$$P(z_i \mid x_i, \pi, \theta) = \frac{P(x_i, z_i \mid \pi, \theta)}{P(x_i \mid \pi, \theta)}$$

$$= \frac{\prod_{k=1}^{K} \left(\pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)\right)^{z_{ik}}}{\sum_{l=1}^{K} \pi_l \mathcal{N}(x_i; \mu_l, \Sigma_l)}$$

$$= \prod_{k=1}^{K} \Bigg(\underbrace{\frac{\pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l \mathcal{N}(x_i; \mu_l, \Sigma_l)}}_{= \tau_{ik}} \Bigg)^{z_{ik}}$$

$$= \prod_{k=1}^{K} \tau_{ik}^{z_{ik}}$$

$$= \mathcal{M}(z_i; 1, \tau_i) \qquad \tau_i = \begin{pmatrix} \tau_{i1} \\ \tau_{i2} \\ \vdots \\ \tau_{iK} \end{pmatrix}$$

$$0 \leq \tau_{ik} \leq 1 \qquad \sum_{k=1}^{K} \tau_{ik} = 1.$$

$$\underset{(x_i)_i}{L}(\pi, \alpha) = \log p\left((x_i)_i \mid \pi, \alpha\right)$$

$$= \sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)\right)$$

$$= \mathscr{L}\left(R((z_i)_i); \pi, \alpha\right) + KL\left(R((z_i)_i) \,\|\, p((z_i)_i \mid (x_i)_i, \pi, \alpha)\right)$$

variational decomposition (mean field approximation)

lower bound of $\underset{(x_i)_i}{L}(\pi, \alpha)$

$\Leftrightarrow$ ELBO: evidence lower bound

Kullback-Leibler divergence between $R((z_i)_i)$ and $p((z_i) \mid (x_i)_i, \pi, \alpha)$

$$\mathscr{L}\left(R((z_i)_i); \pi, \alpha\right) = \sum_{z_1, z_2, \dots, z_n} R((z_i)_i) \log \frac{p((x_i, z_i)_i \mid \pi, \alpha)}{R((z_i)_i)}$$

$$KL\left(R((z_i)_i) \,\|\, p((z_i)_i \mid (x_i)_i, \pi, \alpha)\right) = -\sum_{z_1, z_2, \dots, z_n} R((z_i)_i) \log \frac{p((z_i)_i \mid (x_i)_i, \pi, \alpha)}{R((z_i)_i)}$$

$$\mathscr{L} + KL = \sum_{z_1, \dots, z_n} R((z_i)_i) \log \frac{p((x_i, z_i)_i \mid \pi, \alpha)}{p((z_i)_i \mid (x_i)_i, \pi, \alpha)}$$

$$\mathcal{L} + KL = \sum_{z_1, \dots, z_n} R\big((z^i)_i\big) \log \frac{P\big((x^i, z^i)_i \mid \pi, \alpha\big)}{P\big((z^i)_i \mid (x^i)_i, \pi, \alpha\big)}$$

$$= \sum_{z_1, \dots, z_n} R\big((z^i)_i\big) \log \frac{\cancel{P\big((z^i)_i \mid (x^i)_i, \pi, \alpha\big)}\, P\big((x^i)_i \mid \pi, \alpha\big)}{\cancel{P\big((z^i)_i \mid (x^i)_i, \pi, \alpha\big)}}$$

$$= \sum_{z_1, \dots, z_n} R\big((z^i)_i\big) \log P\big((x^i)_i \mid \pi, \alpha\big)$$

$$= \log P\big((x^i)_i \mid \pi, \alpha\big)$$

$$= L_{(x^i)_i}(\pi, \alpha)$$

$$L_{(x_i)_i}(\pi, \alpha) = \mathcal{L}\left(R((z_i)_i), \pi, \alpha\right) + \underbrace{KL\left(R((z_i)_i) \,\|\, p((z_i)_i \mid (x_i)_i; \pi, \alpha)\right)}_{\geqslant 0}$$

Variational EM algorithm, <u>VEM</u>

Init:

VM step: $R((z_i)_i)$ is fixed. Maximise $\mathcal{L}$ with respect to π, α.

VE step: π, α fixed. Maximise $\mathcal{L}$ with respect to $R((z_i)_i)$.

stop when $\mathcal{L}$ converges

$\iff$ Minimise KL with respect to $R((z_i)_i)$

$\iff R((z_i)_i) = p((z_i)_i \mid (x_i)_i; \pi, \alpha)$

for GMM because we have analytical expressions!!!

($\Rightarrow$ EM!!!)

In the case of GMM:

(E-step) VE step : compute the τ_{ik} (π, θ fixed)

(M-step) VM step : $\hat{\pi}, \hat{\theta} = \underset{\pi, \theta}{\operatorname{argmax}} \mathcal{L}\left(R((z_i)_i), \pi, \theta\right)$

$$\sum_{(z_i)_i} R((z_i)_i) \log p((x_i, z_i)_i \mid \pi, \theta)$$

$$- \sum_{(z_i)_i} R((z_i)_i) \log R((z_i)_i)$$

$$= \underset{\pi, \theta}{\operatorname{argmax}} \sum_{(z_i)_i} R((z_i)_i) \log p((x_i, z_i)_i \mid \pi, \theta)$$

$$= \underset{\pi, \theta}{\operatorname{argmax}} \; \underset{\substack{Z_i \sim R \\ \forall i}}{E} \left[\log p((x_i, Z_i)_i \mid \pi, \theta)\right]$$

$\hookrightarrow$ with $R((z_i)_i) = p((z_i)_i \mid (x_i), \pi, \theta)$

In the case of GMM: VEM reduces to EM

Init: $(\tau^0_{ik})_{ik}$ (kmeans)

M_step: maximise $\mathcal{L}$ with respect to π, α $\Rightarrow \pi^0, \alpha^0$

E-step: maxime $\mathcal{L}$ with respect to R $\Rightarrow (\tau^1_{ik})_{ik}$

$$\left[\mathcal{L}_{(x_i)_i}(\pi^0, \alpha^0) = \mathcal{L}\left((\tau^1_{ik})_{ik}; \pi^0, \alpha^0\right) \geq \mathcal{L}\left((\tau^0_{ik})_{ik}; \pi^0, \alpha^0\right) \right.$$

M-step: maximise $\mathcal{L}$ with respect to π, α $\Rightarrow \pi^1, \alpha^1$

$$\left[\mathcal{L}\left((\tau^1_{ik})_{ik}; \pi^1, \alpha^1\right) \geq \mathcal{L}\left((\tau^1_{ik})_{ik}; \pi^0, \alpha^0\right) \right.$$

E-step: maximise $\mathcal{L}$ with respect to R $\Rightarrow (\tau^2_{ik})_{ik}$

$$\mathcal{L}_{(x_i)_i}(\pi^1, \alpha^1) = \mathcal{L}\left((\tau^2_{ik})_{ik}; \pi^1, \alpha^1\right) \geq \mathcal{L}\left((\tau^1_{ik})_{ik}; \pi^1, \alpha^1\right)$$

Thus: $L_{(x_i)_i}(\pi^1, \theta^1) = \mathcal{L}((\tau^2_{ik})_{ik}; \pi^1, \theta^1)$

$$\geq \mathcal{L}((\tau^1_{ik})_{ik}; \pi^1, \theta^1)$$

$$\geq \mathcal{L}((\tau^1_{ik})_{ik}; \pi^0, \theta^0)$$

$$= L_{(x_i)_i}(\pi^0, \theta^0)$$

so $\quad L_{(x_i)_i}(\pi^1, \theta^1) \geq L_{(x_i)_i}(\pi^0, \theta^0)$

$$\log p(\beta \mid X, Y, \sigma^2)$$

$$= \log p(Y \mid X, \beta, \sigma^2) + \log p(\beta) + \cancel{ct}_1$$

$$= -\frac{1}{2}(Y - X\beta)^T(\sigma^2 I_n)^{-1}(Y - X\beta) - \frac{1}{2}\beta^T\left(\frac{I_p}{\alpha}\right)^{-1}\beta + \cancel{ct}_2$$

$$= -\frac{1}{2\sigma^2}\|Y - X\beta\|^2 - \frac{\alpha}{2}\|\beta\|^2 + \cancel{ct}_2$$

$$\underset{\beta}{\operatorname{argmax}} \log p(\beta \mid X, Y, \sigma^2)$$

$$= \underset{\beta}{\operatorname{argmin}} \left\{ \|Y - X\beta\|^2 + \underbrace{\alpha\sigma^2}_{\lambda} \|\beta\|^2 \right\}$$