02.12.16 бммо лек XIII

$$p(X,T|\Theta), \quad p(X|\Theta)=\int p(X,T|\Theta)\,dT \to \max_\Theta$$

$$\ln p(X|\Theta) \geqslant \mathbb{E}_{q(T)} \ln p(X,T|\Theta) - \mathbb{E}_{q(T)} \ln q(T) = \{q(T)=q(T|\lambda)\} =$$

$$= \mathbb{E}_{q(T|\lambda)} \ln p(X,T|\Theta) - \mathbb{E}_{q(T|\lambda)} \ln p(T|\lambda) \to \max_{\lambda,\Theta}$$

$$\mathbb{E}_{q(T|\lambda)} \ln p(X,T|\Theta) \approx \frac{1}{L}\sum_{l=1}^{L} \ln p(X,T^l|\lambda), \quad T^l \sim q(T|\lambda)$$

$$\nabla_\Theta \mathbb{E}_{q(T|\lambda)} \ln p(X,T|\Theta) = \mathbb{E}_{q(T|\lambda)} \nabla_\Theta \ln p(X,T|\Theta)$$

$$T \sim q(T|\lambda); \quad z \sim q_0(z), \quad T=f(z,\lambda), \quad z^l \sim q_0(z)$$

$$\nabla_\lambda \mathbb{E}_{q(T|\lambda)} \ln p(X,T|\Theta) = \nabla_\lambda \mathbb{E}_{q_0(z)} \ln p(X, f(z,\lambda)|\Theta) =$$

$$= \mathbb{E}_{q_0(z)} \nabla_\lambda \ln p(X, f(z,\lambda)|\Theta) \approx \frac{1}{L}\sum_l \nabla_\lambda \ln p(X, f(z^l,\lambda)|\Theta)$$

$$p(X,T|\Theta) = p(X|T,\Theta)p(T|\Theta) = \left[\prod_{n=1}^{N} p(x_n|T,\Theta)\right] p(T|\Theta)$$

$$\frac{1}{N}\ln p(X|\Theta) \geqslant \mathbb{E}_{q(T|\lambda)}\left[\frac{1}{N}\sum_{n=1}^{N} \ln p(x_n|T,\Theta) + \frac{1}{N}\ln p(T|\Theta) - \frac{1}{N}\ln q(T|\lambda)\right]$$

doubly stochastic VI

$$i \sim \mathrm{Unif}(1\ldots N), \quad z \sim q_0(z), \quad \lambda \leftarrow \lambda - \alpha\left[\nabla_\lambda \ln p(x_n|f(z,\lambda),\Theta) + \right.$$

$$\left. + \nabla_\lambda \mathbb{E}_{q(T|\lambda)} \ln \frac{p(T|\Theta)}{p(T|\lambda)}\right]$$

$$\Theta \leftarrow \Theta - \alpha\left[\nabla_\Theta \ln p(x_n|f(z,\lambda),\Theta) + \nabla_\Theta \mathbb{E}_{q(T|\lambda)} \ln p(T|\Theta)\right]$$

RVM

$$\{x_n, y_n\}_{n=1}^{N}, \quad x_n \in \mathbb{R}^D, \quad y_n \in \{-1,1\}, \quad g(x)=\mathrm{sign}(w^T x)$$

$$p(y,w|X,\alpha) = \left[\prod_{n=1}^{N} \sigma(y_n w^T x_n)\right] \mathcal{N}(w|0,A^{-1})$$

$$A = \mathrm{diag}(\alpha_1 \ldots \alpha_D), \quad \sigma(u) = \frac{1}{1+e^{-u}}$$

$$\ln p(y|X,\alpha) = \ln \int \underbrace{\prod_n \sigma(y_n w^T x_n) N(w|0,A^{-1})}_{Q(w)=\exp(L(w))} dw \approx$$

приближение лапласа

$$\approx \ln \int \exp(L(w_{MP}) + \frac{1}{2}(w-w_{MP})^T \nabla^2 L(w_{MP})(w-w_{MP})) dw =$$

$$= L(w_{MP}) + \frac{D}{2}\ln 2\pi + \frac{1}{2}\ln\det((-\nabla^2 L(w_{MP}))^{-1}) \to \max_L$$

$$w_{MP} = \arg\max_w L(w)$$

$$\ln p(y|X,\alpha) \geq \mathbb{E}_{q(w)}[L(w) - \ln q(w)] = \{q(w) = N(w|\mu,\Sigma)\} =$$

$$= \mathbb{E}_{N(w|\mu,\Sigma)} L(w) - \mathbb{E}_{N(w|\mu,\Sigma)} \ln N(w|\mu,\Sigma) \approx$$

ряд тейлора

$$\approx \mathbb{E}_{N(w|\mu,\Sigma)}[L(w) + \nabla L(v) + \nabla L(v)^T(w-v) +$$

$$+ \frac{1}{2}(w-v)^T \nabla^2 L(v)(w-v)] + \frac{D}{2}\ln 2\pi + \frac{1}{2}\ln\det\Sigma + \frac{D}{2} =$$

$$= L(v) + \nabla L(v)^T(\mu - v) + \frac{1}{2} \operatorname{tr} \nabla^2 L(v)\, \mathbb{E}_{N(w|\mu,\Sigma)}(w-v)(w-v)^T +$$

$$+ \frac{D}{2}\ln 2\pi + \frac{1}{2}\ln\det\Sigma + \frac{D}{2} = L(v) + \nabla L(v)^T(\mu - v) +$$

$$+ \frac{1}{2}\operatorname{tr}\nabla^2 L(v)\Sigma + \frac{1}{2}(\mu-v)^T\nabla^2 L(v)(\mu - v) + \frac{D}{2}\ln 2\pi +$$

$$+ \frac{1}{2}\ln\det\Sigma + \frac{D}{2} \to \max_{\mu,\Sigma,\alpha}$$

$$\frac{\partial}{\partial\mu} = \nabla L(v) + \nabla^2 L(v)(\mu - v) = 0 \Rightarrow \mu = v - (\nabla^2 L(v))^{-1}\nabla L(v)$$

$$\frac{\partial}{\partial\Sigma} = \frac{1}{2}\nabla^2 L(v) + \frac{1}{2}\Sigma^{-1} = 0 \Rightarrow \Sigma = (-\nabla^2 L(v))^{-1}, \boxed{v = w_{MP}}$$

$$\nabla L(v) = 0, \mu = v, \ ELBO = L(w_{MP}) + \frac{1}{2}\ln\det((-\nabla^2 L(w_{MP}))^{-1}) +$$

$$+ \text{const} \to \max_\alpha$$

$$\boxed{v = \mu_{prev}}$$

$$\ln p(y|X,\alpha) \geqslant \mathbb{E}_{q(w)}\Big[\sum_{n=1}^{N} \ln\sigma(y_n w^T x_n) + \ln \mathcal{N}(w|0,A^{-1}) -$$

$$- \ln \mathcal{N}(w|\mu,\Sigma)\Big] = \sum_{n=1}^{N} \mathbb{E}_{\mathcal{N}(w|\mu,\Sigma)} \ln\sigma(y_n w^T x_n) -$$

$$- \frac{D}{2}\ln 2\pi + \frac{1}{2}\sum_{j=1}^{D} \ln\alpha_j - \frac{1}{2}\sum_{j=1}^{D}\alpha_j\, \mathbb{E}_{\mathcal{N}(w|\mu,\Sigma)} w_j^2 + \frac{D}{2}\ln 2\pi +$$

$$+ \frac{D}{2} + \frac{1}{2}\ln\det\Sigma \;\circleddash$$

$$x \sim \mathcal{N}(x|\mu,\Sigma)$$
$$y \sim \mathcal{N}(y|Ax,\Gamma)$$
$$p(y) = \mathcal{N}(y|A\mu, \Gamma + A\Sigma A^T)$$

$$/\!/ \; \mathbb{E}(ww^T)_{ii} = (\Sigma + \mu\mu^T)_{ii} = \Sigma_{ii} + \mu_i^2$$

$$\ln\sigma(u) = \ln\Big(\frac{1}{1+e^{-u}}\Big) = -\ln(1+e^{-u})$$

$$w \sim \mathcal{N}(w|\mu,\Sigma)$$

$$u_i = y_i w^T x_i \sim \mathcal{N}(u_i|m_i, s_i^2), \quad m_i = y_i\mu^T x_i, \quad s_i^2 = \cancel{y_i^2}\, x_i^T \Sigma x_i$$

$$\circleddash \sum_{n=1}^{N} \underbrace{\mathbb{E}_{\mathcal{N}(u_n|m_n,s_n^2)} \ln\sigma(u_n)}_{=\;\mathbb{E}_{\mathcal{N}(z|0,1)}\ln\sigma(s_n z + m_n)} + \frac{1}{2}\sum_{j=1}^{D}\ln\alpha_j - \frac{1}{2}\sum_{j=1}^{D}\alpha_j(\Sigma_{jj} + \mu_j^2) +$$

$$+ \frac{D}{2} + \frac{1}{2}\ln\det\Sigma$$

$$\frac{\partial}{\partial\alpha_j} = \frac{1}{2\alpha_j} - \frac{1}{2}(\Sigma_{jj} + \mu_j^2) = 0, \quad \alpha_j = \frac{1}{\Sigma_{jj} + \mu_j^2}$$

$$\ln p(y|X,\alpha) \geqslant \sum_{n=1}^{N} \mathbb{E}_{\mathcal{N}(z|0,1)} \ln\sigma(s_n z + m_n) -$$

$$- \frac{1}{2}\sum_{j=1}^{D}\ln(\Sigma_{jj} + \mu_j^2) - \cancel{\frac{D}{2}} + \cancel{\frac{D}{2}} + \ln\det\Sigma \to \max_{\mu,\Sigma}$$

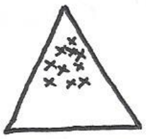
Если $D \gg 1$, то возьмём $q(w) = \prod_{j=1}^{D} \mathcal{N}(w_j|\mu_j,\Sigma_j)$

семейство полностью факторизованное распр-е

$$\mathrm{ELBO} = \sum_{n=1}^{N} \mathbb{E}_{\mathcal{N}(z|0,1)} \ln\sigma(s_n z + m_n) + \frac{1}{2}\sum_{j=1}^{D} \ln\frac{\Sigma_j}{\Sigma_j + \mu_j^2} \to \max_{\mu,\Sigma}$$

Correlated Topic Model

$$\Theta \sim Dir(\Theta|\alpha)\,;\quad \Theta_i = \frac{c_i}{\sum_j c_j}\,,\quad c_i \sim G(c_i|\alpha_i, 1)$$



$$\eta \sim N(\eta|\mu, \Sigma)$$

$$\Theta_i = \frac{\exp(\eta_i)}{\sum_j \exp(\eta_j)} \iff \Theta \sim Logit\text{-}N(\Theta|\mu, \Sigma)$$

$w_{dn} \in \{1 \dots W\}$, $d = 1 \dots D$, $n = 1 \dots N_d$, $z_{dn} \in \{1 \dots T\}$

Θ_{dt} вер-ть тематики t в док-те d, $\Theta_{dt} \geq 0$, $\sum_{t'} \Theta_{dt'} = 1$,

φ_{tw} вер-ть слова w в теме t, $\varphi_{tw} \geq 0$, $\sum_{w'} \varphi_{tw'} = 1$

Модель CTM

для $d = 1, 2, \dots, D$

$\eta_d \sim N(\eta|\mu, \Sigma)$, $\Theta_d = f(\eta_d)$

для $n = 1 \dots N_d$, $z_{dn} \sim Discrete(\Theta_d)$, $w_{dn} \sim Discrete(\varphi_{z_{dn}})$

$$p(W, Z, H|\mu, \Sigma, \Phi) = \prod_{d=1}^{D} N(\eta_d|\mu, \Sigma) \prod_{n=1}^{N} f_t(\eta_d)^{[z_{dn}=t]} \cdot \prod_{w=1}^{W} \varphi_{tw}^{[z_{dn}=t][w_{dn}=w]}\,,\quad p(W|\mu, \Sigma, \Phi) \to \max_{\mu, \Sigma, \Phi}$$

$$\ln p(W|\mu, \Sigma, \Phi) \geq \mathbb{E}_{q(Z,H)}[\ln p(W, Z, H) - \ln q(Z, H)] \ominus$$

$$q(Z, H) \approx q(Z)\, q(H)$$

факторизация $q(Z) = \prod_{d=1}^{D} \prod_{n=1}^{N_d} \prod_{t=1}^{T} \gamma_{dnt}^{[z_{dn}=t]}$

$$q(H) = \prod_{d=1}^{D} N(\eta_d|m_d, S_d)$$

$$\ominus \sum_{d=1}^{D} \Big[\mathbb{E}_{N(\eta_d|m_d, S_d)} \ln N(\eta_d|\mu, \Sigma) + \sum_{n=1}^{N_d} \sum_{t=1}^{T} \Big(\overbrace{\mathbb{E}_{q(Z)}[z_{dn}=t]}^{''\,\gamma_{dnt}} \cdot$$

$$\cdot\, \mathbb{E}_{N(\eta_d|m_d, S_d)} \ln f_t(\eta_d) + \cancel{\ln p(W|\mu, \Sigma, \Phi)} \cdot$$

$$+ \sum_{w=1}^{W} \mathbb{E}_{q(Z)}[z_{dn}=t] \cdot [w_{dn}=w] \ln \varphi_{tw} \Big) \Big] - \mathbb{E}_{q(Z,H)} \ln q(Z, H)$$

$$\mathbb{E}_{\mathcal{N}(\eta_d | m_d, S_d)} \ln f_t(\eta_d) = \mathbb{E}_{\mathcal{N}(\eta_d | m_d, S_d)} \left[\eta_{dt} - \ln \sum_{t=1}^{T} e^{\eta_{dt}} \right] =$$

$$= m_{dt} - \mathbb{E}_{\mathcal{N}(\eta_d | m_d, S_d)} \ln \sum_{t=1}^{T} e^{\eta_{dt}} =$$

$$= m_{dt} - \mathbb{E}_{\mathcal{N}(\zeta | 0, I)} \ln \sum_{t=1}^{T} \exp(C_d \zeta + m_d)_t$$

$$S_d = C_d C_d^T, \quad C_d = (\Delta^0)$$