LDA

$W = \{w^d\} = \{w_i^d\}$ $\quad 1 \le d \le D \quad 1 \le l \le L_d$

$\Phi = \{\varphi^t\} = \{\varphi_v^t\}$ $\quad 1 \le t \le T,\ 1 \le v \le V \quad \sum_v \varphi_v^t = 1$

$\Theta = \{\theta^d\}, \quad 1 \le d \le D$

$\quad = \{\theta_t^d\} \quad 1 \le t \le T \quad \sum_t \theta_t^d = 1$

$Z = \{z^d\} \quad 1 \le d \le D$

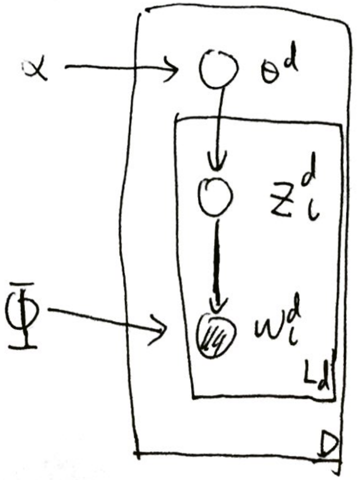
$\quad = \{z_i^d\} \quad 1 \le l \le L_d$

$\theta^d \sim \text{Dirichlet}(\alpha)$

$z_i^d \sim \text{multi}(\cdot \mid \theta^d)$

$w_i^d \sim P(\cdot \mid z_i^d, \Phi) = \text{multi}(\cdot \mid \varphi^{z_i^d})$

概率图



$P(\theta^d \mid \alpha)$

$P(z_i^d \mid \theta^d)$

$P(w_i^d \mid \varphi^{z_i^d})$

$P(\Theta, Z, W \mid \alpha, \Phi)$

$= P(\Theta \mid \alpha) P(Z \mid \Theta) P(W \mid Z, \Phi)$

$= \left[\prod_d P(\theta^d \mid \alpha)\right] \left[\prod_d \prod_i P(z_i^d \mid \theta^d)\right]$

$\left[\prod_d \prod_i P(w_i^d \mid z_i^d, \Phi)\right]$

LDA 推理

问题： $P(\theta, z, w) \Rightarrow P(\theta, z | w)$

变分方法： 变分分布族 $Q(\theta, z | \gamma, \lambda)$

$$\min_{\gamma,\lambda} KL(Q(\theta, z | \gamma, \lambda) \| P(\theta, z | w))$$

$$KL(Q \| P(\theta, z | w))$$

$$= E_Q[\log Q] - E_Q[\log P(\theta, z, w)] + E_Q[\log P(w)]$$

$$ELBO \triangleq E_Q[\log P(\theta, z, w)] \underbrace{- E_Q[\log Q]}_{H(Q) \geq 0 \text{ 熵}}$$

$$KL = -ELBO + \log P(w) \geq 0$$

$$\max ELBO = \max (E_Q[\log P(\theta, z, w)] - E_Q[\log Q(\theta, z)])$$

平均场假设

$$Q(\theta, z | \gamma, \lambda) = Q(\theta | \gamma) \cdot Q(z | \lambda)$$
$$= \prod_d Q(\theta^d | \gamma^d) \prod_d Q(z^d | \lambda^d)$$
$$= \prod_d Q(\theta^d | \gamma^d) \prod_d \prod_i Q(z_i^d | \lambda^d)$$

LDA 平均场变分推理 (1)

1. 问题

$$\operatorname*{argmax}_{\gamma,\lambda} ELBO(\gamma, \lambda | \alpha, \Phi)$$

$$ELBO = E_q[\log P(\theta, z, w | \alpha, \Phi) - \log q(\theta, z | \gamma, \lambda)]$$

~~joint~~

$$P(\theta, z, W) = \prod_{d=1}^{D} \left[P(\theta^d | \alpha) \left[\prod_{i=1}^{L_d} P(z_i^d | \theta^d) P(w_i^d | z_i^d, \Phi) \right] \right]$$

$$P(\theta, z, w) = P(\theta | \alpha) \left[\prod_{i=1}^{L} P(z_i | \theta) P(w_i | z_i, \Phi) \right]$$

$$q(\theta, z | \gamma, \lambda) = q(\theta | \gamma) \cdot q(z | \lambda) = q(\theta | \gamma) \cdot \prod_i q(z_i | \lambda^i)$$

$$P(\theta | \alpha) = Dir(\theta | \alpha) = \frac{\Gamma(\sum_t \alpha_t)}{\prod_t \Gamma(\alpha_t)} \prod_t \theta_t^{\alpha_t - 1}$$

$$P(z_i^d | \theta^d) = \theta^d_{z_i^d}, \quad P(w_i^d | z_i^d, \Phi) = \varphi^{z_i^d}_{w_i^d}$$

$$q(\theta | \gamma) = \frac{\Gamma(\sum_t \gamma_t)}{\prod_t \Gamma(\gamma_t)} \prod_t \theta_t^{\gamma_t - 1}$$

$\cancel{q(z^i | \lambda^i) = \lambda^i_{z^i}}$ $\quad q(z_i | \lambda^i) = \lambda^i_{z_i}$

LDA 平均场变分推理(2)

2. 关于 $E_q[f(x)]$

① 线性：$E_q[af(x)+bg(x)] = aE_q[f(x)] + bE_q[g(x)]$

$$E_q(c) = c$$

② 如果 $X=(x_1,x_2)$，$q(x)=q(x_1,x_2)=q(x_1)q(x_2)$

则 $E_{q(x)}[f(x)] = E_{q(x_1)}[E_{q(x_2)}[f(x_1,x_2)]]$

③ 如 $f(x) = f(x_1)$

$$E_{q(x)}[f(x_1)] = E_{q(x_1)}[f(x_1)]$$

3. 计算 ELBO

$$\text{ELBO}(\gamma,\lambda) = L_1+L_2+L_3+L_4+L_5$$

$$L_1 = E_{q(\theta,z)}[\log p(\theta|\alpha)]$$
$$= \sum_t (\alpha_t-1)E_q[\log\theta_t] + \text{const.}$$

$$L_2 = E_q[\log\prod_i p(z_i|\theta)]$$
$$= \sum_i E_q[\log\theta_{z_i}] = \sum_i E_{q(\theta)}\left[E_{q(z_i)}[\log\theta_{z_i}]\right]$$
$$= \sum_i E_{q(\theta)}\left[\sum_t \lambda^i_t\log\theta_t\right]$$
$$= \sum_i\sum_t \lambda^i_t E_q[\log\theta_t]$$

$$L_3 = E_q[\log\prod_i p(w_i|z_i,\Phi)]$$
$$= \sum_i E_q[\log p(w_i|z_i,\Phi)]$$
$$= \sum_i E_q[\log\varphi^{z_i}_{w_i}]$$
$$= \sum_i\sum_t \lambda^i_t\cdot\log\varphi^t_{w_i}$$

$$L_4 = -E_q[\log q(\theta|\gamma)]$$
$$= -\sum_t(\gamma_t-1)E_q[\log\theta_t] - \log\Gamma(\sum_t\gamma_t) + \sum_t\log\Gamma(\gamma_t)$$

$$L_5 = -E_q[\log\prod_i q(z_i|\lambda^i)]$$
$$= -\sum_i E_q[\log q(z_i|\lambda^i)]$$
$$= -\sum_i\sum_t \lambda^i_t\log\lambda^i_t$$

定义：
$$\Psi(x) \triangleq \frac{d}{dx}(\log\Gamma(x))$$

4. 计算 $E_q[\log\theta_t]$

$$\frac{\prod_t\Gamma(\gamma_t)}{\Gamma(\sum_t\gamma_t)} = \int\prod_t\theta_t^{\gamma_t-1}d\theta$$

$$\sum_t\log\Gamma(\gamma_t) - \log\Gamma(\sum_t\gamma_t) = \log\left[\int\prod_t\theta_t^{\gamma_t-1}d\theta\right]$$

对 γ_t 求导：

$$\Psi(\gamma_t) - \Psi(\sum_t\gamma_t) = \frac{\int\prod_t\theta_t^{\gamma_t-1}\cdot\log\theta_t\,d\theta}{\int[\quad]d\theta}$$

LDA 平均场变分推理 (3)

$$\psi(\gamma_t) - \psi(\sum_t \gamma_t) = \int \frac{\Gamma(\sum_t \gamma_t)}{\prod_t \Gamma(\gamma_t)} \prod_t \theta_t^{\gamma_t - 1} \cdot \log\theta_t \, d\theta$$

$$= E_q[\log\theta_t]$$

$$\underset{\gamma,\lambda}{\operatorname{argmax}}\ E(\gamma,\lambda \mid \alpha, \Phi) = L_1 + L_2 + L_3 + L_4 + L_5$$

$$L_1 = \sum_t (\alpha_t - 1)[\psi(\gamma_t) - \psi(\sum_s \gamma_s)] + \log\Gamma(\sum_s \alpha_s) - \sum_s \log\Gamma(\alpha_s)$$

$$L_2 = \sum_l \sum_t \lambda_t^l [\psi(\gamma_t) - \psi(\sum_s \gamma_s)]$$

$$L_3 = \sum_l \sum_t \lambda_t^l \cdot \log\varphi_{w_l}^t = \sum_l \sum_t \sum_v \lambda_t^l \cdot \log\varphi_v^t \cdot I(v = w_l)$$

$$L_4 = -\sum_t (\gamma_t - 1)[\psi(\gamma_t) - \psi(\sum_s \gamma_s)] - \log\Gamma(\sum_s \gamma_s) + \sum_s \log\Gamma(\gamma_s)$$

$$L_5 = -\sum_l \sum_t \lambda_t^l \log\lambda_t^l$$

对λ求导:

$$L_2 + L_3 + L_5 + \sum_l \mu_l (\sum_t \lambda_t^l - 1)$$

$$\psi(\gamma_t) - \psi(\sum_s \gamma_s) + \log\varphi_{w_l}^t - \log\lambda_t^l - 1 + \mu_l = 0$$

$$\log\lambda_t^l = \psi(\gamma_t) - \psi(\sum_s \gamma_s) + \log\varphi_{w_l}^t - 1 + \mu_l$$

$$\lambda_t^l = \varphi_{w_l}^t \cdot \exp[\psi(\gamma_t)] \cdot C$$

$$\lambda_t^l \propto \varphi_{w_l}^t \cdot \exp(\psi(\gamma_t))$$

对γ求导:

$$L_1 + L_2 + L_4$$

$$(\alpha_t - 1)[\psi'(\gamma_t) - \psi'(\sum_s \gamma_s)] + \sum_l \lambda_t^l [\psi'(\gamma_t) - \psi'(\sum_s \gamma_s)]$$
$$- [\psi(\gamma_t) - \psi(\sum_s \gamma_s)] - (\gamma_t - 1)[\psi'(\gamma_t) - \psi'(\sum_s \gamma_s)] - \psi(\sum_s \gamma_s)$$
$$+ \psi(\gamma_t)$$

$$[\psi'(\gamma_t) - \psi'(\sum_s \gamma_s)](\alpha_t + \sum_l \lambda_t^l - \gamma_t) = 0$$

$$\gamma_t = \alpha_t + \sum_l \lambda_t^l$$

LDA 参数学习（整个文档集）

$$\arg\max_{\alpha,\Phi} \mathrm{ELBO}(\gamma,\lambda \mid \alpha,\Phi)$$

$$L_3 = \sum_d \sum_l \sum_t \sum_v \lambda_t^{dl} \cdot \log \varphi_v^t \cdot I(v = w_l^d)$$

对 Φ 寻优

$$L_3 + \sum_t \mu_t \left(\sum_v \varphi_v^t - 1 \right)$$

$$\sum_d \sum_l \lambda_t^{dl} \cdot \frac{1}{\varphi_v^t} \cdot I(v = w_l^d) + \mu_t = 0$$

$$\varphi_v^t \propto \sum_d \sum_l \lambda_t^{dl} \cdot I(v = w_l^d)$$

对 α 寻优

L_1

$$g_t \triangleq \left[\Psi(\gamma_t) - \Psi\left(\sum_s \gamma_s\right)\right] + \Psi\left(\sum_s \alpha_s\right) - \Psi(\alpha_t) = \frac{\partial L_1}{\partial \alpha_t}$$

$$\frac{\partial^2 L_1}{\partial \alpha_t \partial \alpha_{t'}} = \begin{cases} \Psi'\left(\sum_s \alpha_s\right) - \Psi'(\alpha_t) & t' = t \\ \Psi'\left(\sum_s \alpha_s\right) & t' \neq t \end{cases}$$

$$= \mathrm{diag}(h) + 1 \cdot z \cdot 1^T$$

$$h = \left[-\Psi'(\alpha_t) \right] \qquad z = \Psi'\left(\sum_s \alpha_s\right)$$

矩阵逆引理

$$(I + uv^T)^{-1} = I + a uv^T$$

$$(I + uv^T)(I + a uv^T) = I + a uv^T + uv^T + a u (v^T u) v^T$$

$$= I + (1 + a + a v^T u) uv^T$$

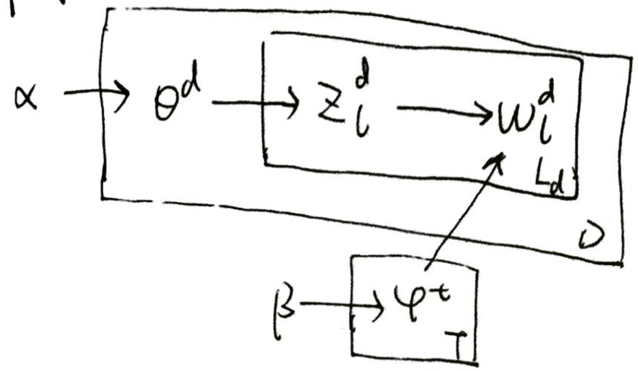
$$a = \frac{-1}{1 + v^T u}$$

$$(A + uv^T)^{-1} = \left[A(I + A^{-1} u v^T)\right]^{-1} = (I + A^{-1} u v^T)^{-1} \cdot A^{-1}$$

$$= \left(I - \frac{1}{1 + v^T A^{-1} u} A^{-1} u v^T\right) \cdot A^{-1}$$

$$= A^{-1} - \frac{1}{1 + v^T A^{-1} u} A^{-1} u v^T A^{-1}$$

平滑化



LDA 随机推理：MCMC

1. MC — Monte-Carlo 方法
2. MC — Markov-chain

① MC 方法

目标：计算 m

方法：(1) 寻找 $p(x)$, $f(x)$ 使 $m = E[f(z)]$

(2) 产生 $z_1, z_2, \cdots, z_n \sim p(x)$

(3) $$\frac{f(z_1) + \cdots + f(z_N)}{N} \doteq m$$

3. 产生指定分布的样本

(1) 一维场合

* $[0,1]$ 上的均匀分布
* 一般的 $p(x)$

$$F(x) \triangleq \int_{-\infty}^{x} p(u)du$$

$X = F^{-1}(u)$，$u \sim [0,1]$ 均匀分布

(2) 多维场合

$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ 概率有向图模型 可以

$$p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2,x_3)p(x_5|x_4)$$

Markov Chain

1. 定义 $x_0, x_1, \cdots, x_n, \cdots$

$$p(x_{n+1}|x_1,\cdots,x_n) = p(x_{n+1}|x_n)$$

如果 $p(x_{n+1}=a|x_n=b)$ 与 n 无关，齐次的/时齐的

2. 初始分布，转移概率，高阶转移概率

3. 可达与不可约；状态的周期与非周期链

不可约链：$x \to y$，$\forall$ 状态 x,y

状态的周期 $C(x) = \{n : p_{x,x}^{(n)} > 0\}$ 的最大公约数

非周期链：$C(x)=1$，$\forall x$.

4. 极限定理，平稳分布，细致平稳分布

平稳分布：$q(x)$

$$q(y) = \sum_x q(x) P_{x,y}$$

性质：极限分布是平稳分布.

细致平稳分布：

$$q(x)P_{x,y} = q(y)P_{y,x}$$

性质：细致平稳分布是平稳分布.

MCMC：Gibbs 采样

$X=\begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix}$，$p(x)=p(x_1,\cdots,x_D)$，$x_d\in[1,2,\cdots,T]$

1. 完全条件概率

$p(x_d|x_1,\cdots,x_{d-1},x_{d+1},\cdots,x_D)\triangleq p(x_d|x_{-d})$，$1\le x_d\le T$

$p(x_d|x_{-d})\ge 0$，$\sum_{x_d=1}^{D} p(x_d|x_{-d})=1$

$p(x_d|x_{-d})=\frac{p(x_d,x_{-d})}{p(x_{-d})}\propto p(x_d,x_{-d})$，$1\le x_d\le T$

2. 用 $p(x_d|x_{-d})$ 来构造马氏链

$P_d(x'|x)=\begin{cases} p(x_d'|x_{-d}'), & x_{-d}'=x_{-d} \\ 0 & 其它 \end{cases}$

$x=(x_d,x_{-d})$

$x'=(x_d',x_{-d}')$

性质：$p(x)$ 是 P_d 的细平稳分布

$p(x)P_d(x'|x)=p(x')P_d(x|x')$

3. 马氏链的复合

$P_1(x'|x)$，$P_2(x'|x)\Rightarrow P=P_1*P_2$

$P(x'|x)\triangleq\sum_z P_1(z|x)P_2(x'|z)$

性质：若 $q(x)$ 是 P_1，P_2 的平稳分布，则 $q(x)$ 也是 P 的平稳分布。

4. $P=P_1*P_2*\cdots*P_D$

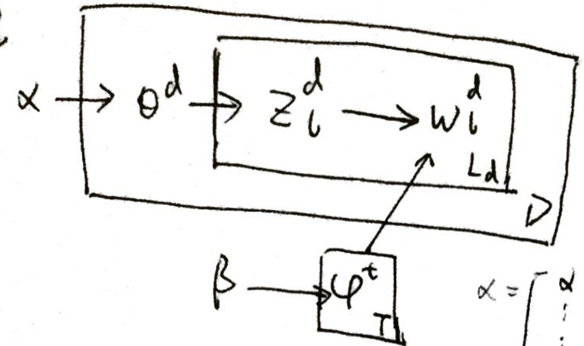
性质：$p(x)$ 是 P 的平稳分布。

P 是不可约，非周期的

$p(x)$ 是 P 的极限分布。

应用于LDA推理

1. LDA联合分布



$$\alpha=\begin{bmatrix}\alpha\\ \vdots\\ \alpha\end{bmatrix}\quad \beta=\begin{bmatrix}\beta\\ \vdots\\ \beta\end{bmatrix}$$

$$P(\Theta,\Phi,Z,W|\alpha,\beta)$$

$$=\left[\prod_d P(\theta^d|\alpha)\right]\left[\prod_t P(\varphi^t|\beta)\right]\cdot\left[\prod_d\prod_i P(z_i^d|\theta^d)\right]\left[\prod_d\prod_i P(w_i^d|z_i^d,\Phi)\right]$$

$$=\left[\prod_t P(\varphi^t|\beta)\right]\cdot\left[\prod_t\prod_v (\phi_v^t)^{n_v^t}\right]\quad n_v^t=\#\{(d,i):z_i^d=t,\ w_i^d=v\}$$

$$\cdot\left[\prod_d P(\theta^d|\alpha)\right]\cdot\left[\prod_d\prod_t (\theta_t^d)^{m_t^d}\right]\quad m_t^d=\#\{i:z_i^d=t\}$$

$$=\prod_t\left[\frac{\Gamma(V\beta)}{\Gamma(\beta)^V}\prod_v(\varphi_v^t)^{\beta-1}\right]\left[\prod_t\prod_v(\phi_v^t)^{n_v^t}\right]$$

$$\cdot\prod_d\left[\frac{\Gamma(T\alpha)}{\Gamma(\alpha)^T}\prod_t(\theta_t^d)^{\alpha-1}\right]\left[\prod_d\prod_t(\theta_t^d)^{m_t^d}\right]$$

$$=\prod_d\left[\frac{\Gamma(T\alpha)}{\Gamma(\alpha)^T}\prod_t(\theta_t^d)^{m_t^d+\alpha-1}\right]\cdot\prod_t\left[\frac{\Gamma(V\beta)}{\Gamma(\beta)^V}\prod_v(\phi_v^t)^{n_v^t+\beta-1}\right]$$

2. 边际分布

$$P(Z,W)=\prod_d\left[\frac{\Gamma(T\alpha)}{\Gamma(\alpha)^T}\cdot\frac{\prod_t\Gamma(m_t^d+\alpha)}{\Gamma(\sum_t m_t^d+T\alpha)}\right]\cdot$$

$$\prod_t\left[\frac{\Gamma(V\beta)}{\Gamma(\beta)^V}\cdot\frac{\prod_v\Gamma(n_v^t+\beta)}{\Gamma(\sum_v n_v^t+V\beta)}\right]$$

3. 推理 $P(Z|W)$ Gibbs 采样

$$P(z_i^d|Z\backslash z_i^d,W)\propto P(z_i^d,Z\backslash z_i^d,W),\quad 1\le z_i^d\le T$$

4. 完全条件概率

$$P(\underset{t_0}{\overset{\|}{z_i^d}}|Z\backslash z_i^d,W)\propto P(z_i^d,Z\backslash z_i^d,W)=P(Z,W)$$

$$=\frac{\hat m_{t_0}^d+\alpha}{\sum_t\hat m_t^d+T\alpha}\cdot\frac{\hat n_v^{t_0}+\beta}{\sum_v\hat n_v^{t_0}+V\beta}$$

$\cancel{\theta_t^d}$ 估计 θ_t^d: $$\hat\theta_t^d=\frac{m_t^d+\alpha}{\sum_t m_t^d+T\alpha},\quad 1\le t\le T,\ 1\le d\le D$$

估计 φ_v^t: $$\hat\varphi_v^t=\frac{n_v^t+\beta}{\sum_v n_v^t+V\beta},\quad 1\le v\le V,\ 1\le t\le T$$

① $X_1 \sim \Gamma(\alpha_1) \quad \cdots \quad X_k \sim \Gamma(\alpha_k)$

$Y = X_1 + \cdots + X_k \sim \Gamma(\alpha_1 + \cdots + \alpha_k)$ ✓

② $X_1 \sim \Gamma(\alpha) \quad X_2 \sim \Gamma(\beta)$

$$Y = \frac{X_1}{X_1 + X_2} \sim B(x \mid \alpha, \beta)$$ ✓

③ $$D(\alpha_1, \cdots, \alpha_{k+1}) = \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{k+1})}{\Gamma(\alpha_1 + \cdots + \alpha_{k+1})}$$ ✓

④ $X_1 \sim \Gamma(\alpha_1) \quad \cdots \quad X_k \sim \Gamma(\alpha_k)$

$$Y_i = \frac{X_i}{X_1 + \cdots + X_k} \quad 1 \le i \le k \sim \mathrm{Dir}$$ O

⑤ Margin

⑥ [illegible]

$$\frac{X_1}{X_1 + \cdots + X_k}, \quad \cdots, \quad \frac{X_{k-1}}{X_1 + \cdots + X_k}, \quad \frac{X_k}{X_1 + \cdots + X_k}$$

$$\cancel{P(y_1, \cdots, y_n) = P(x_1, \cdots, x_n) \left| \frac{\partial x}{\partial y} \right|}$$

$$= \cancel{P(g(y))} \mid$$

$$P(\vec{y}) = P(\vec{g}(\vec{y})) \left| \frac{\partial \vec{x}}{\partial \vec{y}} \right|$$