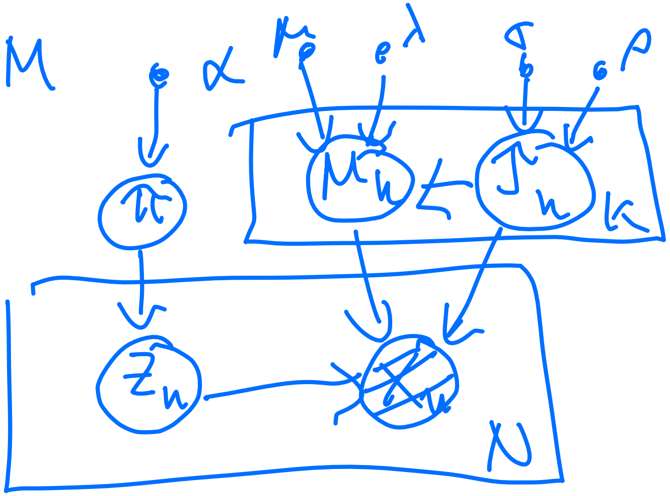
Royal Institute of Technology

VI FOR GMM

GM



R.V.

$X = \{X_n : n \in [N]\}$

$Z = \{Z_n : n \in [N]\}$

$M = \{\mu_k : k \in [K]\}$

$T = \{\tau_k : k \in [K]\}$

$$\pi \mid \alpha \sim \mathrm{Dir}(\alpha) = \prod_k \pi_k^{\alpha_k - 1}$$

$$Z_n \mid \pi \sim \mathrm{Cat}(\pi)$$

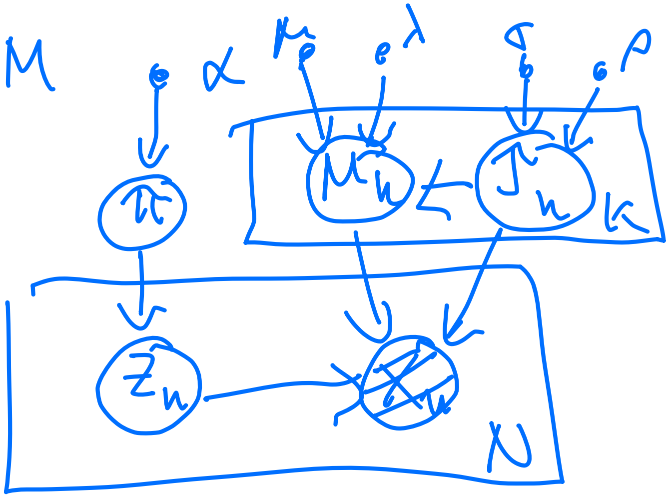
$$\tau_k \mid \sigma, \rho \sim \mathrm{Ga}(\sigma, \rho) = \frac{\rho^{\sigma}}{\Gamma(\sigma)} \tau_k^{\sigma - 1} e^{-\rho \tau_k}$$

$$\mu_k \mid \tau_k, \mu, \lambda \sim N(\mu, (\lambda \tau_k)^{-1}) = \sqrt{\frac{\lambda \tau_k}{2\pi}}\, e^{-\frac{\lambda \tau_k}{2}(\mu_k - \mu)^2}$$

$$X_n \mid Z_n = k, \mu_k, \tau_k \sim N(\mu_k, (\tau_k)^{-1}) = \sqrt{\frac{\tau_k}{2\pi}}\, e^{-\frac{\tau_k}{2}(X_n - \mu_{z_n})^2}$$

Noticed that i forgot half the precision, i.e., $\tau_k/2$

GM



R.V.

$X = \{x_n : n \in [N]\}$

$Z = \{z_n : n \in [N]\}$

$M = \{\mu_k : k \in [K]\}$

$T = \{T_k : k \in [K]\}$

Joint $p(X, Z, \pi, M, T) = p(X, Z \mid \pi, M, T)\, p(\pi)\, p(M, T)$

VI ass $q(Z, \pi, M, T) = q(Z)\, q(\pi, M, T)$

Notation $z_n \leftrightarrow z_{n1}, \dots, z_{nK}$; $z_n = k \Leftrightarrow z_{nk} = 1,\ z_{nk'} = 0\ \forall k' \neq k$

Complete lik

$$p(X, Z \mid \pi, M, T) = \prod_n p(x_n \mid z_n, M, T)\, p(z_n \mid \pi)$$

$$= \prod_{n,k} \left[p(x_n \mid \mu_k, T_k)\, p(z_{nk} = 1 \mid \pi) \right]^{z_{nk}}$$

$$= \prod_{n,k} \left[\mathcal{N}(x_n \mid \mu_k, T_k^{-1})\, \pi_k \right]^{z_{nk}}$$

Joint $p(X, Z, \pi, M, T) = p(X, Z | \pi, M, T)\, p(\pi)\, p(M, T)$

$$p(X, Z | \pi, M, T) = \prod_{n,k} \left[\mathcal{N}(x_n | \mu_k, T_k^{-1})\, \pi_k \right]^{z_{nk}}$$

$$\log q^*(Z) \stackrel{+}{=} \mathbb{E}_{\pi, M, T}\left[\log p(X, Z | \pi, M, T) \right]$$

$$= \sum_{n,k} z_{n,k} \underbrace{\left(\mathbb{E}_{\mu_k, T_k}\left[\log \mathcal{N}(x_n | \mu_k, T_k^{-1}) \right] + \mathbb{E}_{\pi}\left[\log \pi_k \right] \right)}_{\log u_{n,k}}$$

$$= \sum_{n,k} z_{n,k} \log u_{n,k} \quad \therefore \quad q^*(Z) \propto \prod_{n,k} u_{n,k}^{z_{n,k}} \propto \prod_{n,k} \left(\frac{u_{n,k}}{\sum_k u_{n,k}} \right)^{z_{n,k}}$$

Joint $p(X, Z, \pi, M, T) = p(X, Z \mid \pi, M, T)\, p(\pi)\, p(M, T)$

$$p(X, Z \mid \pi, M, T) = \prod_{n,k} \left[\mathcal{N}(x_n \mid \mu_k, T_k^{-1})\, \pi_k \right]^{z_{nk}}$$

$$\log q^*(\pi, M, T) = E_Z[\log p(X, Z \mid \pi, M, T)] + \log p(\pi) + \log p(M, T)$$

$$= \sum_{n,k} \underbrace{E_Z[z_{n,k}]}_{r_{n,k}} \underline{\log \mathcal{N}(x_n \mid \mu_k, T_k^{-1})} + \sum_{n,k} E_Z[z_{n,k}] \log \pi_k$$

$$+ \sum_k (\alpha_k - 1) \log \pi_k + \sum_k \underline{\log p(\mu_k, T_k)}$$

Note: (1) no term with π and (μ_k or T_k)

(2) — " — k and k' for $k \neq k'$

$\therefore q^*(\pi, M, T)$ factorized as $q(\pi) \prod_k q(\mu_k, T_k)$

$$\log GN(\mu, \tau \mid m, \ell, s, r)$$

$$= \log\left(\cancel{\frac{r^s}{\Gamma(s)}} \, \tau^{s-1} e^{-r\tau} \sqrt{\frac{\tau \ell}{\cancel{2\pi}}} \, e^{-\frac{\tau \ell}{2}(\mu - m)^2} \right)$$

$$\stackrel{+}{=} (s-1)\log \tau - r\tau + \frac{1}{2}\log \tau + \cancel{\frac{1}{2}\log \ell} - \frac{\tau \ell}{2}(\mu^2 - 2\mu m + m^2)$$

$$\stackrel{+}{=} (s-1)\log \tau - r\tau + \frac{1}{2}\log \tau - \frac{\tau \ell}{2}\mu^2 + \frac{\tau \ell}{2} 2\mu m - \frac{\tau \ell}{2} m^2$$

$\log q^*(\mu_k, \tau_k)$

$$\stackrel{+}{=} \sum_n r_{n,k}\left(\frac{1}{2}\log \tau_k - \frac{\tau_k}{2}(x_n - \mu_k)^2\right) + (\sigma - 1)\log \tau_k - \rho \tau_k$$

$$+ \frac{1}{2}\log \tau_k - \frac{\tau_k \lambda}{2}(\mu_k - \mu)^2$$

$$\stackrel{+}{=} \overbrace{\left(\sigma - 1 + \frac{1}{2}\sum_n r_{n,k}\right)}^{\sigma^* - 1}\log \tau_k + \frac{1}{2}\log \tau_k - \overbrace{\left(\rho + \frac{\lambda\mu^2}{2} + \frac{1}{2}\sum_n r_{n,k} x_n^2\right)}^{a}\tau_k$$

$$- \frac{\tau_k}{2}\overbrace{\left(\lambda + \sum_n r_{n,k}\right)}^{\lambda^*}\mu_k^2 + \frac{\tau_k}{2}\overbrace{\left(2\lambda\mu + 2\sum_n r_{n,k} x_n\right)}^{2\lambda^*\mu^*}\mu_k$$

$$= \frac{1}{2}\log \tau_k - \frac{\tau_k \lambda^*}{2}(\mu_k - \mu^*)^2 + (\sigma^* - 1)\log \tau_k - \overbrace{\left(a - \frac{\lambda^*(\mu^*)^2}{2}\right)}^{\rho^*}\tau_k$$

$$\sigma^* = \sigma + \frac{1}{2}\sum_n r_{n,k} \qquad \lambda^* = \lambda + \sum_n r_{n,k} \qquad \mu^* = \left(\lambda\mu + \sum_n r_{n,k} x_n\right) / \left(\lambda + \sum_n r_{n,k}\right)$$

$$\rho^* = \rho + \frac{\lambda\mu^2}{2} + \frac{1}{2}\sum_n r_{n,k} x_n^2 - \frac{\lambda^*(\mu^*)^2}{2}$$

$q(\pi, \mu, T)$ factorize as $\underline{q(\pi)} \prod_k q(\mu_k | T_k)$

Joint $p(X, Z, \pi, \mu, T) = \underbrace{p(X, Z | \pi, \mu, T)} p(\pi) p(\mu, T)$

$p(X | Z, \mu, T) \, p(Z | \pi) \qquad p(Z | \pi) = \prod_{n,k} \pi_k^{z_{nk}}$

$$\log q^*(\pi) = E_{-\pi}\left[\log p(X, Z, \pi, \mu, T)\right]$$

$$= E_{-\pi}\left[\sum_{n,k} z_{nk} \log \pi_k + \sum_k (\alpha_k - 1) \log \pi_k\right]$$

$$= \sum_{n,k} r_{nk} \log \pi_k + \sum_k (\alpha_k - 1) \log \pi_k$$

$$= \sum_k \left[\left(\sum_n r_{nk}\right) + (\alpha_k - 1)\right] \log \pi_k$$

So $q^*(\pi)$ is $\mathrm{Dir}(\alpha^*)$

where

$\alpha_k^* = \alpha_k + \sum_n r_{nk}$

$$\log q^*(Z) \stackrel{+}{=} \mathbb{E}_{\pi, \mu, \tau}\left[\log p(X, Z \mid \pi, \mu, \tau)\right]$$

$$= \sum_{n,k} z_{n,k} \underbrace{\left(\mathbb{E}_{\mu_k, \tau_k}\left[\log \mathcal{N}(x_n \mid \mu_k, \tau_k^{-1})\right] + \mathbb{E}_{\pi}\left[\log \pi_k\right]\right)}_{\log u_{n,k}}$$

① $\mathbb{E}_{\pi}[\log \pi_k]$: use digamma

② $\mathbb{E}_{\tau_k}[\log \tau_k]$: use digamma

③ $\mathbb{E}_{\tau_k}[\tau_k]$: is shape/rate

④ $\mathbb{E}_{\tau_k, \mu_k}[\tau_k (x_n - \mu_k)^2]$

④ $\mathbb{E}_{\hat{\tau}_n, \mu_n}\left[\hat{\tau}_n (x_n - \mu_n)^2\right]$

$\frac{1}{\hat{\tau}_n} c = V = \mathbb{E}_{\mu_n}[\mu_n^2 | \hat{\tau}_n] - \mathbb{E}_{\mu_n}[\mu_n | \hat{\tau}_n]^2$

$\mathbb{E}_{\mu_n}[\mu_n | \hat{\tau}_n] = m$

$= \mathbb{E}_{\hat{\tau}_n}\left[\hat{\tau}_n \, \mathbb{E}_{\mu_n}\left[(x_n - \mu_n)^2 | \hat{\tau}_n\right]\right]$

$= \mathbb{E}_{\hat{\tau}_n}\left[\hat{\tau}_n \, \mathbb{E}_{\mu_n}\left[x_n^2 - 2 x_n \mu_n + \mu_n^2 | \hat{\tau}_n\right]\right]$

$= \mathbb{E}_{\hat{\tau}_n}\left[\hat{\tau}_n \left(x_n^2 - 2 x_n m + \frac{1}{\hat{\tau}_n} c + m^2\right)\right]$

$= c + \mathbb{E}_{\hat{\tau}_n}[\hat{\tau}_n] \left(x_n^2 - 2 x_n m + m^2\right)$

use shape/rate