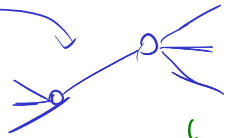
$\log p(w_t | w_{t-1}, \dots, w_1)$



$$p(w_t | w_{t-1}, w_{t-2}) \approx \frac{\#\{w_{t-2}\, w_{t-1}\, w_t\} + 1}{\#\{w_{t-2}\, w_{t-1}\} + |V|}$$

$p(X|\theta) \xrightarrow[\theta]{} \max$

$E_z[p(X,z|\theta)] \xrightarrow[\theta]{} \max$

$$p(x|D) = \int p(x|\theta)\, p(\theta|D)\, d\theta = E_{p(\theta|D)}\, p(x|\theta) \approx \frac{1}{R}\sum p(x|\theta^{(r)})$$

$\theta^{(r)} \sim p(\theta|D)$

---

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

$$\ln p(X,z|\theta) = \ln p(X|\theta) + \ln p(z|X,\theta)$$

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

$E_{q(z)}[\dots]$ $\quad q(z)$

$$\ln p(X|\theta) = E_{q(z)} \ln p(X,z|\theta) - E_{q(z)} \ln p(z|X,\theta)$$

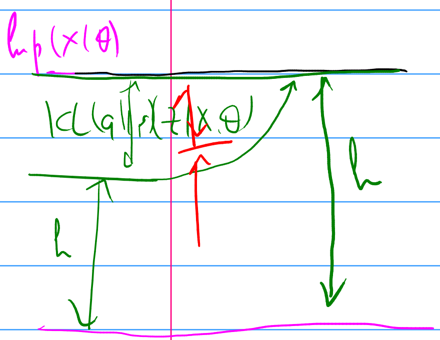
$$= \int \ln p(X,z|\theta)\, q(z)\, dz - \int \ln p(z|X,\theta)\, q(z)\, dz$$

$$\pm \int \ln q(z)\, q(z)\, dz$$

$KL(q\|p) = \int q \cdot \ln \frac{q}{p}\, dz$

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

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




$\mathcal{L}(q,\theta) \le \ln p(X|\theta)$

$\mathcal{L}(q,\theta) = \ln p(x,\theta)$, если $q(z) = p(z|x,\theta)$

$$\theta^{(m)} \xrightarrow{?} \theta^{(m+1)}$$

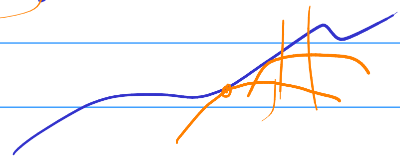
$q(z) := p(z|X,\theta^{(m)})$, т.е.

$$\ln p(x|\theta) = \int \ln \frac{p(X,z|\theta)}{p(z|X,\theta^{(m)})} \cdot p(z|X,\theta^{(m)})\, dz - \int \ln \frac{p(z|X,\theta)}{p(z|X,\theta^{(m)})}\, p(z|X,\theta^{(m)})\, dz$$

$$L(\theta, \theta^{(m)}) = \int \ln \frac{p(X, Z|\theta)}{p(Z|X, \theta^{(m)})} p(Z|X, \theta^{(m)})\, dZ$$

$$L(\theta, \theta^{(m)}) \le \ln p(X|\theta)$$

$$L(\theta^{(m)}, \theta^{(m)}) = \ln p(X|\theta^{(m)})$$



$$\arg\max_\theta L(\theta, \theta^{(m)}) = \arg\max_\theta \int \ln p(X, Z|\theta)\, p(Z|X, \theta^{(m)})\, dZ$$

---

$$\boxed{\ln p(X)} = L(q) + KL(q\|p) \qquad p(X, Z) \quad p(Z|X)$$

скр. перем. ↑ данные

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

$$\mathrm{Const} = L(q) + KL(q\| p(Z|X))$$

max ELBO min

$Z \sim \cancel{p}(Z|X)$

$q = \cancel{p}(Z|X)$

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

$KL(q\|p)$ $Q$ $q^*$ $p(Z|X)$

$$\int f(x)\, p(x) p(y)\, dx\, dy = \int f(x) p(x) dx$$

$$L(q) = \int \ln \frac{p(X,Z)}{q(Z)} q(Z)\, dZ =$$

$$= \int \ln \frac{p(X,Z)}{\prod q_i(Z_i)} \cdot \prod q_i(Z_i)\, dZ = \int \left[\ln p(X,Z) - \sum \ln q_i(Z_i)\right] \cdot \prod q_i(Z_i)\, dZ =$$

$$= \int \ln p(X,Z) \prod q_i(Z_i)\, dZ - \sum_j \int \ln q_j(Z_j) \prod q_i(Z_i)\, dZ$$

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

$$\int \ln q_j \prod_{i=1}^{M} q_i(Z_i)\, dZ = \left(\int \ln q_j\, q_j(Z_j)\, dZ_j\right) \cdot \left(\prod_{i \ne j} \int q_i(Z_i)\, dZ_i\right) = \int \ln q_j\, q_j\, dZ_j$$

$$L(q) \to \max_{q_j}$$

$dZ_j dZ_{-j}$

$$L(q) = \int q_j(Z_j) \cdot \ln p(X,Z) \cdot \prod_{i \ne j} q_i\, dZ - \int \ln q_j\, q_j\, dZ_j + \mathrm{Const}$$

$$= \int \left[\int \ln p(X,Z) \prod_{i\ne j} q_i(Z_i)\, dZ_{-j}\right] q_j(Z_j)\, dZ_j - \int \ln q_j\, q_j\, dZ_j + \mathrm{Const}$$

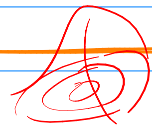
$$\ln \tilde{p}(X, Z_j) = \int \ln p(X,Z) \prod_{i\ne j} q_i\, dZ_{-j} + \mathrm{const}$$

$$\mathcal{L}(q) = \int \ln \tilde{p}(x, z_j)\, q_j(z_j)\, dz_j - \int \ln q_j(z_j)\, q_j(z_j)\, dz_j =$$

$$= \int \ln \frac{\tilde{p}}{q_j}\, q_j\, dz_j = -KL\left(q_j(z_j) \,\|\, \tilde{p}(x, z_j)\right) \to \max$$

$$q_j^*(z_j) = \hat{p}(x, z_j) \qquad q_{-j}(z_{-j}) = \prod_{i \neq j} q_i(z_i)$$

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



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

precision

$$p(\bar{z}) \propto q(\bar{z}) = q_1(z_1)\, q_2(z_2) \qquad \Sigma^{-1} = \Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}$$

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

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

$$\ln q_1^*(z_1) = \mathbb{E}_{q_2^*(z_2)}\left[\cancel{-\ln 2\pi + \frac{1}{2}\ln \det \Lambda} - \frac{1}{2}(\bar{z}-\bar{\mu})^T \Lambda (\bar{z}-\bar{\mu})\right] + \text{const}$$

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

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

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

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

$$\ln \mathcal{N} = \ldots - \frac{1}{2}\lambda(x-\mu)^2 = \ldots - \frac{\lambda}{2} x + \lambda x \mu$$

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

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

$$q_1^*(z_1) = \mathcal{N}(\mu_1, \lambda_{11}^{-1})$$

$$q_2^*(z_2) = \mathcal{N}(\mu_2, \lambda_{22}^{-1})$$

$$KL(p\|q) \to \min \qquad q(z) = \prod_i q_i(z_i) \qquad z_i \cap z_j = \varnothing$$

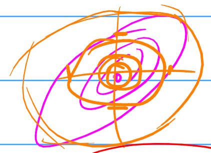
$$\int \ln \frac{p}{q}\, p\, dz = \int \ln \frac{p(z)}{\prod q_i(z_i)}\, p(z)\, dz = \cancel{\int \ln p(z)\, p(z)\, dz} - \underbrace{\sum_{i=1}^{M} \int \ln q_i(z_i)\, p(z)\, dz}_{\to\ \min_q}$$

$$-\int \ln q_i(z_i)\, p(z)\, dz \to \min$$

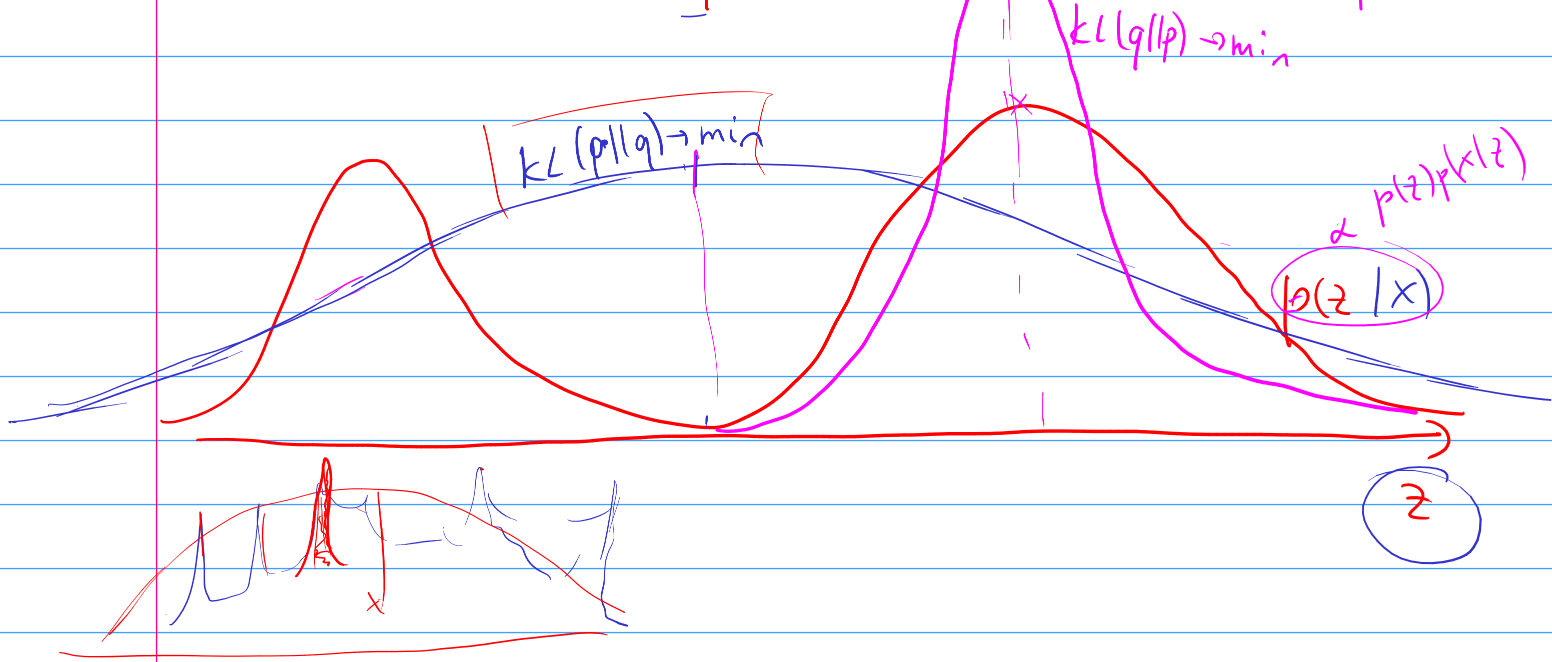
$$\int \ln \hat p\, \hat p\, dz_i \qquad -\int \ln q_i(z_i)\, \hat p(z_i)\, dz_i \xrightarrow[q_i]{} \min \qquad \hat p(z_i) = \int p(z)\, dz_{-i}$$

$$KL(\hat p\|q_i) = \int \ln \frac{\hat p}{q_i}\, \hat p\, dz_i \to \min \qquad q_i^*(z_i) = \int p(z)\, dz_{-i}$$

$$\begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{pmatrix} \longrightarrow \begin{pmatrix} \lambda_{11} & 0 \\ 0 & \lambda_{22} \end{pmatrix}$$



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




② $$p(D|\mu,\tau) = \prod_n N(x_n|\mu,\tau^{-1}) = \prod_{n=1}^{N} p(x_n|\mu,\tau) = \prod_{n=1}^{N} \sqrt{\frac{\tau}{2\pi}}\cdot e^{-\frac{\tau}{2}(x_n-\mu)^2}$$

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

$$\mathrm{Gamma}(\tau|a_0,b_0) \propto \tau^{a_0-1} e^{-b_0\tau}$$

$$p(\mu,\tau|D) \propto p(\mu,\tau)\cdot p(D|\mu,\tau) \approx q(\mu,\tau) = q_\mu(\mu)\cdot q_\tau(\tau)$$

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

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

$$+\frac{1}{2}\ln\tau$$

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

$$-\frac{\tau'}{2}\mu^2 - \tau'\mu'\cdot\mu$$

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

$$\lambda_0\mu^2 - 2\lambda_0\mu\mu_0 + N\mu^2 - 2\sum x_n \cdot \mu + \text{const}$$

$$(\lambda_0+N)\mu^2 - 2(\lambda_0\mu_0 + \sum x_n)\mu$$

$$\begin{cases} \tau_\mu' = \mathbb{E}[\tau](\lambda_0+N) \\ \mu_\mu' = \dfrac{\lambda_0\mu_0 + \sum x_n}{\lambda_0+N} \end{cases}$$

$$(a'-1)\ln\tau - b'\tau$$

$$\ln q_\tau^*(\tau) = \mathbb{E}_\mu[\ldots] + \text{const} = \mathbb{E}_\mu\Big[ \ln\tau\left(a_0 - 1 + \frac{1}{2} + \frac{N}{2}\right) -$$

$$-\tau\left(b_0 + \frac{\lambda_0}{2}(\mu-\mu_0)^2 + \frac{1}{2}\sum_n (x_n-\mu)^2\right)\Big] + \text{const}$$

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

$$\begin{cases} a_\tau' = a_0 + \dfrac{N}{2} + \dfrac{1}{2} \\ b_\tau' = b_0 + \dfrac{1}{2}\mathbb{E}_\mu\left[\sum_n (x_n-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right] \end{cases}$$

$$p(\tau|a,b) \propto \tau^{a-1} e^{-b\tau}$$

$$\mathbb{E}[\tau] = a/b$$

Non-informative priors $\quad \mu_0 = 0,\ \lambda_0 = 0 \quad \sigma_0 = \infty$

$$a_0 = 0,\ b_0 = 0$$

$$\bar{x} = \frac{1}{N}\sum x_n$$

$$\overline{x^2} = \frac{1}{N}\sum x_n^2$$

$$\tau_\mu' = N\cdot\mathbb{E}[\tau] \qquad a_\tau' = \frac{N+1}{2}$$

$$\mu_\mu' = \bar{x} = \frac{1}{N}\sum x_n \qquad b_\tau' = \frac{1}{2}\mathbb{E}_\mu\left[\sum_n (x_n-\mu)^2\right]$$

$$\mathbb{E}[\tau] = \frac{a'}{b'}$$

$$\bar{x} = \frac{1}{N}\sum x_n$$

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

$$= \overline{x^2} - 2(\bar{x})^2 + \frac{\mathbb{E}[\mu^2]}{N} =$$

$$\mathbb{E}[\mu^2] = (\bar{x})^2 + \frac{1}{N\cdot\mathbb{E}\tau}$$

$$= \overline{x^2} - 2(\bar{x})^2 + \frac{1}{N}(\bar{x})^2 + \frac{1}{N^2\,\mathbb{E}\tau}$$

$$(\mathbb{E}\tau)^{-1} = \frac{1}{N-1}\left(\overline{x^2} - (\bar{x})^2\right) = \frac{1}{N-1}\sum(x_n-\bar{x})^2$$