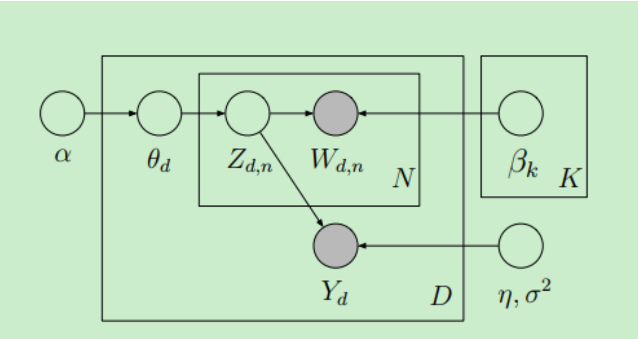
Supervised topic model

概率图



变分参数

Model Parameter	Prior Distribution	Variational Distribution
$\boldsymbol{\theta}_d$	$\mathrm{Dir}(\boldsymbol{\alpha})$	$\mathrm{Dir}(\boldsymbol{\gamma}_d)$
$z_{d,n}$	$\mathrm{Mult}(\boldsymbol{\theta}_d)$	$\mathrm{Mult}(\boldsymbol{\phi}_{d,n})$

ELBO

根据对数似然

$$\begin{aligned}
&\log p(w_{1:N}, y, \mid \boldsymbol{\alpha}, \boldsymbol{\beta}_{1:K}, \eta, \sigma^2) \geq \mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\phi}_{1:N}; \boldsymbol{\alpha}, \boldsymbol{\beta}_{1:K}, \eta, \sigma^2) \\
&\qquad = \mathbb{E}[\log p(\boldsymbol{\theta} \mid \boldsymbol{\alpha})] + \sum_{n=1}^{N} \mathbb{E}[\log p(\mathbf{Z}_n \mid \boldsymbol{\theta})] + \sum_{n=1}^{N} \mathbb{E}[\log p(w_n \mid \mathbf{Z}_n, \boldsymbol{\beta}_{1:K})] + \mathbb{E}[\log p(y \mid \mathbf{Z}_{1:N}, \eta, \sigma^2)] + \mathrm{H}(q)
\end{aligned}$$

第一项

$$\begin{aligned}
\mathbb{E}[\log p(\boldsymbol{\theta} \mid \boldsymbol{\alpha})] &= \log \Gamma\left(\sum_{k=1}^{K} \alpha_k\right) - \sum_{k=1}^{K} \log \Gamma(\alpha_k) + \sum_{k=1}^{K} (\alpha_k - 1) \mathbb{E}_{q(\boldsymbol{\theta}|\boldsymbol{\gamma})}[\log \theta_k] \\
&= \log \Gamma\left(\sum_{k=1}^{K} \alpha_k\right) - \sum_{k=1}^{K} \log \Gamma(\alpha_k) + \sum_{k=1}^{K} (\alpha_k - 1) \left(\Psi(\gamma_k) - \Psi\left(\sum_{k'=1}^{K} \gamma_{k'}\right)\right)
\end{aligned}$$

第二项

$$\begin{aligned}
\sum_{n=1}^{N} \mathbb{E}[\log p(\mathbf{Z}_n \mid \boldsymbol{\theta})] &= \sum_{n=1}^{N} \mathbb{E}_{q(z_n, \boldsymbol{\theta}|\boldsymbol{\phi}_n, \boldsymbol{\gamma})}[\log \prod_{k=1}^{K} \theta_k^{\mathbb{I}(z_n = k)}] \\
&= \sum_{n=1}^{N} \mathbb{E}_{q(z_n, \boldsymbol{\theta}|\boldsymbol{\phi}_n, \boldsymbol{\gamma})}[\sum_{k=1}^{K} \mathbb{I}(z_n = k) \log \theta_k] \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \mathbb{E}_{q(z_n|\boldsymbol{\phi}_n)}[\mathbb{I}(z_n = k)] \mathbb{E}_{q(\boldsymbol{\theta}|\boldsymbol{\gamma})}[\log \theta_k] \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \phi_{n,k} \left(\Psi(\gamma_k) - \Psi\left(\sum_{k'=1}^{K} \gamma_{k'}\right)\right)
\end{aligned}$$

转化为不含 n 的形式（隐含假设相同该文档相同单词来自同一个话题）

第三项

$$\begin{aligned}
\sum_{n=1}^{N} \mathbb{E}[\log p(w_n \mid \mathbf{Z}_n, \boldsymbol{\beta}_{1:K})] &= \sum_{n=1}^{N} \mathbb{E}_{q(z_n|\boldsymbol{\phi}_n)}[\log \prod_{k=1}^{K} \beta_{k,w_n}^{\mathbb{I}(z_n = k)}] \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \mathbb{E}_{q(z_n|\boldsymbol{\phi}_n)}[\mathbb{I}(z_n = k)] \log \beta_{k,w_n} \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \phi_{n,k} \log \beta_{k,w_n}
\end{aligned}$$

第四项

$$
\begin{aligned}
\mathbb{E}[\log p(y \mid \mathbf{Z}_{1:N}, \eta, \sigma^2)] &= \mathbb{E}[\log \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(y-\boldsymbol{\eta}^\top \bar{\mathbf{Z}})^2}{2\sigma^2}] \\
&= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2 - 2y\boldsymbol{\eta}^\top \mathbb{E}[\bar{\mathbf{Z}}] + \boldsymbol{\eta}^\top \mathbb{E}[\bar{\mathbf{Z}}\bar{\mathbf{Z}}^\top]\boldsymbol{\eta}}{2\sigma^2} \\
&= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2} + \frac{y\boldsymbol{\eta}^\top \frac{\sum_{n=1}^N \boldsymbol{\phi}_n}{N}}{\sigma^2} - \frac{\boldsymbol{\eta}^\top \mathbb{E}[\frac{\sum_{n=1}^N \mathbf{Z}_n}{N} \frac{\sum_{n=1}^N \mathbf{Z}_n^\top}{N}]\boldsymbol{\eta}}{2\sigma^2} \\
&= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2} + \frac{y\boldsymbol{\eta}^\top \sum_{n=1}^N \boldsymbol{\phi}_n}{N\sigma^2} - \frac{\boldsymbol{\eta}^\top \mathbb{E}[\sum_{n=1}^N \sum_{m=1}^N \mathbf{Z}_n \mathbf{Z}_m^\top]\boldsymbol{\eta}}{2N^2\sigma^2} \\
&= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2} + \frac{y\boldsymbol{\eta}^\top \sum_{n=1}^N \boldsymbol{\phi}_n}{N\sigma^2} - \frac{\boldsymbol{\eta}^\top \mathbb{E}[\sum_{n=1}^N \sum_{m\neq n}^N \mathbf{Z}_n \mathbf{Z}_m^\top + \sum_{n=1}^N \mathbf{Z}_n \mathbf{Z}_n^\top]\boldsymbol{\eta}}{2N^2\sigma^2} \\
&= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2} + \frac{y\boldsymbol{\eta}^\top \sum_{n=1}^N \boldsymbol{\phi}_n}{N\sigma^2} - \frac{\boldsymbol{\eta}^\top \left(\sum_{n=1}^N \sum_{m\neq n}^N \boldsymbol{\phi}_n \boldsymbol{\phi}_m^\top + \sum_{n=1}^N \mathrm{diag}(\boldsymbol{\phi}_n)\right)\boldsymbol{\eta}}{2N^2\sigma^2}
\end{aligned}
$$

第五项

$$
\begin{aligned}
\mathrm{H}(q) &= \mathrm{H}(q(\boldsymbol{\theta} \mid \boldsymbol{\gamma})) + \sum_{n=1}^N \mathrm{H}(q(\mathbf{Z}_n \mid \boldsymbol{\phi}_n)) \\
&= -\log\Gamma\left(\sum_{k=1}^K \gamma_k\right) + \sum_{k=1}^K \log\Gamma(\gamma_k) - \sum_{k=1}^K (\gamma_k - 1)\left(\Psi(\gamma_k) - \Psi\left(\sum_{k'=1}^K \gamma_{k'}\right)\right) - \sum_{n=1}^N \sum_{k=1}^K \phi_{n,k}\log\phi_{n,k}
\end{aligned}
$$

最后得到所有文档的ELBO

$$
\begin{aligned}
\mathrm{ELBO} =& D\log\Gamma\left(\sum_{k=1}^K \alpha_k\right) - D\sum_{k=1}^K \log\Gamma(\alpha_k) + \sum_{d=1}^D \sum_{k=1}^K (\alpha_k - 1)\left(\Psi(\gamma_{d,k}) - \Psi\left(\sum_{k'=1}^K \gamma_{d,k'}\right)\right) \\
&+ \sum_{d=1}^D \sum_{n=1}^{N_d} \sum_{k=1}^K \phi_{d,n,k}\left(\Psi(\gamma_{d,k}) - \Psi\left(\sum_{k'=1}^K \gamma_{d,k'}\right)\right) \\
&+ \sum_{d=1}^D \sum_{n=1}^{N_d} \sum_{k=1}^K \phi_{d,n,k}\log\beta_{k,w_{d,n}} \\
&- \frac{D}{2}\log(2\pi\sigma^2) - \sum_{d=1}^D \frac{y_d^2}{2\sigma^2} + \sum_{d=1}^D \frac{y_d\boldsymbol{\eta}^\top \sum_{n=1}^{N_d} \boldsymbol{\phi}_{d,n}}{N_d\sigma^2} - \sum_{d=1}^D \frac{\boldsymbol{\eta}^\top \left(\sum_{n=1}^{N_d} \sum_{m\neq n}^{N_d} \boldsymbol{\phi}_{d,n}\boldsymbol{\phi}_{d,m}^\top + \sum_{n=1}^{N_d} \mathrm{diag}(\boldsymbol{\phi}_{d,n})\right)\boldsymbol{\eta}}{2N_d^2\sigma^2} \\
&- \sum_{d=1}^D \log\Gamma\left(\sum_{k=1}^K \gamma_{d,k}\right) + \sum_{d=1}^D \sum_{k=1}^K \log\Gamma(\gamma_{d,k}) - \sum_{d=1}^D \sum_{k=1}^K (\gamma_{d,k} - 1)\left(\Psi(\gamma_{d,k}) - \Psi\left(\sum_{k'=1}^K \gamma_{d,k'}\right)\right) - \sum_{d=1}^D \sum_{n=1}^{N_d} \sum_{k=1}^K \phi_{d,n,k}\log\phi_{d,n,k}
\end{aligned}
$$

EM

E步：更新变分参数

更新 $\boldsymbol{\gamma}_d$：

$$
\begin{aligned}
\mathcal{L}_{[\gamma_d]} =& \sum_{k=1}^K (\alpha_k - 1)\left(\Psi(\gamma_{d,k}) - \Psi\left(\sum_{k'=1}^K \gamma_{d,k'}\right)\right) \\
&+ \sum_{n=1}^{N_d} \sum_{k=1}^K \phi_{d,n,k}\left(\Psi(\gamma_{d,k}) - \Psi\left(\sum_{k'=1}^K \gamma_{d,k'}\right)\right) \\
&- \log\Gamma\left(\sum_{k=1}^K \gamma_{d,k}\right) + \sum_{k=1}^K \log\Gamma(\gamma_{d,k}) - \sum_{k=1}^K (\gamma_{d,k} - 1)\left(\Psi(\gamma_{d,k}) - \Psi\left(\sum_{k'=1}^K \gamma_{d,k'}\right)\right) \\
\frac{\partial\mathcal{L}}{\partial\gamma_{d,k}} =& (\alpha_k - 1)\Psi'(\gamma_{d,k}) - \sum_{k=1}^K (\alpha_k - 1)\Psi'(\sum_{k'=1}^K \gamma_{d,k'}) \\
&+ \sum_{n=1}^{N_d} \phi_{d,n,k}\Psi'(\gamma_{d,k}) - \sum_{n=1}^{N_d} \sum_{k=1}^K \phi_{d,n,k}\Psi'(\sum_{k'=1}^K \gamma_{d,k'}) \\
&- \Psi(\sum_{k=1}^K \gamma_{d,k}) + \Psi(\gamma_{d,k}) - \Psi(\gamma_{d,k}) - (\gamma_{d,k} - 1)\Psi'(\gamma_{d,k}) + \Psi(\sum_{k'=1}^K \gamma_{d,k'}) + \sum_{k=1}^K (\gamma_{d,k} - 1)\Psi'(\sum_{k'=1}^K \gamma_{d,k'})
\end{aligned}
$$

令其偏导为0，可得更新公式

$$
\gamma_{d,k} = \alpha_k + \sum_{n=1}^{N_d} \phi_{d,n,k}
$$

更新 $\boldsymbol{\phi}_{d,n}$：

$$
\begin{aligned}
\mathcal{L}_{[\phi_{d,n}]} =& \sum_{k=1}^{K} \phi_{d,n,k} \left(\Psi(\gamma_{d,k}) - \Psi\left(\sum_{k'=1}^{K} \gamma_{d,k'} \right) \right) \\
&+ \sum_{k=1}^{K} \phi_{d,n,k} \log \beta_{k,w_{d,n}} \\
&+ \frac{y_d \boldsymbol{\eta}^\top \boldsymbol{\phi}_{d,n}}{N_d \sigma^2} - \frac{\boldsymbol{\eta}^\top \left(2 \sum_{m \neq n}^{N_d} \boldsymbol{\phi}_{d,n} \boldsymbol{\phi}_{d,m}^\top + \operatorname{diag}(\boldsymbol{\phi}_{d,n}) \right) \boldsymbol{\eta}}{2 N_d^2 \sigma^2} \\
&- \sum_{k=1}^{K} \phi_{d,n,k} \log \phi_{d,n,k} \\
\frac{\partial \mathcal{L}}{\partial \phi_{d,n,k}} =& \Psi(\gamma_{d,k}) - \Psi\left(\sum_{k'=1}^{K} \gamma_{d,k'} \right) \\
&+ \log \beta_{k,w_{d,n}} \\
&+ \frac{y_d \eta_k}{N_d \sigma^2} - \frac{2 \sum_{k'=1}^{K} \eta_{k'} \sum_{m \neq n}^{N_d} \phi_{d,m,k'} \eta_k + \eta_k^2}{2 N_d^2 \sigma^2} \\
&- \log \phi_{d,n,k} - 1
\end{aligned}
$$

考虑 $\sum_{k=1}^{K} \phi_{d,n,k} = 1$ 约束，令拉格朗日函数偏导为0

$$
\phi_{d,n,k} \propto \exp\left(\Psi(\gamma_{d,k}) - \Psi\left(\sum_{k'=1}^{K} \gamma_{d,k'} \right) + \log \beta_{k,w_{d,n}} + \frac{y_d \eta_k}{N_d \sigma^2} - \frac{2 \sum_{k'=1}^{K} \eta_{k'} \sum_{m \neq n}^{N_d} \phi_{d,m,k'} \eta_k + \eta_k^2}{2 N_d^2 \sigma^2} \right)
$$

M步：更新模型参数

更新 $\boldsymbol{\beta}_k$：

$$
\begin{aligned}
\mathcal{L}_{[\boldsymbol{\beta}_k]} &= \sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{d,n,k} \log \beta_{k,w_{d,n}} \\
\frac{\partial \mathcal{L}}{\partial \beta_{k,v}} &= \sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{d,n,k} \mathbb{I}(w_{d,n} = v) \frac{1}{\beta_{k,v}}
\end{aligned}
$$

考虑 $\sum_{v=1}^{V} \beta_{k,v} = 1$ 约束，令拉格朗日函数偏导为0

$$
\beta_{k,v} \propto \sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{d,n,k} \mathbb{I}(w_{d,n} = v)
$$

更新 $\boldsymbol{\eta}$，利用展开的形式不可取：

$$
\begin{aligned}
\mathcal{L}_{[\boldsymbol{\eta}]} &= \sum_{d=1}^{D} \frac{y_d \boldsymbol{\eta}^\top \sum_{n=1}^{N_d} \boldsymbol{\phi}_{d,n}}{N_d \sigma^2} - \sum_{d=1}^{D} \frac{\boldsymbol{\eta}^\top \left(\sum_{n=1}^{N_d} \sum_{m \neq n}^{N_d} \boldsymbol{\phi}_{d,n} \boldsymbol{\phi}_{d,m}^\top + \sum_{n=1}^{N_d} \operatorname{diag}(\boldsymbol{\phi}_{d,n}) \right) \boldsymbol{\eta}}{2 N_d^2 \sigma^2} \\
&= \sum_{k=1}^{K} \eta_k \sum_{d=1}^{D} \frac{y_d}{N_d \sigma^2} \sum_{n=1}^{N} \phi_{d,n,k} - \sum_{d=1}^{D} \frac{1}{2 N_d^2 \sigma^2} \sum_{i=1}^{K} \sum_{k=1}^{K} \eta_i \eta_k \sum_{n=1}^{N_d} \phi_{d,n,k} \sum_{m \neq n}^{N_d} \phi_{d,m,i} - \sum_{d=1}^{D} \frac{1}{2 N_d^2 \sigma^2} \sum_{n=1}^{N_d} \sum_{k=1}^{K} \eta_k^2 \phi_{d,n,k} \\
\frac{\partial \mathcal{L}}{\partial \eta_k} &= \sum_{d=1}^{D} \frac{y_d}{N_d \sigma^2} \sum_{n=1}^{N} \phi_{d,n,k} - XXX - \eta_k \sum_{d=1}^{D} \frac{1}{N_d^2 \sigma^2} \sum_{n=1}^{N_d} \phi_{d,n,k}
\end{aligned}
$$

需要使用矩阵和向量操作，便于编程实现：

$$
\begin{aligned}
\mathcal{L}_{[\boldsymbol{\eta}]} &= \sum_{d=1}^{D} \frac{y_d \boldsymbol{\eta}^\top \sum_{n=1}^{N_d} \boldsymbol{\phi}_{d,n}}{N_d \sigma^2} - \sum_{d=1}^{D} \frac{\boldsymbol{\eta}^\top \left(\sum_{n=1}^{N_d} \sum_{m \neq n}^{N_d} \boldsymbol{\phi}_{d,n} \boldsymbol{\phi}_{d,m}^\top + \sum_{n=1}^{N_d} \operatorname{diag}(\boldsymbol{\phi}_{d,n}) \right) \boldsymbol{\eta}}{2 N_d^2 \sigma^2} \\
\frac{\partial \mathcal{L}}{\partial \boldsymbol{\eta}} &= \sum_{d=1}^{D} \frac{y_d}{N_d \sigma^2} \sum_{n=1}^{N_d} \boldsymbol{\phi}_{d,n} - \sum_{d=1}^{D} \frac{1}{N_d^2 \sigma^2} \sum_{n=1}^{N_d} \left(\sum_{m \neq n}^{N_d} \boldsymbol{\phi}_{d,n} \boldsymbol{\phi}_{d,m}^\top + \operatorname{diag}(\boldsymbol{\phi}_{d,n}) \right) \boldsymbol{\eta}
\end{aligned}
$$

令其偏导为0，可得

$$
\boldsymbol{\eta} = \left(\sum_{d=1}^{D} \frac{1}{N_d^2} \sum_{n=1}^{N_d} \left(\sum_{m \neq n}^{N_d} \boldsymbol{\phi}_{d,n} \boldsymbol{\phi}_{d,m}^\top + \operatorname{diag}(\boldsymbol{\phi}_{d,n}) \right) \right)^{-1} \left(\sum_{d=1}^{D} \frac{y_d}{N_d} \sum_{n=1}^{N_d} \boldsymbol{\phi}_{d,n} \right)
$$

更新 σ^2：

$$
\begin{aligned}
\mathcal{L}_{[\sigma^2]} =& -\frac{D}{2} \log(2\pi\sigma^2) - \sum_{d=1}^{D} \frac{y_d^2}{2\sigma^2} + \sum_{d=1}^{D} \frac{y_d \boldsymbol{\eta}^\top \sum_{n=1}^{N_d} \boldsymbol{\phi}_{d,n}}{N_d \sigma^2} - \sum_{d=1}^{D} \frac{\boldsymbol{\eta}^\top \left(\sum_{n=1}^{N_d} \sum_{m \neq n}^{N_d} \boldsymbol{\phi}_{d,n} \boldsymbol{\phi}_{d,m}^\top + \sum_{n=1}^{N_d} \operatorname{diag}(\boldsymbol{\phi}_{d,n}) \right) \boldsymbol{\eta}}{2 N_d^2 \sigma^2} \\
\frac{\partial \mathcal{L}}{\partial \sigma^2} =& -\frac{D}{2} \frac{1}{\sigma^2} + \frac{\sum_{d=1}^{D} y_d^2}{2} \frac{1}{(\sigma^2)^2} - \sum_{d=1}^{D} \frac{y_d \boldsymbol{\eta}^\top}{N_d} \sum_{n=1}^{N_d} \boldsymbol{\phi}_{d,n} \frac{1}{(\sigma^2)^2} + \boldsymbol{\eta}^\top \sum_{d=1}^{D} \frac{\sum_{n=1}^{N_d} \sum_{m \neq n}^{N_d} \boldsymbol{\phi}_{d,n} \boldsymbol{\phi}_{d,m}^\top + \sum_{n=1}^{N_d} \operatorname{diag}(\boldsymbol{\phi}_{d,n})}{2 N_d^2} \boldsymbol{\eta} \frac{1}{(\sigma^2)^2} \quad \text{note: the upate of } \boldsymbol{\eta} \\
=& \frac{-D\sigma^2 + \sum_{d=1}^{D} y_d^2 - 2\boldsymbol{\eta}^\top \sum_{d=1}^{D} \frac{y_d}{N_d} \sum_{n=1}^{N_d} \boldsymbol{\phi}_{d,n} + \boldsymbol{\eta}^\top \sum_{d=1}^{D} \frac{y_d}{N_d} \sum_{n=1}^{N_d} \boldsymbol{\phi}_{d,n}}{2(\sigma^2)^2}
\end{aligned}
$$

令其偏导为0，可得

$$\sigma^2 = \frac{1}{D}\left(\sum_{d=1}^{D} y_d^2 - \boldsymbol{\eta}^\top \sum_{d=1}^{D} \frac{y_d}{N_d} \sum_{n=1}^{N_d} \boldsymbol{\phi}_{d,n}\right)$$