① Var. approx.

$$p(X) = \frac{p(X, Z)}{p(Z|X)}$$

latent vars, model params

$p(\theta, Z | X)$

$$\log p(X) = \log p(X, Z) - \log p(Z|X) \quad \Big| \; \mathbb{E}_{q(Z)}$$

$$\log p(X) = \mathbb{E}_q\left[\log p(X, Z) - \log p(Z|X) + \log q(Z) - \log q(Z)\right]$$

$$\log p(X) = \int \log \frac{p(X, Z)}{q(Z)} q(Z)\, dZ + \int \log \frac{q(Z)}{p(Z|X)} q(Z)\, dZ$$

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

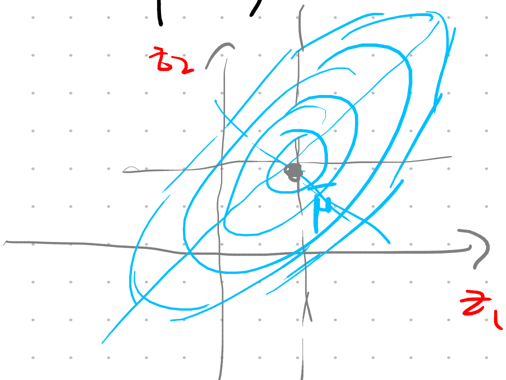
$\mathcal{L}(q) \to \max \iff KL \to \min$

$\mathcal{L}(q) \longrightarrow \max$

$$q(Z) = \prod_{i=1}^{N} q_i(Z_i), \quad Z_i \subseteq Z, \; Z_i \cap Z_j = \emptyset$$

$$\Rightarrow \log q_j^*(Z_j) = \mathbb{E}_{q_i^*(Z_i), i \neq j}\left[\log p(X, Z)\right] + \text{Const}$$

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



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

$$\begin{cases} \log q_1^*(z_1) = \mathbb{E}_{q_2^*(z_2)}\left[\log p(\bar{z})\right] + \text{Const} \\ \log q_2^*(z_2) = \mathbb{E}_{q_1^*(z_1)}\left[\log p(\bar{z})\right] + \text{Const} \end{cases}$$

$$\log q_1^*(z_1) = \mathbb{E}_{q_2^*(z_2)}\left[-\log 2\pi + \frac{1}{2}\log\det\Lambda - \frac{1}{2}(z-\mu)^T \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{pmatrix} (z-\mu) \right] + \text{Const}$$

$$= \mathbb{E}_{q_2^*(z_2)}\left[-\frac{1}{2}\left((z_1-\mu_1)^2 \lambda_{11} + 2\lambda_{12}(z_1-\mu_1)(z_2-\mu_2) + \lambda_{22}(z_2-\mu_2)^2 \right) \right] + \text{Const}$$

$$= -\frac{1}{2}\mathbb{E}_{q_2^*}\left[\lambda_{11} z_1^2 - 2\lambda_{11}\mu_1 z_1 + 2\lambda_{12}(z_2-\mu_2)\cdot z_1 \right] + \text{const}$$

$$= -\frac{1}{2}\left(\lambda_{11} z_1^2 - 2\lambda_{11}\mu_1 z_1 + 2\lambda_{12}\left(\mathbb{E}_{q_2^*}[z_2] - \mu_2\right) z_1 \right) + \text{Const}$$

$$= -\frac{1}{2}\cdot\lambda_{11}\left(z_1 - \left(\mu_1 - \frac{\lambda_{12}}{\lambda_{11}}\left(\mathbb{E}_{q_2^*}[z_2] - \mu_2\right)\right)\right)^2 + \text{Const}$$

$$q_1^*(z_1) = \mathcal{N}\left(z_1 \,\middle|\, \underbrace{\mu_1 - \frac{\lambda_{12}}{\lambda_{11}}\left(\overset{m_2}{\mathbb{E}_{q_2^*}[z_2]} - \mu_2\right)}_{m_1},\ \lambda_{11} \right)$$

$$\log q_2^*(z_2) = \mathbb{E}_{q_1^*(z_1)}\left[-\frac{1}{2}\left(\lambda_{22} z_2^2 - 2\lambda_{22}\mu_2 z_2 + 2\lambda_{12} z_2 (z_1-\mu_1) \right) \right] + \text{Const}$$

$$= -\frac{1}{2}\left(\lambda_{22} z_2^2 - 2\lambda_{22}\mu_2 z_2 + 2\lambda_{12} z_2 \left(\mathbb{E}_{q_1^*}[z_1] - \mu_1\right) \right) + \text{Const}$$

$$q_2^*(z_2) = \mathcal{N}\left(z_2 \,\middle|\, \underbrace{\mu_2 - \frac{\lambda_{12}}{\lambda_{22}}\left(\overset{m_1}{\mathbb{E}_{q_1^*}[z_1]} - \mu_1\right)}_{m_2},\ \lambda_{22} \right)$$

$$\begin{cases} m_1 = \mu_1 - \dfrac{\lambda_{12}}{\lambda_{11}}(m_2 - \mu_2) \\ m_2 = \mu_2 - \dfrac{\lambda_{12}}{\lambda_{22}}(m_1 - \mu_1) \end{cases} \Rightarrow \begin{cases} m_1 = \mu_1 \\ m_2 = \mu_2 \end{cases}$$

$$q(z) = \prod_i q_i(z_i)$$

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

$$= \int p(z) \log \frac{p(z)}{q(z)} dz = \text{Const} - \overbrace{\int p(z) \cdot \sum_i \log q_i(z_i)\, dz}^{\to \max}$$

$$q_j^*(z_j): \quad \int p(z) \log q_j(z_j)\, dz =$$

$$= \int \left(\int p(z)\, dz_{-j} \right) \cdot \log q_j(z_j)\, dz_j \longrightarrow \max$$

$$\boxed{q_j^*(z_j) = \int p(z)\, dz_{-j}}$$

② Обучение гауссиана

$$p(x \mid \mu, \tau) = \sqrt{\frac{\tau}{2\pi}}\, e^{-\frac{\tau}{2}(x-\mu)^2}$$

$$= \sqrt{\frac{\lambda_0 \tau}{2\pi}}\, e^{-\frac{\lambda_0 \tau}{2}(\mu-\mu_0)^2}$$

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

$$\overbrace{\mu, \tau}^{z} \mid \overbrace{D}^{x}$$

$$p(\mu, \tau \mid D) \propto \underbrace{p(\mu, \tau)}_{= Gam(\tau \mid \alpha_0, \beta_0) \cdot \mathcal{N}(\mu \mid \mu_0, \lambda_0 \tau)} \cdot \underbrace{p(D \mid \mu, \tau)}_{= \prod_n p(x_n \mid \mu, \tau)} = p(\overbrace{\mu, \tau}^{z}, \overbrace{x_1, \dots, x_N}^{X})$$

$$\approx q(\mu, \tau) = q_\mu(\mu)\, q_\tau(\tau)$$ (предположение)

$$\begin{cases} \log q_\mu^*(\mu) = E_{q_\tau^*}\left[\log p(\mu, \tau, x_1 \dots x_N)\right] + \text{Const} \\ \log q_\tau^*(\tau) = E_{q_\mu^*}\left[\log p(\mu, \tau, x_1 \dots x_N)\right] + \text{Const} \end{cases}$$

$$\log p(\tau) + \log p(\mu \mid \tau) + \sum_n \log p(x_n \mid \mu, \tau) =$$

$$= \text{Const} + (\alpha_0 - 1)\log \tau - \beta_0 \tau + \frac{1}{2}\log \tau - \frac{\lambda_0 \tau}{2}(\mu - \mu_0)^2 +$$

$$+ \sum_n \left(\frac{1}{2}\log\tau - \frac{\tau}{2}(x_n-\mu)^2 \right) \qquad \sum_n x_n^2 - 2\sum_n x_n\mu + N\cdot\mu^2$$

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

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

$$= -\frac{1}{2}\left(\mathbb{E}[\tau](\lambda_0+N)\cdot\mu^2 - 2\,\mathbb{E}[\tau]\cdot\left(\lambda_0\mu_0 + \sum_n x_n\right)\cdot\mu \right) + \text{Const}$$

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

$$\log q^*_\tau(\tau) = \mathbb{E}_{q^*_\mu}\left[(\alpha_0-1)\log\tau - \beta_0\tau + \frac{1}{2}\log\tau + \frac{N}{2}\log\tau - \frac{\tau}{2}\left(\lambda_0(\mu-\mu_0)^2 + \sum_n (x_n-\mu)^2\right)\right] + \text{const}$$

$$= \text{Const} + \left(\alpha_0 + \frac{N+1}{2} - 1\right)\log\tau - \tau\left(\beta_0 + \frac{1}{2}\mathbb{E}_{q^*_\mu}\left[\lambda_0(\mu-\mu_0)^2 + \sum_n (x_n-\mu)^2\right]\right)$$

$$q^*_\tau(\tau) = \text{Gam}\left(\tau \,\middle|\, \alpha_0 + \frac{N+1}{2},\ \beta_0 + \frac{1}{2}\mathbb{E}_{q^*_\mu}\left[\lambda_0(\mu-\mu_0)^2 + \sum_n (x_n-\mu)^2\right]\right)$$

Non-informative prior: $\alpha_0 = \beta_0 = \mu_0 = \lambda_0 = 0$

$$\begin{cases} q^*_\mu(\mu) = \mathcal{N}\left(\mu \,\middle|\, \frac{1}{N}\sum x_n,\ N\cdot\mathbb{E}[\tau]\right) \\ q^*_\tau(\tau) = \text{Gam}\left(\tau \,\middle|\, \frac{N+1}{2},\ \frac{1}{2}\mathbb{E}_{q^*_\mu}\left[\sum_n (x_n-\mu)^2\right]\right) \end{cases}$$

$1/\sigma^2$ "

$$\mathbb{E}[\tau] = \mathbb{E}_{\text{Gam}(\alpha,\beta)}[\tau] = \frac{\alpha}{\beta}$$

$$\mathbb{E}_\tau[\tau] = \frac{N+1}{\mathbb{E}_{q^*_\mu}\left[\sum_n (x_n-\mu)^2\right]}$$

$$\sum_n (x_n-\mu)^2 = \sum_n x_n^2 - 2\mu \cdot \sum_n x_n + N \cdot \mu^2$$

$$\mathbb{E}_{q^*_\mu}[\mu] = \frac{1}{N}\sum x_n$$

$$\mathbb{E}_{q^*_\mu}[\mu^2] = (\mathbb{E}[\mu])^2 + \mathrm{Var}[\mu] = \left(\frac{1}{N}\sum x_n\right)^2 + \frac{1}{N \cdot \mathbb{E}_\tau[\tau]}$$

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

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

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

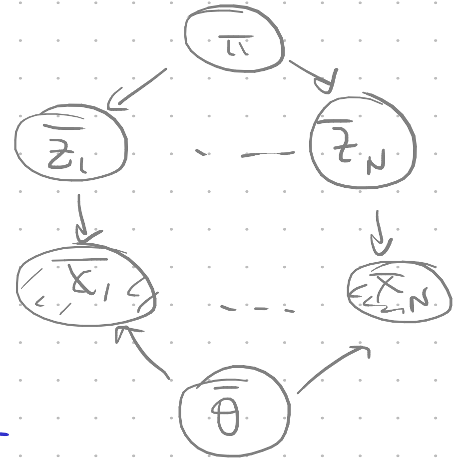
$$\mathbb{E}[\tau] = \frac{N}{\sum_n x_n^2 - \frac{1}{N}\left(\sum_n x_n\right)^2}$$

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

$$\text{"}\sigma^2\text{"} = \frac{1}{N^2}\left(\sum_n x_n^2 - \frac{1}{N}\left(\sum_n x_n\right)^2\right) = \frac{1}{N^2}\sum_n (x_n - \bar{x})^2$$

$$q^*_\mu(\mu) = \mathcal{N}\left(\mu \,\middle|\, \bar{x}, \frac{N^2}{\sum_n (\bar{x}-x_n)^2}\right), \quad q^*_\tau(\tau) = \mathrm{Gam}\left(\tau \,\middle|\, \frac{N+1}{2}, \frac{1}{2}\left(\bar{x}^2 + \frac{1}{N^2}\sum_n (x_n-\bar{x})^2\right)\right)$$

(3) Обучение смеси гауссианов



$X = \{\bar x_1 - \bar x_N\}$

$Z = \{(\bar z_1 - \bar z_N\}$ $\qquad \bar z_n = (\dots z_{nk} \dots)$, $\in \{0,1\}$

$$p(Z \mid \bar\pi) = \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}}$$

$$p(X \mid Z, \bar\mu, \bar\Lambda) = \prod_{n=1}^{N} \prod_{k=1}^{K} \mathcal{N}(\bar x_n \mid \bar\mu_k, \Lambda_k^{-1})^{z_{nk}}$$

$$p(x_n \mid \mu_k, \lambda_k) = \sqrt{\frac{\lambda_k}{2\pi}}\, e^{-\frac{\lambda_k}{2}(x_n - \mu_k)^2}$$

$\sum_Z p(X, Z \mid \bar\pi, \bar\mu, \bar\lambda)$ $=$

$$p(X, Z \mid \bar\pi, \bar\mu, \bar\lambda) \quad ; \quad p(\bar\pi, \bar\mu, \bar\lambda \mid X) \propto p(\bar\pi, \bar\mu, \bar\lambda) \cdot p(X \mid \bar\pi, \bar\mu, \bar\lambda)$$

$$p(\bar\pi, \bar\mu, \bar\lambda) = \mathrm{Dir}(\bar\pi \mid \alpha_0) \cdot \prod_{k=1}^{K} \mathrm{Gam}(\lambda_k \mid a_0, b_0) \cdot \mathcal{N}(\mu_k \mid m_{0k}, \beta_{0k}\lambda_k)$$

$\propto \pi_1^{\alpha_0 - 1} \dots \pi_K^{\alpha_0 - 1}$

предположение

$$p(Z, \bar\pi, \bar\mu, \bar\lambda \mid X) \approx q(Z, \bar\pi, \bar\mu, \bar\lambda) = q(Z) \cdot q(\bar\pi, \bar\mu, \bar\lambda)$$

$$\log q^*(Z) = \mathbb{E}_{\bar\pi, \bar\mu, \bar\lambda}\left[\log p(X, Z, \bar\pi, \bar\mu, \bar\lambda)\right] + \mathrm{Const} =$$

$$= \mathbb{E}_{\pi, \mu, \lambda}\Big[\log p(\bar\pi) + \sum_{k=1}^{K} \log p(\mu_k, \lambda_k) + \sum_{n=1}^{N} \log p(\bar z_n \mid \bar\pi)$$

$$+ \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \log p(x_n \mid \mu_k, \lambda_k)\Big] + \mathrm{Const}$$

$\sum_k z_{nk} \log \pi_k$ $\parallel$

$$= \mathbb{E}_\pi\left[\log p(\pi) + \sum_n \log p(z_n \mid \pi)\right] + \mathbb{E}_{\mu,\lambda}\left[\sum_k \log p(\mu_k, \lambda_k)\right]$$

$$+ \mathbb{E}_{\pi,\mu,\lambda}\left[\sum_n \sum_k z_{nk} \log p(x_n \mid \mu_k, \lambda_k)\right] + \mathrm{Const}$$

$$= \text{Const} + \sum_n \left[\sum_k z_{nk} \left(\mathbb{E}_\pi [\log \pi_k] + \mathbb{E}_{\mu,\lambda} [\log p(x_n | \mu_k, \lambda_k)] \right) \right]$$

($\tilde{z}_n$)

$$q^*(z) = \prod_{n=1}^N q^*(z_n) = \prod_{n=1}^N \left(\prod_{k=1}^K r_{nk}^{z_{nk}} \right), \text{ где}$$

$$r_{nk} \propto e^{\mathbb{E}_\pi [\log \pi_k] + \mathbb{E}_{\mu,\lambda} [\log p(x_n | \mu_k, \lambda_k)]}$$

$$\mathbb{E}_{q^*}[z_{nk}] = r_{nk}$$

$$\mathbb{E}_{\mu,\lambda}\left[\frac{1}{2}\log\lambda_k - \frac{1}{2}\log 2\pi - \frac{\lambda_k}{2}(x_n - \mu_k)^2\right]$$

$$\log q^*(\bar{\pi}, \bar{\mu}, \bar{\lambda}) = \mathbb{E}_{q^*_z} [\log p(X, Z, \bar{\pi}, \bar{\mu}, \bar{\lambda})] + \text{Const} =$$

$$= \log p(\bar{\pi}) + \log p(\bar{\mu}, \bar{\lambda}) + \mathbb{E}_z [\log p(Z | \bar{\pi})] + \mathbb{E}_z [\log p(X | Z, \bar{\mu}, \bar{\lambda})] + \text{Const}$$

$$q^*(\bar{\pi}, \bar{\mu}, \bar{\lambda}) = q^*(\bar{\pi}) \cdot q^*(\bar{\mu}, \bar{\lambda})$$

$$\log q^*(\bar{\pi}) = (\alpha_0 - 1) \sum_{k=1}^K \log \pi_k + \mathbb{E}_z \left[\sum_n \sum_k z_{nk} \log \pi_k \right] + \text{Const}$$

$$= \text{Const} + \sum_{k=1}^K \left(\alpha_0 + \sum_n \mathbb{E}_z[z_{nk}] - 1 \right) \log \pi_k =$$

$$= \text{Const} + \sum_{k=1}^K \left(\alpha_0 + \sum_n r_{nk} - 1 \right) \log \pi_k$$

$$q^*(\bar{\pi}) = \text{Dir}\left(\bar{\pi} \mid \ldots, \alpha_0 + \sum_n r_{nk}, \ldots\right)$$

$$\log q^*(\bar{\mu}, \bar{\lambda}) = \sum_k \log p(\mu_k, \lambda_k) + \mathbb{E}_z \left[\sum_n \sum_k z_{nk} \log p(x_n | \mu_k, \lambda_k) \right] + \text{Const}$$

$$q^*(\bar{\mu}, \bar{\lambda}) = \prod_{k=1}^K q^*(\mu_k, \lambda_k)$$

$$q^*(\mu_k, \lambda_k) = \text{Const} + (a_0 - 1)\log\lambda_k - b_0\lambda_k + \frac{1}{2}\log\lambda_k - \frac{\lambda_k \cdot \beta_0}{2}(\mu_k - m_{0k})^2$$

$$+ \sum_n \mathbb{E}[z_{nk}] \cdot \left(\frac{1}{2}\log\lambda_k - \frac{\lambda_k}{2}(x_n - \mu_k)^2\right) =$$

$$= \text{Const} + \left(a_0 - 1 + \frac{1}{2} + \frac{1}{2}\sum_n \mathbb{E}[z_{nk}]\right)\log\lambda_k - \lambda_k\left(b_0 + \frac{\beta_0}{2}(\mu_k - m_{0k})^2 + \frac{1}{2}\sum_n r_{nk}(x_n - \mu_k)^2\right)$$

$$q^*(\mu_k, \lambda_k) = \text{Gam}(\lambda_k \mid a, b) \cdot \mathcal{N}(\mu_k \mid m_k, \beta \cdot \lambda_k)$$

$$+ \frac{1}{2}\log\lambda_k$$

$$\frac{1}{2}\left(\beta_0\mu_k^2 - 2\beta_0 m_{0k}\mu_k + \beta_0 m_{0k}^2 + \left(\sum_n r_{nk}\right)\cdot\mu_k^2 - 2\mu_k \cdot \sum_n r_{nk}x_n + \sum_n r_{nk}x_n^2\right) =$$

$$= \frac{\beta_0 + \sum_n r_{nk}}{2}\left(\mu_k - \frac{\beta_0 m_{0k} + \sum_n r_{nk}x_n}{\beta_0 + \sum_n r_{nk}}\right)^2 +$$

$$+ \frac{1}{2}\left(\beta_0 m_{0k}^2 + \sum_n r_{nk}x_n^2 - \frac{(\beta_0 m_{0k} + \sum_n r_{nk}x_n)^2}{\beta_0 + \sum_n r_{nk}}\right)$$

$$m_k = \frac{\beta_0 m_{0k} + \sum_n r_{nk}x_n}{\beta_0 + \sum_n r_{nk}}, \quad \beta = \beta_0 + \sum_n r_{nk}$$

r_{nk}
$\downarrow$
$q^*(\pi)$
$q^*(\mu_k, \lambda_k)$

$$a = a_0 + \frac{1}{2}\sum_n r_{nk}, \quad b = b_0 + \frac{1}{2}\left(\beta_0 m_{0k}^2 + \sum_n r_{nk}x_n^2 - \beta \cdot m_k^2\right)$$

$$q^*(\pi, \mu, \lambda) \longrightarrow r_{nk}?$$

$$\mathbb{E}_{\lambda_k, \mu_k}\left[\frac{1}{2}\log\lambda_k - \frac{1}{2}\lambda_k(x_n - \mu_k)^2\right] =$$

$$\int p(\bar{x}|\bar{\eta}) = \int h(\bar{x})\, e^{\bar{\eta}^T t(\bar{x}) - a(\bar{\eta})}\, d\bar{x} = 1$$

$$\nabla_{\bar{\eta}} \ldots \;=\; \ldots \nabla_{\bar{\eta}} 1 = 0$$

$$\int h(\bar{x}) \left(t(\bar{x}) - \nabla_\eta a(\bar{\eta})\right) \cdot e^{\bar{\eta}^T t(\bar{x}) - a(\bar{\eta})}\, d\bar{x} = 0$$

($h(\bar{x})\, e^{\ldots} = p(\bar{x}|\bar{\eta})$)

"

$$E_{p(\bar{x}|\bar{\eta})}\left[t(\bar{x}) - \nabla_\eta a(\bar{\eta}) \right] = 0$$

$$\nabla_\eta a(\bar{\eta}) = E_{p(\bar{x}|\bar{\eta})}\left[t(\bar{x}) \right]$$

$$Dir(\bar{x}|\bar{\alpha}) = \frac{1}{B(\bar{\alpha})} \cdot x_1^{\alpha_1 - 1} \ldots x_k^{\alpha_k - 1} =$$

$$= e^{\sum_k (\alpha_k - 1)\cdot \log x_k - \log B(\bar{\alpha})}$$

$h(\bar{x}) = 1$

$\bar{\eta} = \bar{\alpha} - \bar{1}$

$t(\bar{x}) = \log \bar{x}$

$a(\bar{\eta}) = \log B(\bar{\alpha})$

$$E_{Dir(\bar{x}|\bar{\alpha})}\left[\log x_k\right] = \frac{\partial \log B(\bar{\alpha})}{\partial \alpha_k} =$$

$$B(\bar{\alpha}) = \frac{\Gamma(\alpha_1)\ldots\Gamma(\alpha_k)}{\Gamma(\alpha_1 + \ldots + \alpha_k)}$$

$$= \frac{d \log \Gamma(\alpha_k)}{d\alpha_k} - \frac{d \log \Gamma(\sum \alpha_\ell)}{d\alpha_k}$$

digamma function $\quad \psi(x) = \dfrac{d \log \Gamma(x)}{dx}$

$$\boxed{E_{Dir(\bar{x}|\bar{\alpha})}\left[\log x_k\right] = \psi(\alpha_k) - \psi\left(\sum_\ell \alpha_\ell\right)}$$

$$\mathbb{E}_{q^*(\bar\pi)}\left[\log \pi_k\right] = \psi\left(\alpha_0 + \sum_n r_{nk}\right) - \psi\left(\sum_l \left(\alpha_0 + \sum_n r_{nl}\right)\right)$$

$$\text{Gam}(\lambda \mid a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} = e^{\underbrace{\begin{pmatrix}\log\lambda \\ \lambda\end{pmatrix}^T}_{\bar t(\lambda)} \underbrace{\begin{pmatrix} a-1 \\ -b\end{pmatrix}}_{\bar\eta} - \underbrace{\left(-a\log b + \log\Gamma(a)\right)}_{a(\bar\eta)}}$$

$$\mathbb{E}_{\text{Gam}(\lambda \mid a, b)}\left[\log\lambda\right] = \frac{\partial a(\bar\eta)}{\partial \eta_1} = \psi(a) - \log b$$

$$\mathbb{E}_{q^*(\lambda_k, \mu_k)}\left[\frac{1}{2}\log\lambda_k - \frac{\lambda_k}{2}(x_n - \mu_k)^2\right] = \frac{1}{2}\left(\psi(a_k) - \log b_k\right) - \ldots$$

$$\mathbb{E}\left[\lambda_k (x_n - \mu_k)^2\right] = \mathbb{E}_{\lambda_k, \mu_k}\left[\lambda_k x_n^2 - 2\lambda_k x_n \mu_k + \lambda_k \mu_k^2\right] =$$

$$= \mathbb{E}_{\lambda_k}\left[\mathbb{E}_{\mu_k \mid \lambda_k}\left[\lambda_k x_n^2 - 2\lambda_k x_n \mu_k + \lambda_k \mu_k^2\right]\right] =$$

$$= \mathbb{E}_{\lambda_k}\left[\lambda_k x_n^2 - 2\lambda_k x_n m_k + \lambda_k\left(m_k^2 + \frac{1}{\beta_k \lambda_k}\right)\right] =$$

$$= \frac{1}{\beta_k} + \overbrace{\mathbb{E}_{q^*(\lambda_k)}\left[\lambda_k\right]}^{= a_k / b_k} \left(x_n^2 - 2x_n m_k + m_k^2\right) = \frac{1}{\beta_k} + \frac{a_k}{b_k}(x_n - m_k)^2$$

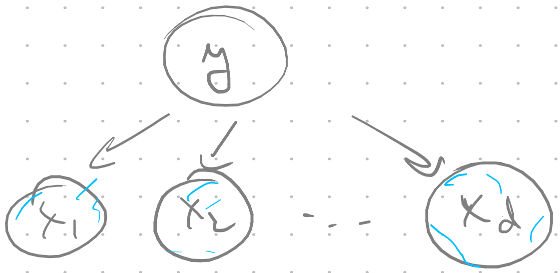
$$\log r_{nk} = \text{Const} + \Big[\psi\big(\overbrace{\alpha_0 + \sum_n r_{nk}}^{\alpha_k}\big) - \psi\big(\sum_l \overbrace{(\alpha_0 + \sum_n r_{nl})}^{\alpha_l}\big) +$$

$$+ \frac{1}{2}\left(\psi(a_k) - \log b_k\right) - \frac{1}{2}\left(\frac{1}{\beta_k} + \frac{a_k}{b_k}(x_n - m_k)^2\right)\Big]$$

(4) Naive Bayes

$D = \{(\bar{x}, y)\}$

$$p(\bar{x} \mid y) = \prod_{i=1}^{d} p(x_i \mid y) \quad \leftarrow \text{naive}$$



$$p(\bar{x}, y) = p(y) \cdot \prod_{i} p(x_i \mid y)$$

Nd / V

multinomial — bag-of-words

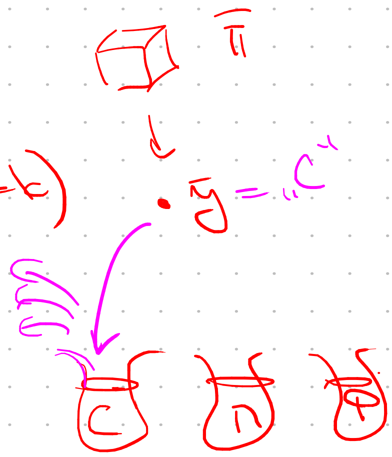
multivariate — [- - - -]

$C \in \{$Спорт, Политика, Финансы$\}$

$$D = \{(\bar{x}_n, y_n)\}_{n=1}^{N}$$

$\theta_k = p(y = k)$

$\theta_{wk} = p(w \mid y = k)$



$$p(D \mid \bar{\theta}) = \prod_{n} p(y_n) \prod_{i=1}^{L_n} p(x_{ni} \mid y_n)$$

$$\log p(D \mid \bar{\theta}) = \sum_{n} \sum_{k} y_{nk} \left(\log \theta_k + \sum_{i=1}^{L_n} \log \theta_{w_{ni}, k} \right) \xrightarrow[\bar{\theta}]{} \max$$

План

1) Clustering

$D = \{\bar{x}_n\}_{n=1}^{N}$, T – число тем

$$p(D \mid \bar{\theta}) = \prod_{n} \sum_{t} \left(p(y = t) \cdot \prod_{i=1}^{L_n} p(w_{ni} \mid t) \right)$$

$$Z = \{\bar{z}_n\}_{n=1}^{N}$$

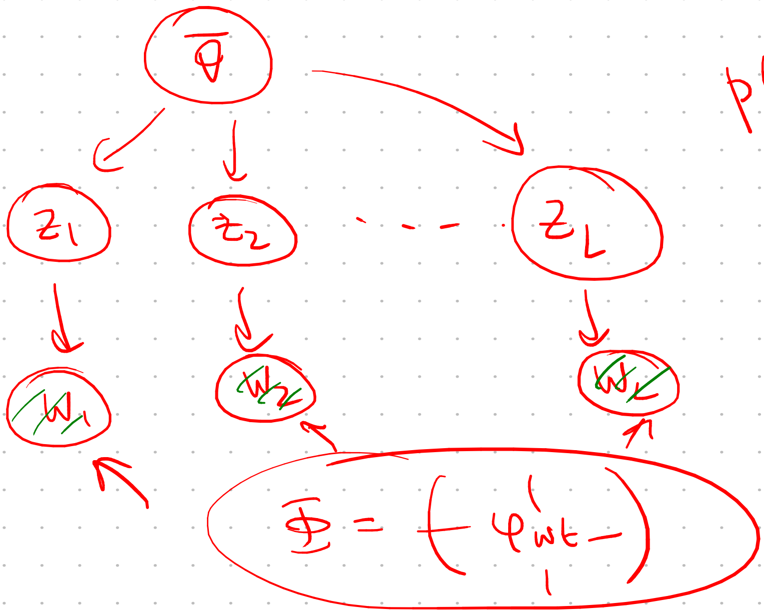
$$p(Z, D \mid \bar{\theta}) = \prod_{n} \prod_{t} \left(\theta_t \prod_{i=1}^{L_n} \theta_{w_{ni}, t} \right)^{z_{nt}}$$

$$Q(\bar{\theta}, \bar{\theta}^{(m)}) = \mathbb{E}_{Z \mid \bar{\theta}^{(m)}} \left[\log p(D, Z \mid \bar{\theta}) \right] =$$

$$= \mathbb{E}_{Z \mid \bar{\theta}^{(m)}} \left[\sum_{n} \sum_{t} z_{nt} \left(\log \theta_t + \sum_{i} \log \theta_{w_{ni}, t} \right) \right]$$

2) Topic modeling

document — $\bar{\theta} = (\theta_1 \dots \theta_T)$



$$p(d \mid \bar{\theta}, \Phi) =$$

$$= \prod_{i=1}^{L} p(w_i \mid \bar{\theta}, \Phi) =$$

$$= \prod_{i=1}^{L} \sum_{t=1}^{T} p(w_i, t \mid \bar{\theta}, \Phi)$$

$$= \prod_{i=1}^{L} \sum_{t=1}^{T} p(t \mid \bar{\theta})\, p(w_i \mid \varphi_{*t})$$