14.10.16 было лекция

VII

mean-field approximation

$$p(z|x) \approx q(z) = \underset{q \in Q}{\arg\min}\ KL(q(z) \| p(z|x))$$

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

вариационный байесовский вывод, variational Bayes

$$Q = \{ q(z) \mid q(z) = \prod_j q(z_j) \}, \quad z = \{ z_1 \dots z_m \}$$

непересекающиеся подмножества

задача ~~min~~ $KL(q(z) \| p(z|x)) \to \min_{p \in Q} \iff$

$$\mathcal{L}(q) = \int q(z) \ln \frac{p(x,z)}{q(z)} dz \to \max_{q \in Q}$$

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

$$\mathcal{L}(q) = \int q(z) \ln \frac{p(x,z)}{q(z)} dz = \int \prod_j q(z_j) \ln \frac{p(x,z)}{q(z)} dz =$$

$$= \int \prod_j q(z_j) \ln p(x,z) dz \int \prod_j q(z_j \sum_k \ln q(z_k) dz =$$

$$= \int \prod_j q(z_j) \ln p(x,z) dz \sum_k \int q(z_k) \ln q(z_k) dz_k =$$

$= \{$фиксируем все $q(z_k)$ кроме $q(z_l)\} \overset{const}{=}$

$$= \int q(z_l) \Big(\underbrace{\int \prod_j q(z_j) \ln p(x,z) dz_{\setminus l}}_{h(z_l)} \Big) dz_l - \int q(z_l) \ln q(z_l) dz_l$$

$$= \left\{ e^{h(z_l)} > 0, \quad \frac{e^{h(z_l)}}{\int e^{h(z_l)} dz_l} = \hat{p}(z_l), \quad h(z_l) = \ln \hat{p}(z_l) + const \right\} =$$

$$= \int q(z_l) \ln \hat{p}(z_l) dz_l + const - \int q(z_l) \ln q(z_l) dz_l \overset{const}{=}$$

$$= -\int q(z_l) \ln \frac{q(z_l)}{\hat{p}(z_l)} dz_l = -KL(q(z_l) \| \hat{p}(z_l)) \to \max_{q(z_l)}$$

$$\boxed{q(z_l) = \hat{p}(z_l)}$$

$$q(z_\ell)=\hat{p}(z_\ell)=\frac{e^{h(z_\ell)}}{\int e^{h(z_\ell)}\,dz_\ell}=\frac{e^{\mathbb{E}_{j\neq\ell}\ln p(x,z)}}{\int e^{\mathbb{E}_{j\neq\ell}\ln p(x,z)}\,dz_\ell}$$

$$\boxed{\ln q(z_\ell)=\mathbb{E}_{j\neq\ell}\ln p(x,z)+const}$$

условная сопряжённость

блочная сопряжённость

conditional conjugacy

$p(x,z_1,\dots,z_m)$ $\quad p(x,z_1,\dots,z_{k-1}|z_k,\dots,z_m)\sim p(z_k|z_{k+1},\dots,z_m)$

если верно, что $\cancel{p(x,z_1,\dots,z_k)\sim p(z_k|z_{k+1},\dots,z_m)}$

сопрягаются для $\forall k$, то выполнено св-во условной сопряжённости относительно $z_1\dots z_m$.

Т Пусть верна усл. сопр. $z_1\dots z_m$. Тогда формула $q(z)=\prod_{j=1}^{m}q(z_j)\approx p(z_1\dots z_m|x)$ может быть получена аналитически.

$$N(x|\mu,A^{-1})=\frac{\det^{\frac{1}{2}}A}{(2\pi)^{\frac{d}{2}}}\,e^{-\frac{1}{2}(x-\mu)^T A(x-\mu)}=$$

$$=\frac{\det^{\frac{1}{2}}A}{(2\pi)^{\frac{d}{2}}}\,e^{-\frac{1}{2}\operatorname{tr}A\underbrace{(x-\mu)(x-\mu)^T}_{d\times d}}$$

$$p(A)=\det^{\nu}A\;e^{\operatorname{tr}AW}\frac{1}{const(\nu,W)}$$ распределение Уишарта

$A=A^T\geq 0$ матричное обобщение экспоненциального распределения

$$p(\mu,A)=NW(\mu,A|m_0,\beta_0,W_0,\nu_0)=$$

$$=N(\mu|m_0,(\beta_0A)^{-1})\,W(A|W_0,\nu_0)$$

$X = (x_1 \dots x_n)$, $x_i \in \mathbb{R}^d$

$Z = (z_1 \dots z_n)$, $z_i \in \{0,1\}^k : \sum_{k=1}^{K} z_{ik} = 1$

$\pi = (\pi_1 \dots \pi_k)$, $\cancel{Z}$ $\sum_{k=1}^{K} \pi_k = 1$, $\pi_k \geq 0$

$\mu_k \in \mathbb{R}^d$, $A_k \in \mathbb{R}^{d \times d}$

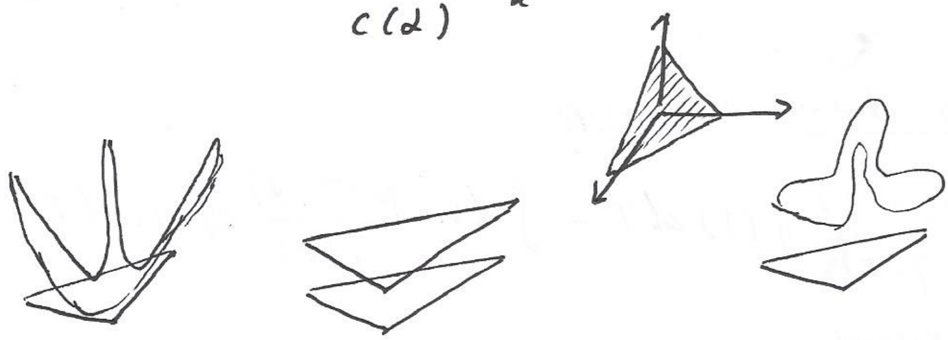
$$p(X, Z, \pi, \mu, A) = p(\cancel{Z} X \mid Z, \mu, A)\, p(Z \mid \pi) \cdot p(\pi) \cdot p(\mu, A) =$$

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

$$= \prod_n [p(x_n \mid z_n, \mu, \Lambda)\, p(z_n \mid \pi)]\, p(\pi) \prod_k [p(\mu_k \mid \Lambda_k)\, p(\Lambda_k)] =$$

$$= \prod_n \Big(\prod_k [N(x_n \mid \mu_k, \Lambda_k^{-1})]^{z_{nk}} \prod_k \pi_k^{z_{nk}} \Big) \cdot Dir(\pi \mid \alpha) \cdot$$
$$\cdot \prod_k NW(\mu_k, \Lambda_k \mid m_0, \beta_0, \nu_0, W_0)$$

$$Dir(\pi \mid \alpha) = \frac{1}{C(\alpha)} \prod_k \pi_k^{\alpha_k - 1} \; ; \; \pi_k \geq 0, \; \sum_k \pi_k = 1$$



$p(Z, \pi, \mu, \Lambda \mid X)$? $p(X \mid Z, \pi, \mu, \Lambda)$ не
$p(Z, \pi, \mu, \Lambda)$ сопряжены !

факторизация $\cancel{z}$

$$q(\pi, Z, \mu, \Lambda) = q(Z)\, q(\pi, \mu, \Lambda)$$

выполнена условная сопряжённость

$$\ln q(Z) \overset{const}{=} \mathbb{E}_{\pi, \mu, \Lambda} \ln p(X, Z, \pi, \mu, \Lambda) + const =$$

$$= \sum_n \sum_k \mathbb{E}_{\pi, \mu, \Lambda} z_{nk} (\ln N(x_n \mid \mu_k, \Lambda_k^{-1}) + \ln \pi_k) +$$

$$+ \sum_k \mathbb{E}_{\pi, \mu, \Lambda} \overset{(\alpha-1)}{\cancel{z_{nk}}} \ln \pi_k + \sum_k \mathbb{E}_{\pi, \mu, \Lambda} \ln NW(\mu_k, \Lambda_k \mid m_0, \beta_0, \nu_0, W_0) =$$

$$\overset{const}{=} \sum_n \sum_k z_{nk} (\mathbb{E}_{\mu_k, \Lambda_k} \ln N(x_n \mid \mu_k, \Lambda_k^{-1}) + \mathbb{E}_{\pi} \ln \pi_k)$$

$$q(Z) = \prod_n q(z_n) = \prod_n \prod_k (\dots)^{z_{nk}} / const =$$

$$= \prod_n \prod_k \left(\frac{e^{\mathbb{E}_{\mu_k,\Lambda_k} \ln N(x_n | \mu_k, \Lambda_k^{-1}) + \mathbb{E}_{\pi_k} \ln \pi_k}}{\sum_k e^{\mathbb{E}_{\mu_k,\Lambda_k} \ln N(x_n | \mu_k, \Lambda_k^{-1}) + \mathbb{E}_{\pi_k} \ln \pi_k}} \right)^{z_{nk}}$$

$$q(\mu, \Lambda, \pi) = Dir(\pi | \tilde{\alpha}) \prod_k NW(\mu_k, \Lambda_k | \dots)$$

достаточная статистика Dir:

$$\frac{1}{C(\alpha)} \prod_k \pi_k^{\alpha_k - 1} = \frac{1}{C(\alpha)} e^{\sum_k (\alpha_k - 1) \ln \pi_k}$$

$$\mathbb{E} \ln \pi_k = \frac{\partial \ln C(\alpha)}{\partial \alpha_k}$$

$$\ln q(\mu, \Lambda, \pi) = \mathbb{E}_Z \ln p(X, Z, \pi, \mu, \Lambda) + const$$

<u>14. 10. 16 было семинар</u>

$$\ln p(x) = \int \ln \frac{p(x,T)}{q(T)} q(T) dT - \int \ln \frac{p(T|x)}{q(T)} q(T) dT$$

распределение Уишерта

$$Wi(\Lambda | W, \nu) = B(W, \nu) |\Lambda|^{(\nu - d - 1)/2} e^{-\frac{1}{2} tr \Lambda W^{-1}}$$

модель: $p(X, \mu, \lambda) = \prod_n [p(x_n | \mu, \lambda)] \, p(\mu, \lambda)$

$$p(x_n | \mu, \lambda) = N(x_n | \mu, \lambda^{-1})$$

$$p(\mu, \lambda) = N(\mu | m_0, (\beta_0 \lambda)^{-1}) \, G(\lambda | a_0, b_0)$$

$$q(\mu) q(\lambda) \approx p(\mu, \lambda | X)$$

$$\ln q(\mu) = \mathbb{E}_{q(\lambda)} \left(\sum_n \ln p(x_n | \mu, \lambda) + \ln p(\mu, \lambda) \right) + const =$$

$$= \sum_n \mathbb{E}_{q(\lambda)} \left(\ln \sqrt{\lambda} - \frac{\lambda}{2} (x_n - \mu)^2 \right) + \mathbb{E}_{q(\lambda)} \ln \sqrt{\beta_0 \lambda} -$$

$$- \mathbb{E}_{q(\lambda)} \left(\frac{\beta_0 \lambda}{2} \right) (\mu - m_0)^2 + \mathbb{E}_{q(\lambda)} (a_0 - 1) \ln \lambda - \mathbb{E}_{q(\lambda)} (\beta_0 - 1) \lambda \overset{const}{=}$$

$\ominus$ $\sum_n \mathbb{E}_{q(\lambda)} \frac{\lambda}{2} (x_n-\mu)^2 - \mathbb{E}_{q(\lambda)}\left(\frac{\beta_0 \lambda}{2}\right)(\mu-m_0)^2 \overset{const}{=}$

const

$= \mathbb{E}_{q(\lambda)}\lambda \left(\sum_n \underline{(x_n-\mu)^2} - \beta_0 \underline{(\mu-m_0)^2}\right)$

$\underline{q(\mu) \sim N(\mu | a, b)}$

$\ln q(\lambda) = \mathbb{E}_{q(\mu)}\left(\sum_n \ln p(x_n|\mu,\lambda) + \ln p(\mu,\lambda)\right) + const \overset{const}{=}$

$= \mathbb{E}_{q(\mu)}\Big(\frac{1}{2} n \ln \lambda - \frac{\lambda}{2}\sum_n (x_n-\mu)^2 + \frac{1}{2}\ln \beta_0 \lambda - \frac{\beta_0}{2}\lambda(\mu-m_0)^2 +$

$+ (a_0-1)\ln\lambda - (\beta_0-1)\lambda\Big) = \frac{1}{2} n \ln \underset{\sim}{\lambda} + \frac{1}{2}\ln \beta_0 \underset{\sim}{\lambda} +$

$+ (a_0-1)\ln \underset{\sim}{\lambda} - (\beta_0-1)\underset{\sim}{\lambda} \cancel{+} - \frac{\underset{\sim}{\lambda}}{2}\mathbb{E}_{q(\mu)}\left(\cancel{\ast} \sum_n (x_n-\mu)^2 + \beta_0(\mu-m_0)^2\right)$

$\ln\lambda$ и λ $\Rightarrow$ $q(\lambda) \sim G(\lambda | c, d)$

1) инициализируем $q(\mu)$ a и b

2) $q(\lambda) \leftarrow \mathbb{E}_{q(\mu)}$

3) $q(\mu) \leftarrow \mathbb{E}_{q(\lambda)}$ $\qquad \ln p(x) \geq \int \ln \frac{p(x,T)}{q(T)} q(T)\, dT$

4) ...

$\mathbb{E}_{q(\mu,\lambda)} \ln p(x,\mu,\lambda) - \mathbb{E}_{q(\mu,\lambda)} \ln q(\mu,\lambda) = \mathbb{E}_{q(\mu,\lambda)}\Big(\ln p(\mu,\lambda) +$

$+ \sum_n \ln\sqrt{\frac{\lambda}{2\pi}} - \frac{\lambda}{2}(x_n-\mu)^2\Big) - \ldots = \frac{n}{2}\mathbb{E}_{q(\lambda)}\ln\frac{\lambda}{2\pi} - \frac{\lambda}{2}\sum_n \mathbb{E}_{q(\mu)}(x_n-\mu)^2 +$

$+ \frac{1}{2}\mathbb{E}_{q(\lambda)} \ln\frac{\beta_0\lambda}{2\pi} - \frac{\beta_0\lambda}{2}\mathbb{E}_{q(\mu)}(m_0-\mu)^2 + \ldots$

\# $p(X, Z, \pi, \{\mu_k, \lambda_k\}) = \prod_{n,k}\left(N(x_n|\mu_k, \Lambda_k)\pi_k\right)^{z_{nk}} p(\pi)\prod_k p(\mu_k,\lambda_k)$

$p(\pi|\alpha) = C\prod_k \pi_k^{\alpha-1}$

$p(\mu_k, \Lambda_k) = N(\mu_k | m_0, (\beta_0\Lambda_k)^{-1})\, W(\Lambda_k | W_0, V_0)$

$q(Z) = \prod_{n,k} r_{nk}^{z_{nk}}$, $p(\mu, \Lambda, \pi, Z|X) \approx q(\mu,\Lambda,\pi)\, q(Z)$

$$\ln q(\Pi,\mu,\lambda) = \mathbb{E}_{q(z)} \ln p(X,Z,\Pi,\mu,\lambda) + const \Leftrightarrow$$

$$\underset{Z}{q(T)} = q(T_1)\,\underset{\{\mu,\Lambda,\Pi\}}{q(T_2)}, \quad \ln q(T_i) = E_{i\ell} \ln p(X,T) + const$$

$$\Leftrightarrow \mathbb{E}_{q(z)} \Big(\sum_{k,n} z_{nk} (\ln N(x_n|\mu_k,\Lambda_k) + \underset{z_{nk}}{\cancel{Z}} \ln \Pi_k) + \sum_k (\alpha-1)\ln\Pi_k +$$

$$+ \sum_k \ln NW(\mu_k,\lambda_k|m_0,\beta_0,\nu_0,w_0)\Big) = \Big\{ q(Z) = \prod_{n,k} r_{nk}^{z_{nk}} \Big\} =$$

$$\mathbb{E}\, z_{nk} = r_{nk}$$

$$= \sum_{n,n} r_{nk} \ln N(x_n|\mu_k,\Lambda_k) + \overset{r_{nk}}{\cancel{N}} \sum_k \ln\Pi_k + \sum_k (\alpha-1)\ln\Pi_k +$$

$$+ \sum_k \ln NW(\mu_k,\lambda_k|m_0,\beta_0,\nu_0,w_0)$$

$$q(\Pi,\mu,\lambda) = q(\Pi)\,q(\mu,\lambda); \quad \sum_{n,k} r_{nk}\ln\Pi_k + \sum_k(\alpha-1)\ln\Pi_k$$

$$\cancel{(N+\alpha-1)\sum_k \ln\Pi_k, \quad q(\Pi) = Dir(\Pi|N+\alpha-1)}$$

$$\sum_{k,n} r_{nk} \ln N(x_n|\mu_k,\Lambda_k) + \sum_k \ln NW(\mu_k,\lambda_k|m_0,\beta_0,\nu_0,w_0)$$

$$\sum_k \ln\Pi_k \cdot \sum_n r_{nk} + \sum_k \ln\Pi_k(\alpha-1) = \sum_k \ln\Pi_k\Big(\sum_n r_{nk} + \alpha - 1\Big)$$

$$q(\Pi) = Dir\Big(\Pi \,\Big|\, \alpha \cancel{-1} + \sum_n r_{nk}\Big)$$

$$\sum_{k,n} r_{nk}, \quad q(\Pi,\mu,\lambda) = q(\Pi)\,q(\mu,\lambda) = q(\Pi)\sum_k q(\mu_k,\lambda_k)$$