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

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

$$\mathcal{L}(q) \to \max \qquad + \qquad KL(q\,\|\,p(Z|X)) \to \min$$

$$q(Z) = \prod_i q_i(Z_i), \qquad Z_i \cap Z_j = \emptyset$$

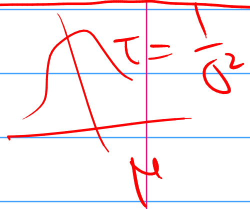
$$\ln q_j^*(Z_j) = \mathbb{E}_{q(Z_{-j})}\left[\ln p(X,Z)\right] + \text{const}$$

$$p(\bar{x}) = \mathcal{N}(\bar{x}\,|\,\mu, \Lambda^{-1}) \approx q(\bar{x}) = q_1(x_1)\,q_2(x_2)$$

$$\begin{pmatrix} \lambda_{11} & \cancel{\lambda_{12}} \\ \cancel{\lambda_{12}} & \lambda_{22} \end{pmatrix} \qquad m_1 = \mu_1,\; m_2 = \mu_2, \quad \tau_1 = \lambda_{11},\; \tau_2 = \lambda_{22}$$

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

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



$$D = \{x_1, \dots, x_N\}$$

$$\underbrace{p(\tau,\mu) \times p(D|\mu,\tau)}_{= \ln p(D,\mu,\tau)} \propto p(\tau,\mu|D)$$

$$p(\tau)\,p(\mu|\tau) = \mathrm{Gam}(\tau|a_0,b_0)\cdot \mathcal{N}(\mu|\mu_0,(\lambda_0\tau)^{-1})$$

$$= \frac{1}{\Gamma(\cdot)}\,\tau^{a_0-1} e^{-b_0\tau}\cdot\sqrt{\frac{\lambda_0\tau}{2\pi}}\; e^{-\frac{\lambda_0\tau}{2}(\mu-\mu_0)^2}$$

$$\prod_{n=1}^{N} p(x_n|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{N/2} e^{-\frac{\tau}{2}\sum_n (x_n-\mu)^2}$$

$$q(\mu,\tau) = \underbrace{q_\mu(\mu)}\,\underbrace{q_\tau(\tau)} \approx p(\tau,\mu|D)$$

$$\ln q_\mu^*(\mu) = \mathbb{E}_\tau\left[\ln p(\tau,\mu,D)\right] + \text{const}$$

$$\ln q_\tau^*(\tau) = \mathbb{E}_\mu\left[\ln p(\tau,\mu,D)\right] + \text{const}$$

$$\ln p(\tau,\mu,D) = (a_0-1)\ln\tau - b_0\tau + \frac{1}{2}\ln\tau - \frac{\lambda_0\tau}{2}(\mu-\mu_0)^2 + \frac{N}{2}\ln\tau - \frac{\tau}{2}\sum_n (x_n-\mu)^2 + \text{const}$$

$$\ln q_\mu^*(\mu) = \mathbb{E}_\tau\left[-\frac{\lambda_0\tau}{2}(\mu-\mu_0)^2 - \frac{\tau}{2}\sum_n (x_n-\mu)^2\right] + \text{const} =$$

$$= -\frac{\mathbb{E}_{q^*}[\tau]}{2}\left(\lambda_0(\mu-\mu_0)^2 + \sum_n (x_n-\mu)^2\right) + \text{const}$$

$$\lambda_0\mu^2 - 2\lambda_0\mu_0\mu + N\cdot\mu^2 - 2\left(\sum_n x_n\right)\mu$$

$$(\lambda_0+N)\left(\mu - \frac{\lambda_0\mu_0 + \sum_n x_n}{\lambda_0+N}\right)^2$$

$$q_\mu^*(\mu) = \mathcal{N}\left(\mu \,\middle|\, \frac{\lambda_0\mu_0 + \sum_n x_n}{\lambda_0+N},\ (\lambda_0+N)\cdot\mathbb{E}_{q^*}[\tau]\right)$$

$$\ln q_\tau^*(\tau) = (a_0-1)\ln\tau - b_0\tau + \frac{1}{2}\ln\tau + \frac{N}{2}\ln\tau - \frac{\tau}{2}\cdot\mathbb{E}_\mu\left[\lambda_0(\mu-\mu_0)^2 + \sum_n (x_n-\mu)^2\right] + \text{const}$$

$$\left(a_0 - 1 + \frac{N+1}{2}\right)\ln\tau - \left(b_0 + \mathbb{E}[\ldots]\right)\tau$$

$$q_\tau^*(\tau) = \text{Gam}\left(\tau \,\middle|\, a_0 + \frac{N+1}{2},\ b_0 + \frac{1}{2}\mathbb{E}_\mu[\ldots]\right)$$

$$\mathbb{E}[\tau] = \frac{a}{b} = \frac{a_0 + N+1/2}{b_0 + \frac{1}{2}\mathbb{E}[\ldots]}$$

$\mu_0 = 0 = a_0 = b_0 = \lambda_0$

non-informative priors

$$\mathbb{E}_\mu\left[\lambda_0(\mu-\mu_0)^2 + \sum_n (x_n-\mu)^2\right] =$$

$\mathbb{E}[x^2] = \mu^2 + \frac{1}{\tau}$

$x_n \sim \mathcal{N}(x|\mu,\tau)$

$$= \mathbb{E}_\mu\left[\mu^2(\lambda_0 + N) - 2\mu\left(\lambda_0\mu_0 + \sum_n x_n\right) + \lambda_0\mu_0^2 + \sum_n x_n^2\right]$$

$$= (\lambda_0+N)\,\mathbb{E}_\mu[\mu^2] - 2\left(\lambda_0\mu_0 + \sum_n x_n\right)\mathbb{E}[\mu] + \ldots =$$

$$= \cancel{(\lambda_0+N)}\left(\frac{(\lambda_0\mu_0 + \sum x_n)^2}{(\lambda_0+N)^2} + \frac{1}{\cancel{(\lambda_0+N)}\mathbb{E}\tau}\right) - \frac{2(\lambda_0\mu_0 + \sum_n x_n)^2}{\lambda_0+N} + \lambda_0\mu_0^2 + \sum_n x_n^2$$

$$= \frac{(\lambda_0\mu_0 + \sum x_n)^2}{\lambda_0+N} + \frac{1}{\mathbb{E}\tau} - \frac{2(\lambda_0\mu_0+\sum_n x_n)^2}{\lambda_0+N} + \lambda_0\mu_0^2 + \sum_n x_n^2$$

$$\mathbb{E}\tau = \frac{(N+1)/\cancel{2}}{\cancel{\frac{1}{2}}\left(\frac{(\sum x_n)^2}{N} + \frac{1}{\mathbb{E}\tau} - \frac{2}{N} + \sum_n x_n^2\right)}$$

$$\frac{1}{\mathbb{E}\tau} = \frac{1}{N+1}\cdot\frac{1}{\mathbb{E}\tau} + \frac{1}{N+1}\left(\frac{(\sum x_n)^2 - 2 + N\sum_n x_n^2}{N}\right)$$

$$\frac{1}{\mathbb{E}\tau}\frac{N}{\cancel{N+1}} = \frac{1}{N\cancel{(N+1)}}\left(\frac{\quad}{\quad}\right)$$

$$\frac{1}{\mathbb{E}\tau} = \frac{1}{N}\left(\cancel{\frac{1}{N}\left(\sum x_n\right)^2} + \sum_n x_n^2 - \frac{1}{N}\left(\sum_n x_n\right)^2\right)$$

σ^2

$$\sigma = \sqrt{\frac{1}{N}\sum_n (x_n - \bar{x})^2}$$

$$= \sum_n x_n^2 - 2\sum x_n \cdot \frac{\sum x_n}{N} + \left(\frac{\sum x_n}{N}\right)^2 \cdot N$$

$$= \sum_n x_n^2 - \frac{2}{N}\left(\sum x_n\right)^2 + \frac{1}{N}\left(\sum x_n\right)^2$$

$$= \sum_n x_n^2 - \frac{1}{N}\left(\sum x_n\right)^2$$

$\bar{x}_1, \dots, \bar{x}_n$

$$p(\bar{x}) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(\bar{x} \mid \bar{\mu}_k, \Lambda_k^{-1})$$

$\bar{x}_n \sim \bar{z}_n = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

$$p(\bar{x}_n \mid \bar{z}_n, \bar{\mu}, \Lambda) = \prod_k \mathcal{N}(\bar{x}_n \mid \bar{\mu}_k, \Lambda_k^{-1})^{z_{nk}}$$

$\theta = (\bar{\pi}, \mu_1 \dots \mu_k, \Lambda_1 \dots \Lambda_k)$

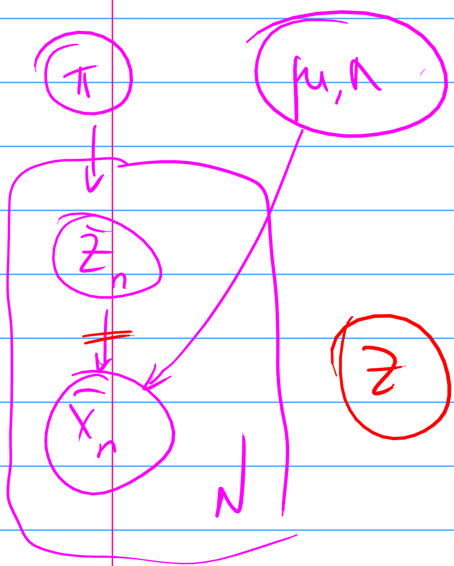
$$p(X, Z \mid \theta) = p(Z \mid \theta)\, p(X \mid Z, \theta) = \left(\prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} \right) \left(\prod_n \prod_k \mathcal{N}(\bar{x}_n \mid \bar{\mu}_k, \Lambda_k^{-1}) \right)^{z_{nk}}$$

$$p(X, Z, \bar{\pi}, \bar{\mu}, \Lambda) = p(Z \mid \pi)\, p(X \mid Z, \mu, \Lambda)\, p(\pi, \mu, \Lambda)$$

wishart

$$p(\bar{\pi}) = \mathrm{Dir}(\bar{\pi} \mid \bar{\alpha}_0) \propto \prod_k \pi_k^{\alpha_0 - 1}$$

$$p(\mu, \Lambda) = p(\mu \mid \Lambda)\, p(\Lambda) = \prod_{k=1}^{K} \mathcal{N}(\bar{\mu}_k \mid \bar{m}_0, (\beta_0 \Lambda_k)^{-1}) \cdot \mathcal{W}(\Lambda_k \mid W_0, \nu_0)$$



$$p(Z, \pi, \mu, \Lambda \mid X) \approx q(Z, \pi, \mu, \Lambda) = q(Z)\, q(\pi, \mu, \Lambda)$$

$$\ln q^*(Z) = \mathbb{E}_{\pi, \mu, \Lambda}\left[\ln p(X, Z, \pi, \mu, \Lambda) \right] + \text{const}$$

$$\ln q^*(\pi, \mu, \Lambda) = \mathbb{E}_Z\left[\ln p(X, Z, \pi, \mu, \Lambda) \right] + \text{const}$$

$$\ln p(X,Z,\pi,\mu,\Lambda) = \ln p(\pi) + \ln p(\mu,\Lambda) + \ln p(Z|\pi) + \ln p(X|Z,\mu,\Lambda)$$

$$\ln q^*(Z) = E_{\pi,\mu,\Lambda}\left[\ln p(Z|\pi) + \ln p(X|Z,\mu,\Lambda)\right] + const$$

$$= E_\pi\left[\sum_n\sum_k z_{nk}\ln\pi_k\right] + E_{\pi,\mu,\Lambda}\left[\sum_n\sum_k z_{nk}\left(\frac{1}{2}\ln\det\Lambda_k - \frac{D}{2}\ln 2\pi - \frac{1}{2}(x_n-\mu_k)^T\Lambda_k(x_n-\mu_k)\right)\right]$$

$$= \sum_n\sum_k z_{nk} E_\pi[\ln\pi_k] + \sum_n\sum_k z_{nk} E_{\pi,\mu,\Lambda}[\text{—}] + const$$

$$= \sum_n\sum_k z_{nk}\left(\text{———} \quad \text{—}\right) + const$$

$E z_{nk} = r_{nk}$ где

(1) $$\boxed{q^*(Z) \propto \prod_n\prod_k \rho_{nk}^{z_{nk}}},$$ где

$$\boxed{r_{nk} = \frac{\rho_{nk}}{\sum_{k'}\rho_{nk'}}}$$

$$\ln\rho_{nk} = E[\ln\pi_k] + \frac{1}{2}E[\ln\det\Lambda] - \frac{D}{2}\ln 2\pi - \frac{1}{2}E\left[(-)^T\Lambda_k(-)\right]$$

$$\ln q^*(\pi,\mu,\Lambda) = E_Z[\text{—}] + const_Z$$

$$= \ln p(\pi) + \underbrace{\ln p(\mu,\Lambda)}_{=\sum\ln p(\mu_k,\Lambda_k)} + \sum_n\sum_k(\ln\pi_k)E[z_{nk}] +$$

$$\sum_k(\alpha_0 - 1)\ln\pi_k \quad + \sum_n\sum_k[E z_{nk}]\cdot\ln p(x_n|\mu_k,\Lambda_k) =$$

$$= \left(\ln p(\pi) + \sum_{n,k}(E z_{nk})\ln\pi_k\right) + \sum_k\left[\left(\ln p(\mu_k,\Lambda_k) + \sum_n[E z_{nk}]\ln p(x_n|\mu_k,\Lambda_k)\right)\right] + const$$

(2) $$q^*(\pi,\mu,\Lambda) = q^*(\pi)\cdot\prod_{k=1}^K q^*(\mu_k,\Lambda_k)$$

$$\ln q^*(\pi) = \sum_k \left(\alpha_{0k} + \sum_n \mathbb{E} z_{nk} - 1\right) \ln \pi_k + const$$

$$q^*(\pi) \propto \prod_k \pi_k^{\alpha_{0k} + \sum_n \mathbb{E} z_{nk} - 1}$$

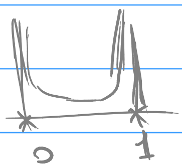
$$q^*(\pi) = Dir(\bar{\pi} \mid \bar{\alpha}), \quad \alpha_k = \alpha_{0k} + \sum_n \mathbb{E} z_{nk}$$

1. $q^{(0)} = p(\pi, \mu, \Lambda)$

2. Повторить:

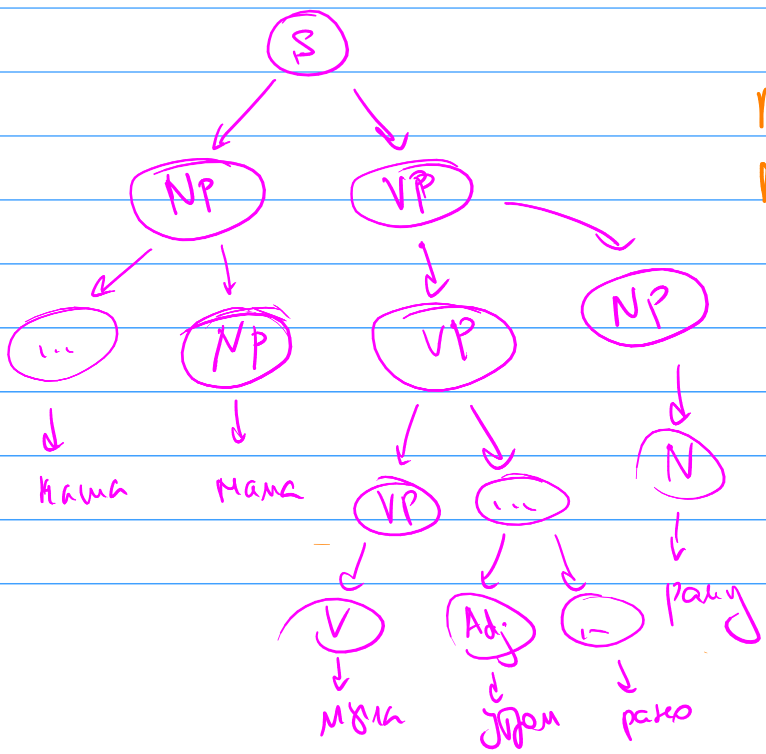
(E-шаг) $z_{nk}^{(i)} \leftarrow q^{(i-1)}(\pi, \mu, \Lambda)$ по ур. (1)

(M-шаг) $q^{(i)}(\pi, \mu, \Lambda) \leftarrow z_{nk}^{(i)}$ по ур. (2)



$p(\pi) = Dir(\bar{\pi} \mid \alpha_0)$ $\quad \bar{\alpha}_0 = \left(\frac{1}{100}, \dots, \frac{1}{100}\right)$

Chomsky кама мама мыла раму утром рано



S → NP VP

NP → Def NP

NP → NP Def

VP → VP NP

NP → N

VP → V

VP → VP Adj

Pron N V N Adj Adj

Наша мама мыла раму узлом рано

наш мама мыть рама узлом рано

Великий поэт А. С. Пушкин (р. 1799, с. Михайл)

O O B-Pers Pers Pers O B-Date B-Loc Loc

$x_1\ x_2\ x_3 \dots x_{t-1}\ x_t\ x_{t-1} \dots$

$(x_1, x_2 \dots \cancel{x_t} \dots x_n)$

$p(x_t \mid x_1 \dots x_{t-1}) \approx p(x_t \mid x_{t-1})$

n-граммы $p(x_t \mid x_{t-1}, x_{t-2}, \dots x_{t-n})$

(Obj, Pred, Subj) – knowledge bases

Freebase ЗВ

(X, is, poet)

(X, born_in, 1799)