$p(X|\theta) \xrightarrow[\theta]{} \max$ $\qquad p(X|\theta) = \int p(X,Z|\theta)\,dZ$

$$p(X,Z|\theta) = p(X|\theta)\,p(Z|X,\theta)$$

$$\ln p(X|\theta) = \ln p(X,Z|\theta) - \ln p(Z|X,\theta)$$

$$\ln p(X|\theta) = E_q\left[\ln p(X,Z|\theta) - \ln p(Z|X,\theta)\right] =$$

$$\pm \ln q(Z)$$

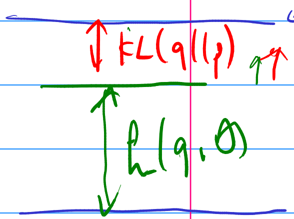
$$= E_q\left[\ln \frac{p(X,Z|\theta)}{q(Z)} - \ln \frac{p(Z|X,\theta)}{q(Z)}\right]$$

$$\underline{\ln p(X|\theta)} = \underbrace{\boxed{\int \ln \frac{p(X,Z|\theta)}{q(Z)}\, q(Z)\,dZ}} - \underbrace{\int \ln \frac{p(Z|X,\theta)}{q(Z)}\, q(Z)\,dZ}$$

ELBO

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

$\max_q \mathcal{L}$ $\qquad q \to \min$



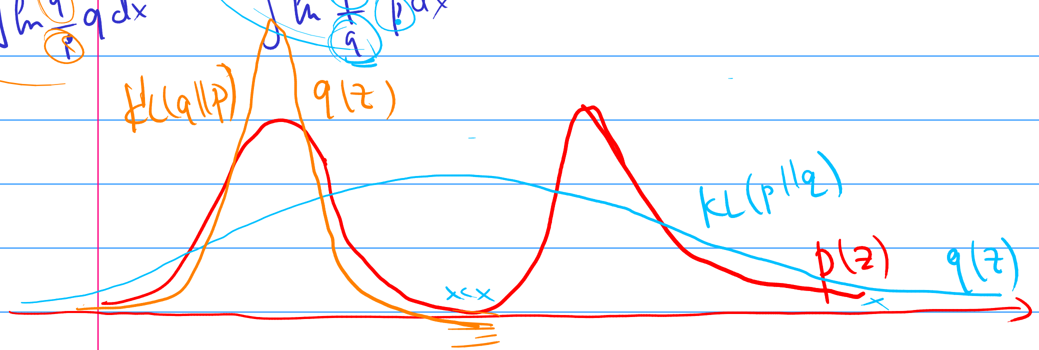
$\boxed{EM}$ $\theta^{(m)} \to \theta^{(m+1)}$ $\qquad q(Z) = p(Z|X,\theta^{(m)})$

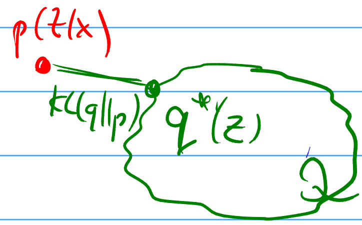
Var. approx.: $\ln p(X) = \mathcal{L}(q) + KL(q\|p) =$

$$= \underbrace{\int \ln \frac{p(X,Z)}{q(Z)}\, q(Z)\,dZ}_{\to \max} - \underbrace{\int \ln \frac{p(Z|X)}{q(Z)}\, q(Z)\,dZ}_{KL(q\|p)\ \to \min}$$

posterior

$KL(q\|p) = \int \ln \frac{q}{p}\, q\,dx$ $\qquad KL(p\|q) = \int \ln \frac{p}{q}\, p\,dx$





$$q(Z) = \prod_{i=1}^{M} q_i(Z_i), \qquad Z_1 \sqcup \dots \sqcup Z_M = Z, \quad Z_i \cap Z_j = \emptyset$$

$$\mathcal{L}(q) = \int \ln \frac{p(X,Z)}{q(Z)}\, q(Z)\,dZ = \int \ln \frac{p(X,Z)}{\prod_i q_i} \prod_i q_i\,dZ = \int \left(\int \prod_{i\neq j} q_i\,dZ_{-j}\right)$$

$$= \int \ln p(X,Z) \prod_i q_i(Z_i)\,dZ - \sum_{j=1}^{M} \int \ln q_j(Z_j) \prod_i q_i(Z_i)\,dZ$$

$$L(q) = \int \ln p(x,z) \prod_i q_i \, dz - \sum_{j=1}^{M} \int \ln q_j(z_j)\, q_j(z_j)\, dz_j$$

$$L(q) \xrightarrow[q_j]{} \max \qquad q_j^*(z_j) = ?$$

$$L(q_j) = \int q_j(z_j) \left[\ln p(x,z) \cdot \prod_{i\neq j} q_i(z_i)\, dz_{-j} \right] dz_j - \int \ln q_j \cdot q_j \, dz_j$$

$$= \int \underbrace{\left[\int \ln p(x,z) \cdot \prod_{i \neq j} q_i(z_i)\, dz_{-j} \right]}\, q_j(z_j)\, dz_j - \int q_j \ln q_j \, dz_j =$$

$$\ln \tilde p(x,z_j) = \int \ln p(x,z) \prod_{i\neq j} q_i \, dz_{-j} + const$$

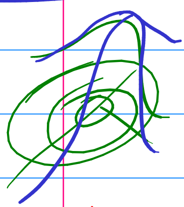
$$= \int \ln \frac{\tilde p(x,z_j)}{q_j(z_j)}\, q_j(z_j)\, dz_j \longrightarrow \max$$

$$= -KL(q_j \,\|\, \tilde p)$$

$$q_j^*(z_j) = \tilde p(x, z_j)$$

$$\boxed{\ln q_j^*(z_j) = \mathbb{E}_{q_{-j}(z_{-j})}\left[\ln p(x,z)\right] + const}$$

(1)



$$p(\bar z) = \mathcal{N}(\bar z \,|\, \bar\mu, \Sigma) = \frac{1}{(2\pi)\sqrt{\det \Sigma}} \cdot e^{-\frac12 (\bar z - \bar\mu)^T \Sigma^{-1} (\bar z - \bar\mu)}$$

$$\Sigma^{-1} = \Lambda = \begin{pmatrix}\lambda_{11} & \lambda_{12}\\ \lambda_{12} & \lambda_{22}\end{pmatrix}$$

$$p(\bar z) \approx q(\bar z) = q_1(z_1)\, q_2(z_2)$$

$$\ln q_1^*(z_1) = \mathbb{E}_{q_2^*(z_2)}\left[\ln p(\bar z)\right] + const$$

$$\ln q_2^*(z_2) = \mathbb{E}_{q_1^*(z_1)}\left[\ln p(\bar z)\right] + const$$

$$= \cancel{-\ln 2\pi + \frac12 \ln\det\Lambda} - \frac12 (\bar z - \bar\mu)^T \Lambda (\bar z - \bar\mu)$$

$$= -\frac12 \lambda_{11}(z_1-\mu_1)^2 - \lambda_{12}(z_1-\mu_1)(z_2-\mu_2) - \frac12 \lambda_{22}(z_2-\mu_2)^2$$

$$\ln q_1^*(z_1) = \mathbb{E}_{q_2^*(z_2)}\left[-\frac12\lambda_{11}(z_1-\mu_1)^2 - \lambda_{12} z_1 (z_2-\mu_2)\right] + const$$

$$= \mathbb{E}_{q_2^*(z_2)}\left[-\frac12\lambda_{11}z_1^2 + \lambda_{11}\mu_1 z_1 - \frac12\lambda_{11}\mu_1^2 - \lambda_{12}z_1(z_2-\mu_2)\right] + const$$

$$= -\frac12 \lambda_{11} z_1^2 + \lambda_{11}\mu_1 z_1 - \lambda_{12} z_1 \left(\mathbb{E}_{q_2}[z_2] - \mu_2\right) + const$$

$$-\frac{1}{2\sigma^2}(z-\mu)^2 \qquad \lambda_{11} z_1 \left(\mu_1 - \frac{\lambda_{12}}{\lambda_{11}}(-)\right)$$

$$q_1^*(z_1) = \mathcal{N}\Big(z_1 \,\Big|\, \underbrace{\mu_1 - \frac{\lambda_{12}}{\lambda_{11}}\big(\mathbb{E}_{q_2^*}[z_2] - \mu_2\big)}_{=m_1},\ \lambda_{11}^{-1}\Big) = \mathcal{N}(z_1 \mid m_1, \lambda_{11}^{-1})$$

$$q_2^*(z_2) = \mathcal{N}\Big(z_2 \,\Big|\, \underbrace{\mu_2 - \frac{\lambda_{12}}{\lambda_{22}}\big(\mathbb{E}_{q_1^*}[z_1] - \mu_1\big)}_{=m_2},\ \lambda_{22}^{-1}\Big) = \mathcal{N}(z_2 \mid m_2, \lambda_{22}^{-1})$$

$$\begin{cases} m_1 = \mu_1 - \dfrac{\lambda_{12}}{\lambda_{11}}(m_2 - \mu_2) \\[2mm] m_2 = \mu_2 - \dfrac{\lambda_{12}}{\lambda_{22}}(m_1 - \mu_1) \end{cases} \qquad m_1 = \mu_1,\ m_2 = \mu_2$$

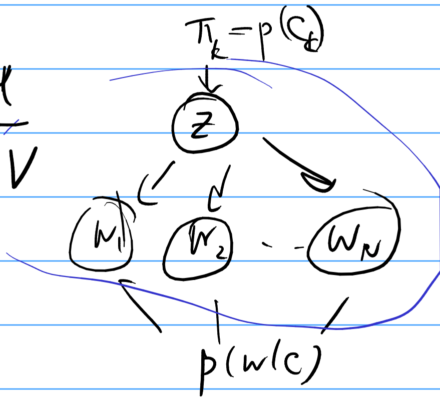
Naive Bayes $\quad D \quad d_1 d_2 \dots d_N \qquad d_1 = \{w_1 \dots w_{N_1}\} \qquad y_1, y_2, \dots$

Idiot's Bayes $\qquad\qquad y_1 = k \Leftrightarrow d_1 \in C_k$

Multinomial

$$p(C \mid d) \propto p(C)\,p(d \mid C) = p(C) \prod_{w \in d} p(w \mid C)$$

$$p(C_k) = \frac{\#\{d \in C_k\}}{\#\{d\}} \ ; \quad p(w \mid C) = \frac{\#\{w \in d \in C\} + 1}{\sum_{w'} \#\{w' \in d \in C\} + V}$$



Multivariate

$$p(C \mid d) \propto p(C) \cdot \prod_{w \in V} p(w \mid C)^{[w \in d]} \big(1 - p(w \mid C)\big)^{[w \notin d]}$$

Clustering NB

$$p(C \mid d) \propto \underbrace{p(C)}_{\pi_c} \prod_{w \in d} \underbrace{p(w \mid C)}_{\varphi_{wc}}$$

$\bar{z}_d \qquad z_{kd} = 1 \iff d \in C_k$

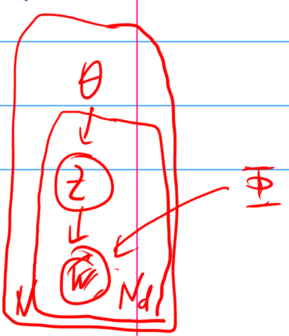
M-step

$$\pi_k = \frac{\sum_d z_{kd}}{N} \qquad \underbrace{\varphi_{wk}}_{p(w \mid C_k)} = \frac{\sum_d z_{kd} \cdot n_{dw} + 1}{\sum_d z_{kd} \cdot n_{d*} + V}$$

E-step: $\quad z_{kd} \propto \pi_k \cdot \prod_{w \in d} \varphi_{wk}$

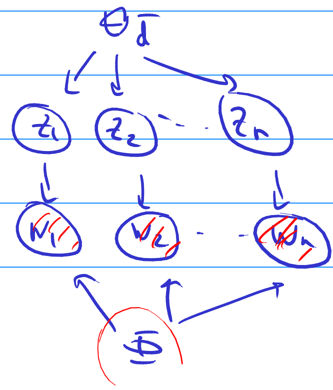
Topic Modeling $\qquad d \leftrightarrow \bar{\theta}_d = (\theta_{td}) \qquad \theta_{td} = p(t \mid d)$

pLSI LSA $\qquad \bar{\varphi}_t \qquad \varphi_{vt} = p(v \mid t)$



$$p(D \mid \Phi, \Theta) = \prod_{d \in D} \prod_{w \in d} p(w \mid \Phi, \Theta)$$

$$= \prod_{d \in D} \prod_{w \in d} \sum_{t=1}^{T} \theta_{td}\, \varphi_{wt} \ \xrightarrow[\Theta, \Phi]{} \max$$



EM: E-шаг $z_{dtw} = p(t \mid d, w) = \frac{p(w,t \mid d)}{p(w \mid d)} = \frac{\theta_{td}\varphi_{wt}}{\sum_s \theta_{sd}\varphi_{ws}}$

$z_{dtw} \propto \theta_{td}\varphi_{wt}$

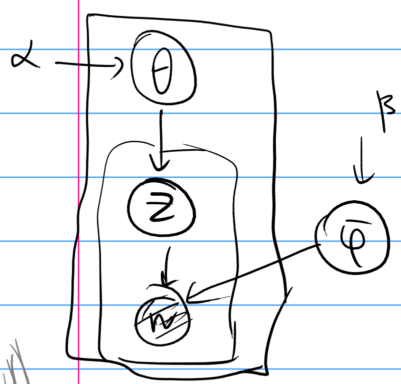
M-шаг: $\theta_{td} = \frac{E[\text{слов в } d \text{ из } t]}{n_d} = \frac{\sum_{w \in d} z_{dtw}}{n_d}$

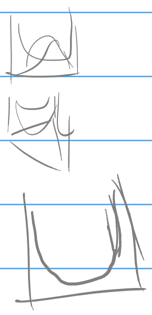
$$\varphi_{wt} = \frac{E[\text{слов } w \text{ из } t]}{E[\text{слов из } t]} = \frac{\sum_d z_{dtw}}{\sum_d \sum_w z_{dtw}}$$

ARTM

$$\ln p(W \mid \Theta, \Phi) - R(\Theta, \Phi) \longrightarrow \max$$

LDA Latent Dirichlet Allocation





$$p(W, \Theta, \Phi \mid \alpha, \beta) = \sum_z \underline{p(W, z, \Theta, \Phi \mid \alpha, \beta)}$$

$$p(W, z, \Theta, \Phi \mid \alpha, \beta) =$$

$$= \left(\prod_{t=1}^{T} \underline{p(\bar\varphi_t \mid \beta)}\right) \left[\prod_{j=1}^{M} \underline{p(\bar\theta_j \mid \alpha)} \cdot \prod_{n=1}^{N_j} \overbrace{p(z_{jn} \mid \bar\theta_j)}^{\text{Mult}}\, \underbrace{p(w_{jn} \mid z_{jn}, \Phi)}_{\text{Mult}}\right]$$

$$p(\bar\varphi_t \mid \bar\beta) = \mathrm{Dir}(\bar\varphi_t \mid \bar\beta) = \frac{1}{B(\bar\beta)} \prod_{v=1}^{V} \varphi_{tv}^{\beta_v - 1} = \frac{\Gamma(\sum \beta_v)}{\prod_v \Gamma(\beta_v)} \prod_v \varphi_{tv}^{\beta_v - 1}$$

$$p(\bar\theta_j \mid \bar\alpha) \propto \prod_{t=1}^{T} \theta_{jt}^{\alpha_t - 1}$$

$$p(z, \Theta, \Phi \mid W, \alpha, \beta) \approx q(z, \Theta, \Phi)$$

$$KL(q \,\|\, p) \to \min$$

$$L(q) = \sum_z \log \frac{p(W, z, \Theta, \Phi \mid \alpha, \beta)}{\underline{q(z, \Theta, \Phi)}}\, \underline{q(z, \Theta, \Phi)} \longrightarrow \max$$

$$q(z, \Theta, \Phi \mid \Lambda, \Gamma, \Pi) = \left(\prod_t \underbrace{q_t(\bar\varphi_t \mid \bar\lambda_t)}_{\mathrm{Dir}(\bar\varphi_t \mid \bar\lambda_t)}\right) \cdot \prod_{j=1}^{M} \left[\underbrace{q_j(\bar\theta_j \mid \bar\gamma_j)}_{\mathrm{Dir}(\bar\theta_j \mid \bar\gamma_j)} \cdot \prod_{n=1}^{N_j} \underbrace{q_j(z_{jn} \mid \bar\pi_j)}_{\mathrm{Mult}(z_{jn} \mid \bar\pi_j)}\right]$$

Упрощение: fix Φ $\quad q(z, \Theta \mid \Gamma, \Pi) \approx p(z, \Theta \mid W, \Phi, \alpha, \beta)$

$$KL(q_j \| p_j) = \int_\theta \sum_z q(\bar{z}_j, \bar{\theta}_j \mid \bar{\gamma}_j, \bar{\pi}_j) \cdot \log \frac{q(\ldots)}{p(\bar{z}_j, \bar{\theta}_j \mid \bar{w}_j, \Phi, \alpha, \beta)} d\theta_j$$

max Γ, γ, π

$$L(\bar{\gamma}_j, \bar{\pi}_j) = \int_\theta \sum_z \log \frac{p(\bar{w}_j, \bar{z}_j, \bar{\theta}_j \mid \Phi, \alpha, \beta)}{q(\bar{z}_j, \bar{\theta}_j \mid \bar{\gamma}_j, \bar{\pi}_j)} q(\ldots) d\theta_j =$$

$$= \mathbb{E}_q[\log p(\bar{w}_j, \bar{z}_j, \bar{\theta}_j \mid \Phi, \alpha, \beta)] - \mathbb{E}_q[q(\bar{z}_j, \bar{\theta}_j \mid \bar{\gamma}_j, \bar{\pi}_j)]$$

$$= \mathbb{E}_q[\log p(\bar{\theta}_j \mid \alpha)] + \mathbb{E}_q[\log p(\bar{z}_j \mid \bar{\theta}_j)] + \mathbb{E}_q[\log p(\bar{w}_j \mid \bar{z}_j, \Phi)]$$

$$- \mathbb{E}_q[\log q_j(\bar{\theta}_j \mid \bar{\gamma}_j)] - \mathbb{E}_q[\log q_j(\bar{z}_j \mid \bar{\pi}_j)] \to \max$$

$$\mathbb{E}_q[\log p(\bar{\theta}_j \mid \bar{\alpha})] = \mathbb{E}_{q(z,\theta)} \log\left(\frac{\Gamma(\sum \alpha_s)}{\prod \Gamma(\alpha_s)} \prod_t \theta_{jt}^{\alpha_t - 1} \right) =$$

$$= \log \Gamma\left(\sum_{s=1}^T \alpha_s\right) - \sum_{s=1}^T \log \Gamma(\alpha_s) + \sum_{t=1}^T (\alpha_t - 1) \mathbb{E}_q[\log \theta_{jt}] =$$

$$\mathbb{E}_{Dir(\bar{x} \mid \bar{\eta})}[\log x_i] = \psi(\eta_i) - \psi\left(\sum_j \eta_j\right), \quad \psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}$$

$$= \log \Gamma(\sum \alpha_s) - \sum \log \Gamma(\alpha_s) + \sum_t (\alpha_t - 1)\left(\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right)\right)$$

$$\mathbb{E}_q[\log p(\bar{w}_j \mid \bar{z}_j, \Phi)] = \mathbb{E}_q\left[\sum_{n=1}^{N_j} \sum_{t=1}^T \sum_{v=1}^V [z_{jn} = t] \cdot [w_{jn} = v] \cdot \log \varphi_{tv}\right]$$

$$= \sum_n \sum_t \sum_v [w_{jn} = v] \pi_{jnt} \cdot \log \varphi_{tv}$$

$$\begin{cases} \pi_{jnt} \propto e^{\log \varphi_{t, w_{jn}} + \psi(\gamma_{jt}) - \psi(\sum_s \gamma_{js})} \\ \gamma_{jt} = \alpha_t + \sum_{n=1}^{N_j} \pi_{jnt} \end{cases} \quad \Bigg| \quad \varphi_{tv} \propto \sum_{j=1}^M \sum_{n=1}^{N_j} [w_{jn} = v] \pi_{jnt}$$