n} &= 1.^T \cdot \left\{ [\boldsymbol{\phi}^T \cdot \boldsymbol{x_n}] \odot [\boldsymbol{A} \cdot \hat{\boldsymbol{\alpha}}_{\boldsymbol{n-1}}] \right\} \\
c_{n+1} \neq 1.^T c_{n+1} \cdot \boldsymbol{\beta_n} &= 1.^T \cdot \boldsymbol{A}^T \cdot \left([\boldsymbol{\phi}^T \cdot \boldsymbol{x_{n+1}}] \odot \hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}} \right)
\end{aligned}
$$

also

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}_{\boldsymbol{1}} &= \frac{1}{c_1} [\boldsymbol{\phi}^T \cdot \boldsymbol{x_1}] \odot \boldsymbol{\pi} \\
\hat{\boldsymbol{\alpha}}_{\boldsymbol{n}} &= \frac{1}{c_n} [\boldsymbol{\phi}^T \cdot \boldsymbol{x_n}] \odot [\boldsymbol{A} \cdot \hat{\boldsymbol{\alpha}}_{\boldsymbol{n-1}}] \\
\hat{\boldsymbol{\beta}}_{\boldsymbol{N}} &= 1. \\
\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} &= \frac{1}{c_{n+1}} \boldsymbol{A}^T \cdot \left([\boldsymbol{\phi}^T \cdot \boldsymbol{x_{n+1}}] \odot \hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}} \right)
\end{aligned}
$$

thus

$$
\begin{aligned}
\boldsymbol{\gamma_n} &= \boldsymbol{\alpha_n} \odot \boldsymbol{\beta_n} \odot \frac{1.}{\vec{p}(\boldsymbol{X} | \boldsymbol{\theta}^{\text{ old }})} \\
&= [\hat{\boldsymbol{\alpha}}_{\boldsymbol{n}} \odot \hat{\boldsymbol{\beta}}_{\boldsymbol{n}}] \\
\boldsymbol{\xi_n} &= \frac{1}{p(\boldsymbol{X} | \boldsymbol{\theta}^{\text{ old }})} \left\{ \Big[[\boldsymbol{\phi}^T \cdot \boldsymbol{x_n}] \odot \boldsymbol{\beta_n} \Big] \cdot \boldsymbol{\alpha}_{\boldsymbol{n-1}}^T \right\} \odot \boldsymbol{A} \\
&= \frac{1}{c_n} \left\{ \Big[[\boldsymbol{\phi}^T \cdot \boldsymbol{x_n}] \odot \hat{\boldsymbol{\beta}}_{\boldsymbol{n}} \Big] \cdot \hat{\boldsymbol{\alpha}}_{\boldsymbol{n-1}}^T \right\} \odot \boldsymbol{A}
\end{aligned}
$$

close form of Q

$$
\begin{aligned}
\max_{\boldsymbol{\theta}} \quad \boldsymbol{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{ old }}\right) &\equiv \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z} | \boldsymbol{X}, \boldsymbol{\theta}^{\text{ old }}\right) \ln p(\boldsymbol{X}, \boldsymbol{Z} | \boldsymbol{\theta}) \\
&= \sum_{n=1}^{N} \boldsymbol{\gamma}_{\boldsymbol{n}}^T < \ln \boldsymbol{\phi}, \boldsymbol{x_n} > + \sum_{n=2}^{N} \operatorname{tr}\left(\boldsymbol{\xi}_{\boldsymbol{n}}^T \ln \boldsymbol{A}\right) + \boldsymbol{\gamma}_{\boldsymbol{1}}^T \ln \boldsymbol{\pi}
\end{aligned}
$$

subject to

$$\text{s.t.} \qquad 1.^T \cdot \boldsymbol{\phi} = 1.^T \quad 1.^T \cdot \boldsymbol{A} = 1.^T \quad 1.^T \cdot \boldsymbol{\pi} = 1$$

update $\boldsymbol{\theta}^{k+1} \equiv \{\boldsymbol{\phi}, \boldsymbol{A}, \boldsymbol{\pi}\}$

$$\begin{aligned}
\boldsymbol{\phi} &\Leftarrow \left\{\sum_{n=1}^{N}[\boldsymbol{x_n} \cdot \boldsymbol{\gamma_n^T}]\right\} \odot \frac{1.}{1. \cdot 1.^T \left\{\sum_{n=1}^{N}[\boldsymbol{x_n} \cdot \boldsymbol{\gamma_n^T}]\right\}} \\
&= \left\{\sum_{n=1}^{N}[\boldsymbol{x_n} \cdot \boldsymbol{\gamma_n^T}]\right\} \odot \frac{1.}{1.\left[\sum_{n=1}^{N} \boldsymbol{\gamma_n}\right]^T} \\
\boldsymbol{A} &\Leftarrow [\sum_{n=2}^{N} \boldsymbol{\xi_n}] \odot \frac{1.}{1. \cdot 1.^T[\sum_{n=2}^{N} \boldsymbol{\xi_n}]} \\
&= [\sum_{n=2}^{N} \boldsymbol{\xi_n}] \odot \frac{1.}{1. \cdot [\sum_{n=2}^{N} \boldsymbol{\gamma_{n-1}}]^T} \\
&= [\sum_{n=2}^{N} \boldsymbol{\xi_n}] \odot \frac{1.}{1. \cdot [\sum_{n=1}^{N-1} \boldsymbol{\gamma_n}]^T} \\
\boldsymbol{\pi} &\Leftarrow \boldsymbol{\gamma_1} \odot \frac{1.}{1. \cdot 1.^T \boldsymbol{\gamma_1}} = \boldsymbol{\gamma_1}
\end{aligned}$$

because

$$\begin{aligned}
1.^T[\boldsymbol{x_n} \cdot \boldsymbol{\gamma_n^T}] &= (1.^T \boldsymbol{x_n}) \cdot \boldsymbol{\gamma_n^T} = 1 \cdot \boldsymbol{\gamma_n^T} = \boldsymbol{\gamma_n^T} \\
1.^T \boldsymbol{\xi_n} = 1.^T \vec{p}(\boldsymbol{z_n z_{n-1}^T} | \boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}) &= \vec{p}(\boldsymbol{z_{n-1}} | \boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}})^T = \boldsymbol{\gamma_{n-1}^T} \\
1.^T \boldsymbol{\gamma_n} &= 1
\end{aligned}$$