Purpose: 期望最大化算法

$$egin{aligned} \max && \ln p(oldsymbol{X}|oldsymbol{ heta}) = \left\{\sum_{oldsymbol{Z}} q(oldsymbol{Z})
ight\} \ln p(oldsymbol{X}|oldsymbol{ heta}) \ &= \left\{\sum_{oldsymbol{Z}} q(oldsymbol{Z})
ight\} \ln rac{p(oldsymbol{X}, oldsymbol{Z}|oldsymbol{ heta})}{p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta})} \ &= \mathcal{L}(q, oldsymbol{ heta}) + \mathrm{KL}(q||p) \end{aligned}$$

where X is **constant**, parameter q(Z), θ , **random variable** Z;

X, Z are discrete matrix

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

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
- KL(q||p) 概率分布 $q(\mathbf{Z})$ 的一个泛函, $m{ heta}$ 函数,KL散度 of $p(Z|X, heta),\ q(Z),\geq 0$

EM法

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

[M步骤]

$$egin{aligned} & ext{fix } q \equiv q^k = p\left(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^k) = p\left(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{ ext{old}}
ight) \ & ext{θ} = rgmax \sum_{oldsymbol{ heta}} q^k \ln \left\{ rac{p(oldsymbol{X}, oldsymbol{Z}|oldsymbol{ heta})}{q^k}
ight\} \ & ext{θ} = rgmax \sum_{oldsymbol{ heta}} q^k \ln p(oldsymbol{X}, oldsymbol{Z}|oldsymbol{ heta}) - \sum_{oldsymbol{Z}} q^k \ln (q^k)} \mathcal{L}(q^k, oldsymbol{ heta}^{k+1}) > \mathcal{L}(q^k, oldsymbol{ heta}^k) = \ln p(oldsymbol{X}|oldsymbol{ heta}^k) \ & ext{θ} = rgmax \sum_{oldsymbol{ heta}} q^k \ln p(oldsymbol{X}, oldsymbol{Z}|oldsymbol{ heta}) \end{aligned}$$

where $oldsymbol{ heta}$ $^{ ext{old}}$ $\equiv oldsymbol{ heta}^k$

$$oldsymbol{Q}\left(oldsymbol{ heta},oldsymbol{ heta}^{
m old}
ight) \equiv \sum_{oldsymbol{Z}} q^k \ln p(oldsymbol{X},oldsymbol{Z}|oldsymbol{ heta})$$

[E步骤]

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

$$egin{align*} q^{k+1} &\equiv rgmax \mathcal{L}(q, oldsymbol{ heta}^{k+1}) \ &\equiv rgmin \mathrm{KL}(q|p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{k+1})) = 0 \ &= p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{k+1}) \ &= p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{k+1}) \ &= \mathcal{L}(q, oldsymbol{ heta}^{k+1}) + \mathrm{KL}(q|p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{k+1})) \ &= \mathcal{L}(q^k, oldsymbol{ heta}^{k+1}) + \mathrm{KL}(q^k|p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{k+1})) \ &= \mathcal{L}(q^{k+1}, oldsymbol{ heta}^{k+1}) + \mathrm{KL}(q^{k+1}|p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}^{k+1})) \ &= \mathcal{L}(q^{k+1}, oldsymbol{ heta}^{k+1}) \ \end{pmatrix}$$

in all

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

So

$$egin{aligned} \lim_{k o \infty} \ln p(oldsymbol{X} | oldsymbol{ heta}^k) &= \lim_{k o \infty} \mathcal{L}(q^k, oldsymbol{ heta}^{k+1}) = \max_{oldsymbol{ heta}} \ln p(oldsymbol{X} | oldsymbol{ heta}) \ oldsymbol{ heta}_{max} &\equiv \max_{oldsymbol{ heta}} \ln p(oldsymbol{X} | oldsymbol{ heta}) \ &= \lim_{k o \infty} lpha^k \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}) \ &= \lim_{k o \infty} rgmax oldsymbol{Q} \left(oldsymbol{ heta}, oldsymbol{ heta}^k
ight) \end{aligned}$$

formula

To sum up

$$egin{aligned} oldsymbol{lpha_1} &= [oldsymbol{\phi}^T \cdot oldsymbol{x_1}] \odot oldsymbol{\pi} \ oldsymbol{lpha_n} &= [oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot [oldsymbol{A} \cdot oldsymbol{lpha_{n-1}}] \ oldsymbol{eta_N} &= oldsymbol{A}^T \cdot \left([oldsymbol{\phi}^T \cdot oldsymbol{x_{n+1}}] \odot oldsymbol{eta_{n+1}}
ight) \end{aligned}$$

So if α_n, β_n is obtained, we could get γ_n, ξ_n

$$oldsymbol{lpha_n} egin{aligned} oldsymbol{lpha_n} & \equiv ec{p}(oldsymbol{z_n}, oldsymbol{x_1}, \dots, oldsymbol{x_N} | oldsymbol{ heta}^{
m old}) \ oldsymbol{eta_n} & \equiv ec{p}(oldsymbol{x_{n+1}}, \dots, oldsymbol{x_N} | oldsymbol{z_n}, oldsymbol{ heta}^{
m old}) \end{aligned}$$

Here

$$m{\gamma_n} \equiv \left[\sum_{m{z_n}} p\left(m{z_n}|m{X},m{ heta}^{
m old}
ight)m{z_n}
ight]$$
 为常数,是期望 $E[m{z_n}|m{X},m{ heta}^{
m old}] = ec{p}(m{z_n}|m{X},m{ heta}^{
m old})$,大小 $K imes 1$

$$oldsymbol{\xi_n} \equiv [\sum_{oldsymbol{z_{n-1}, z_n}} oldsymbol{z_n p} \left(oldsymbol{z_{n-1}, z_n} | oldsymbol{X}, oldsymbol{ heta} \, \mathrm{old} \,
ight) oldsymbol{z_{n-1}^T}]$$
 为常数,是期望 $E[oldsymbol{z_n z_{n-1}^T} | oldsymbol{X}, oldsymbol{ heta} \, \mathrm{old} \,] = ec{p}(oldsymbol{z_n z_{n-1}^T} | oldsymbol{X}, oldsymbol{ heta} \, \mathrm{old} \,)$,大小 $K imes K$

$$egin{aligned} oldsymbol{\gamma_n} &= oldsymbol{lpha_n} \odot oldsymbol{eta_n} \odot rac{1.}{ec{p}(oldsymbol{X} | oldsymbol{ heta}^{
m old})} \ &= rac{1}{1.^T \cdot oldsymbol{lpha_N}} igl[oldsymbol{lpha_n} \odot oldsymbol{eta_n} igr] \ oldsymbol{\xi_n} &= rac{1}{p(oldsymbol{X} | oldsymbol{ heta}^{
m old})} iggl\{ igl[oldsymbol{\phi}^T \cdot oldsymbol{x_n} igr] \odot oldsymbol{eta_n} igr] \cdot oldsymbol{lpha_{n-1}} igr\} \odot oldsymbol{A} \ &= rac{1}{1.^T \cdot oldsymbol{lpha_N}} iggl\{ igl[oldsymbol{\phi}^T \cdot oldsymbol{x_n} igr] \odot oldsymbol{eta_n} igr] \cdot oldsymbol{lpha_{n-1}} igr\} \odot oldsymbol{A} \end{aligned}$$

其中
$$p(m{X}|m{ heta}^{
m old}\,) = \sum_{Z_N} p(m{z_N},m{X}|m{ heta}^{
m old}\,) = \sum m{lpha_N} = 1.^T \cdot m{lpha_N}$$
 is constant

$$egin{aligned} ec{p}(oldsymbol{X}|oldsymbol{ heta}^{ ext{ old }}) &= p(oldsymbol{X}|oldsymbol{ heta}^{ ext{ old }}) \cdot 1. \ p(oldsymbol{X}|oldsymbol{ heta}^{ ext{ old }}) &= 1.^T \cdot p(oldsymbol{z_N}, oldsymbol{X}|oldsymbol{ heta}^{ ext{ old }}) \ &= 1.^T \cdot oldsymbol{lpha}_{oldsymbol{N}} \end{aligned}$$

after fixation

we divide $p(oldsymbol{X}|oldsymbol{ heta}^{
m old}$) into 2 part

$$egin{aligned} p(oldsymbol{X}|oldsymbol{ heta}^{
m old}) &= p(oldsymbol{x_1}, \cdots, oldsymbol{x_n}|oldsymbol{ heta}^{
m old}) p(oldsymbol{x_{n+1}}, \cdots, oldsymbol{x_N}|oldsymbol{x_1}, \cdots, oldsymbol{x_n}, oldsymbol{ heta}^{
m old}) \ &= \left\{ \prod_{k=1}^n p(oldsymbol{x_k}|oldsymbol{x_1}, \cdots, oldsymbol{x_{k-1}}, oldsymbol{ heta}^{
m old})
ight\} \ &= \left\{ \prod_{k=1}^n c_k
ight\} \left\{ \prod_{k=n+1}^N c_k
ight\} \end{aligned}$$

define

where we get c_n by

$$egin{aligned} c_n &\equiv 1.^T c_n \cdot oldsymbol{lpha_n} = 1.^T \cdot \left\{ [oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot [oldsymbol{A} \cdot \hat{oldsymbol{lpha_{n-1}}}]
ight\} \ c_{n+1}
eq 1.^T c_{n+1} \cdot oldsymbol{eta_n} = 1.^T \cdot oldsymbol{A}^T \cdot \left([oldsymbol{\phi}^T \cdot oldsymbol{x_{n+1}}] \odot \hat{oldsymbol{eta}_{n+1}}
ight) \end{aligned}$$

also

$$egin{aligned} oldsymbol{\hat{lpha}_1} &= rac{1}{c_1} [oldsymbol{\phi}^T \cdot oldsymbol{x_1}] \odot oldsymbol{\pi} \ oldsymbol{\hat{lpha}_n} &= rac{1}{c_n} [oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot [oldsymbol{A} \cdot oldsymbol{\hat{lpha}_{n-1}}] \ oldsymbol{\hat{eta}_N} &= 1. \ oldsymbol{\hat{eta}_n} &= rac{1}{c_{n+1}} oldsymbol{A}^T \cdot \left([oldsymbol{\phi}^T \cdot oldsymbol{x_{n+1}}] \odot oldsymbol{\hat{eta}_{n+1}}
ight) \end{aligned}$$

thus

$$egin{aligned} oldsymbol{\gamma_n} &= oldsymbol{lpha_n} \odot oldsymbol{eta_n} \odot oldsymbol{eta_n} \odot oldsymbol{eta_n} \ &= [\hat{oldsymbol{lpha_n}} \odot \hat{oldsymbol{eta_n}}] \ oldsymbol{\xi_n} &= rac{1}{p(oldsymbol{X} | oldsymbol{ heta}
m{ old })} iggl\{ igl[oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot oldsymbol{eta_n} igr] \cdot oldsymbol{lpha_{n-1}} igr\} \odot oldsymbol{A} \ &= rac{1}{c_n} iggl\{ igl[oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot \hat{oldsymbol{eta_n}} igr] \cdot \hat{oldsymbol{lpha_{n-1}}} igr\} \odot oldsymbol{A} \end{aligned}$$

close form of Q

$$egin{aligned} &\max_{oldsymbol{ heta}} &oldsymbol{Q}\left(oldsymbol{ heta}, oldsymbol{ heta} ext{ old }
ight) \equiv \sum_{oldsymbol{Z}} p\left(oldsymbol{Z} | oldsymbol{X}, oldsymbol{ heta} ext{ old }
ight) \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}) \ &= \sum_{n=1}^N oldsymbol{\gamma}_{oldsymbol{n}}^T < \ln oldsymbol{\phi}, oldsymbol{x}_{oldsymbol{n}} > + \sum_{n=2}^N \operatorname{tr}\left(oldsymbol{\xi}_{oldsymbol{n}}^T \ln oldsymbol{A}
ight) + oldsymbol{\gamma}_{oldsymbol{1}}^T \ln oldsymbol{\pi} \end{aligned}$$

subject to

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

update $oldsymbol{ heta}^{k+1} \equiv \{oldsymbol{\phi}, oldsymbol{A}, oldsymbol{\pi}\}$

$$egin{aligned} oldsymbol{\phi} &\Leftarrow \left\{ \sum_{n=1}^{N} [oldsymbol{x}_{n} \cdot oldsymbol{\gamma}_{n}^{T}]
ight\} \odot rac{1}{1 \cdot \cdot 1 \cdot^{T} \left\{ \sum_{n=1}^{N} [oldsymbol{x}_{n} \cdot oldsymbol{\gamma}_{n}^{T}]
ight\}} \ &= \left\{ \sum_{n=1}^{N} [oldsymbol{x}_{n} \cdot oldsymbol{\gamma}_{n}^{T}]
ight\} \odot rac{1}{1 \cdot \left[\sum_{n=1}^{N} oldsymbol{\gamma}_{n} \right]^{T}} \ &= [\sum_{n=2}^{N} oldsymbol{\xi}_{n}] \odot rac{1}{1 \cdot \cdot \left[\sum_{n=2}^{N} oldsymbol{\gamma}_{n-1} \right]^{T}} \ &= [\sum_{n=2}^{N} oldsymbol{\xi}_{n}] \odot rac{1}{1 \cdot \cdot \left[\sum_{n=1}^{N-1} oldsymbol{\gamma}_{n} \right]^{T}} \ &= \left[\sum_{n=2}^{N} oldsymbol{\xi}_{n} \right] \odot rac{1}{1 \cdot \cdot \left[\sum_{n=1}^{N-1} oldsymbol{\gamma}_{n} \right]^{T}} \ &\pi \Leftarrow oldsymbol{\gamma}_{1} \odot rac{1}{1 \cdot \cdot 1 \cdot^{T} oldsymbol{\gamma}_{1}} = oldsymbol{\gamma}_{1} \end{aligned}$$

because

$$egin{aligned} 1.^T [oldsymbol{x_n} \cdot oldsymbol{\gamma_n}^T] &= (1.^T oldsymbol{x_n}) \cdot oldsymbol{\gamma_n}^T = 1 \cdot oldsymbol{\gamma_n}^T = oldsymbol{\gamma_n}^T \ 1.^T oldsymbol{\xi_n} &= 1.^T ec{p}(oldsymbol{z_n} oldsymbol{z_{n-1}}^T | oldsymbol{X}, oldsymbol{ heta}^{
m old}) = ec{p}(oldsymbol{z_{n-1}} | oldsymbol{X}, oldsymbol{ heta}^{
m old})^T = oldsymbol{\gamma_{n-1}}^T \ 1.^T oldsymbol{\gamma_n} &= 1 \end{aligned}$$