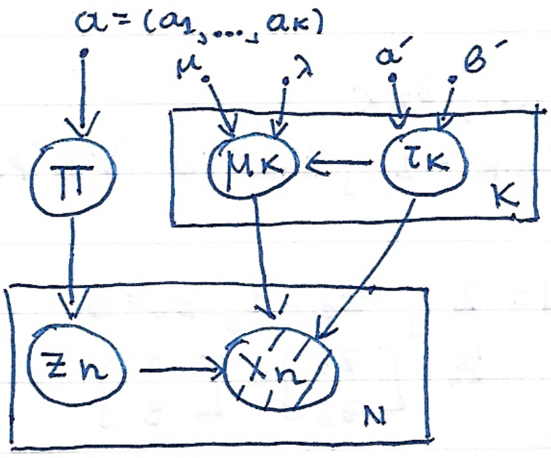
VI GMM



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

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

$M = \{\mu_K : K \in [K]\}$

$T = \{\tau_K : K \in [K]\}$

Note: $[N] = \{1, \dots, N\}$

CPD: $\pi \mid a \sim \text{Dir}(a) = \prod_{K=1}^{K} \pi_K^{a_K - 1}$

$\tau_K \mid a', b' \sim \text{Ga}(a', b') = \frac{(b')^{a'}}{\Gamma(a')} \tau_K^{a'-1} e^{-b' \tau_K}$

$\mu_K \mid \tau_K, \mu, \lambda \sim N(\mu_K \mid \mu, (\lambda \tau_K)^{-1}) =$

$= \sqrt{\frac{\lambda \tau_K}{2\pi}} \, e^{-\frac{\lambda \tau_K}{2}(\mu_K - \mu)^2}$

$z_n \mid \pi \sim \text{Cat}(\pi)$

$x_n \mid z_n = K, M, T \sim N(x_n \mid \mu_K, \tau_K) = \sqrt{\frac{\tau_K}{2\pi}} \, e^{-\frac{\tau_K}{2}(x_n - \mu_K)^2}$

this condition tells us
which $K \in [K]$ we consider
for μ_K, τ_K

Joint:

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

see graphical model

$$= p(X \mid Z, M, T)\, p(Z \mid \pi) \quad p(\pi) \underbrace{p(M \mid T)\, p(T)}_{p(M, T)}$$

VI <u>assumption</u> $\quad q(Z, \pi, M, T) = q(Z)\, q(\pi, M, T)$

The assumption is that Z is independent of (M, Λ, T)

Notation $\quad z_n \to z_{n1}, \dots, z_{nK}$

where $\quad z_n = k \iff z_{nk} = 1 \,,\; z_{nk'} = 0 \quad \forall k' \neq k$

example: $\quad K = 3 \,,\quad N = 2 \,,\quad z_1 = 1 \,,\; z_2 = 3$

Then, instead of $\quad Z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

we write
$$Z = \begin{array}{c} \begin{array}{ccc} 1 & & \end{array} \\ \begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} \end{bmatrix} \\ \begin{array}{ccc} & & 3 \end{array} \end{array}$$

Complete ~~log~~ likelihood:

$$\cancel{\left(p(X, Z \mid \Pi, M, T) = p(X \mid Z, \Pi, M, T)\, p(Z)\right)}$$

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

$$= \prod_n p(x_n \mid z_n, \Pi, M, T) \cdot p(z_n \mid \Pi, M, T)$$

$$\overset{\text{see gr.}}{\underset{\text{model}}{=}} \prod_n p(x_n \mid z_n, M, T)\, p(z_n \mid \Pi)$$

$$= \prod_{n,k} \Big[p(x_n \mid \mu_k, T_k)\, \underbrace{p(z_{nk} = 1 \mid \Pi)}_{\to \text{ recall } z \mid \Pi \sim \text{Categorical}} \Big]^{z_{n,k}}$$

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

Using $z_{n,k}$ is just a trick so that we write our likelihood more compactly

In previous example:

$$p(X, Z \mid \pi, M, T) = \prod_{n=1}^{2} p(x_n \mid z_n = k_n, M, T) \cdot p(z_n = k_n \mid \pi)$$

$$= [p(x_1 \mid z_1 = 1, M, T) \cdot \underbrace{p(z_2 = 1 \mid \pi)}_{/\!/}] \, [p(x_2 \mid z_2 = 3, M, T) \, p(z_2 = 3 \mid \pi)]$$

$$= [N(x_1 \mid \mu_1, \tau_1) \cdot \pi_1] \cdot [N(x_2 \mid \mu_3, \tau_3) \, \pi_3]$$

$$= [N(x_1 \mid \mu_1, \tau_1) \cdot \pi_1]^{1 \, '' \, r_{n1}} \cdot [N(x_1 \mid \mu_2, \tau_2) \, \pi_2]^{0 \, '' \, r_{n2}} [N(x_1 \mid \mu_3, \tau_3)$$

$$\cdot [N(x_2 \mid \mu_1, \tau_1) \cdot \pi_1]^{0 = r_{n2}} \ldots\ldots [N(x_2 \mid \mu_3, \tau_3) \, \pi_3)^{1}$$

$$= \prod_{n=1}^{2} [p(x_n \mid \mu_k, \tau_k) \, p(z_{nk} = 1 \mid \pi)]^{z_{nk}}$$

Update equations

Latent variables (not observed): $\{Z, \pi, M, T\}$

In general, if $\Theta = (\theta_1, \ldots, \theta_m)$ our latent variables, then $\log q^*(\theta_j) = \underset{i \neq j}{E}[\log p(X, \Theta)]$ + const

← we ignore

Here $\Theta = (Z, \pi, M, T)$:

$$\log \underset{\text{the rest}}{q^*(z)} = \underset{\pi, M, T}{E} [\log p(X, Z, \pi, M, T)] + \text{const}$$

$$= \underset{\pi, M, T}{E} [\log p(X, Z \mid \pi, M, T) + \log p(\pi, M, T)] + \text{const}$$

However, $p(M, \Pi, T)$ does not depend on Z,

$\rightarrow$ $\boxed{\text{we can ignore it}}$

Thus $\log q^*(Z) \doteq E_{\Pi, M, T}[\log p(X, Z \mid \Pi, M, T)]$

$$\log p(X, Z \mid \Pi, M, T) = \log \prod_{n,k} [N(x_n \mid \mu_k, \tau_k^{-1})\, \pi_k]^{z_{n,k}}$$

$$= \sum_{n,k} z_{n,k} [\log(N(x_n \mid \mu_k, \tau_k^{-1})) + \log \pi_k]$$

$$\doteq \sum_{n,k} z_{n,k} \left[\log\left[\sqrt{\frac{\tau_k}{2\pi}}\, e^{-\frac{\tau_k}{2}(x_n - \mu_k)^2}\right] + \log \pi_k\right]$$

$$\doteq \sum_{n,k} z_{n,k} \left[\frac{\tau_k}{2} - \frac{\tau_k}{2}(x_n - \mu_k)^2 + \log \pi_k\right]$$

Expectation is linear: ~~since~~ ~~ΠM~~

$$\log q^*(Z) = \sum_{n,k} z_{n,k} \left[\frac{E_T[\tau_k]}{2} - \frac{\overbrace{E_{T,M}[\tau_k (x_n - \mu_k)^2]}^{E_T[\tau_k] \cdot E_M[(x_n - \mu_k)^2]}}{2} + E_\Pi[\log \pi_k]\right]$$

$$= \sum_{n,k} z_{n,k} \underbrace{\left[\frac{E_{\tau_k}[\tau_k]}{2} - \frac{E_{\tau_k}[\tau_k]}{2} E_{\mu_k}[(x_n - \mu_k)^2] + E_{\pi_k}[\log \pi_k]\right]}_{\log \rho_{n,k}}$$

$$= \sum_{n,k} z_{n,k} \log \rho_{n,k}$$

So $q^*(z_n) = \prod_k r_{nk}^{z_{nk}}$, $r_{nk} = \dfrac{\rho_{nk}}{\sum_k \rho_{nk}}$

Why? $\quad \left(q(Z) = \prod_n q(z_n) \right)$

$$\log q(Z) = \sum_{n,K} z_{n,K} \log \rho_{n,K} = \sum_{n,K} \log \rho_{n,K}^{z_{n,K}} =$$

$$= \log \prod_{n,K} \rho_{nK}^{z_{nK}} = \log \prod_n q(z_n) \Rightarrow$$

$$q(z_n) = \prod_K \rho_{nK}^{z_{nK}}$$

(Not sure why he decides / normalizes for r_{nK})

Now for $q^*(\pi, M, T)$

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

$$= E_Z[\log p(X, Z \mid \pi, M, T) + \underbrace{\log P(\pi, M, T)}_{\text{not depend on } z}]$$

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

$$\overset{\text{recall}}{\underset{\log p(X, Z \mid \pi M, T)}{=}} E_Z[\log(\prod_{n,K} [N(x_n \mid \mu_K, T_K^{-1}) \pi_K]^{z_{nK}})]$$

$$+ \log[P(\pi) \cdot p(M, T)]$$

$$= \sum_{n,K} E_Z[z_{n,K}] \log N(x_n \mid \mu_K, T_K^{-1})$$

$$+ \sum_{n,K} E_Z[z_{n,K}] \log \pi_K + \overbrace{\log p(\pi)}^{\log \prod_K \pi_K^{(\alpha_K - 1)} = \sum_K (\alpha_K - 1) \log \pi_K} + \log P(M, T)$$

~~Note how π appears~~

$$E_Z[z_{nk}] = r_{nk} \quad \text{(no idea)}$$

$$q(\pi, \mu, \tau) = q(\pi) \prod_k q(\mu_k, \tau_k)$$

$$(\pi \perp\!\!\!\perp \mu_k, \tau_k)$$

Lets examine $q^*(\mu_k, \tau_k)$:
We ~~ne~~ examine previous expression and extract the terms containing μ_k, τ_k:

$$\sum_{n,k} r_{nk} \underbrace{\log\left[\sqrt{\tfrac{\tau_k}{2\pi}}\, e^{-\frac{\tau_k}{2}(x_n-\mu_k)^2}\right]}_{\tau_k, \mu_k} + \overbrace{\sum_{n,k} r_{nk} \log \pi_k + \sum_k (a_k-1)\log \pi_k}^{\text{only } \pi_k}$$

$$+ \sum_k \underbrace{\log p(\mu_k, \tau_k)}_{\shortparallel} \quad = \mathcal{N}(\mu, (\lambda\tau_k)^{-1})$$

$$\log p(\mu_k|\tau_k) + \log p(\tau_k) \overset{+}{=} \frac{\log \tau_k}{2} - \frac{\lambda\tau_k}{2}(\mu_k-\mu)^2$$

$$+ (a'-1)\log\tau_k - b'\tau_k$$

Thus (you fix "k", so there are no $\sum_k$)

$$\log q^*(\mu_k, \tau_k) \overset{+}{=} \sum_n r_{nk}\left(\frac{\log\tau_k}{2} - \frac{\tau_k}{2}(x_n-\mu_k)^2\right)$$

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

$$- b'\tau_k$$

$$= -\frac{1}{2}[\lambda \tau_K + (\sum_n r_{n,K})\tau_K]\mu_K^2 + [\lambda \tau_K \mu + \tau_K(\sum_n r_{nK} x_n)]$$

$$+ [a' - 1 + \frac{1}{2} + \frac{1}{2}\sum_n r_{nK}] \log \tau_K - (b' + \frac{\lambda \mu^2}{2} + \frac{1}{2}\sum_n r_{nK}x$$

$$\tau_K$$

log of gaussian: $-\frac{\tau^*}{2}Y^2 + \tau \mu Y$

log of gamma: $(a-1)\log Y - bY$

Thus: $q^*(\mu_K | \tau_K) \simeq N(\mu_K | \mu^*, \tau^*)$

$$\cancel{\mu^*} \cdot \tau^* = \lambda \tau_K + \tau_K \sum_n r_{nK} = \tau_K(\lambda + \sum_n r_{nK})$$

$$\mu^* = \frac{\lambda \tau_K \mu + \tau_K \sum_n r_{nK} x_n}{\tau^*}$$

$$q^*(\tau_K) = Ga(\tau_K | a'^*, b'^*)$$

$$a'^* = a' + \frac{1}{2} + \frac{1}{2}\sum_n r_{nK}$$

$$b'^* = b' + \frac{\lambda \mu^2}{2} + \frac{1}{2}\sum_n r_{nK} x_n^2$$

For $q^*(\pi)$: recall that only the following has π:

$$\sum_{n,K} r_{n,K} \log \pi_K + \sum_K (a_K - 1)\log \pi_K$$

$$= \sum_K [(a_K - 1) + \sum_n r_{nK}] \log \pi_K$$

log of dir: $(a-1)\cancel{\#}\log \pi$

Thus: $a^* = a + \sum_n r_{nk}$

For $\log \rho_{nk}$

$E_{\tau_k}[\log \tau_k]$, $E_\pi[\log \pi_k]$ ~ digamma *

$E_{\tau_k}[\tau_k] = \frac{a^*}{b^*}$

$E_{\mu_k}[\cancel{\log}(x_n - \mu_k)^2]$

$$= x_n^2 - 2x_n \overbrace{E_{\mu_k}[\mu_k]}^{\mu^*} + \overbrace{E_{\mu_k}[\mu_k^2]}^{\frac{1}{\tau^*} + (\mu^*)^2}$$

$$* \left(\begin{array}{l} E_{\tau_k}[\log \tau_k] \doteq \psi(\pi_k) - \psi\left(\sum_{j=1}^{K} \pi_j\right) \\ E_\pi[\log \pi_k] = \psi(\cancel{a_k - 1}) - \psi\left(\sum_{j=1}^{K} [a_j - 1]\right) \\ \psi\text{: digamma function} \end{array} \right)$$

I think :)

$$* E_{\tau_k}[\log \tau_k] = -b^* + \psi(a^*)$$

$$E_\pi[\log \pi_k] = \psi(a_k) - \psi\left(\sum_{j=1}^{K} a_j\right)$$

stackexhange :)

ψ: digamma function