EM with GMMs:

MLE

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log p_\theta(x_i)$$

The case of mixture models & assuming the densities are Gaussians.

$$\theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^{m}$$

$$\sum_{i=1}^{n} \log p_\theta(x_i) = \sum_{i=1}^{n} \log \left[\sum_{j=1}^{m} \alpha_j \, p_{\theta_j}(x_i) \right]$$

log of sums

$$p(x_i) = \sum_{k=1}^{m} \alpha_k \, \mathcal{N}(x_i; \mu_k, \Sigma_k)$$

$$p(z=k) = \alpha_k$$

$$\sum_{k=1}^{m} \alpha_k = 1$$

$$p(x_i \mid z=k) = \mathcal{N}(x_i; \mu_k, \Sigma_k)$$

Recall $$F_\theta(q) = \mathop{\mathbb{E}}_{q(z)} \log \frac{p_\theta(x, z)}{q(z)}$$

The optimal value

$$q(z) = p_\theta(z/x)$$

$$p_\theta(Z=k \mid x_i) = \frac{p(x_i, Z=k)}{p(x_i)}$$

$$= \frac{p(Z=k)\, p(x_i \mid Z=k)}{\sum_{j=1}^{m} \alpha_j N(x_i ; \mu_j, \Sigma_j)}$$

$$= \frac{\alpha_k \quad N(x_i ; \mu_k, \Sigma_k)}{\sum_{j=1}^{m} \alpha_j N(x_i ; \mu_j, \Sigma_j)}$$

$$= \gamma_{ik}$$

for $t=1$ to Convergence :

$$q^{t+1}(z) = p_{\theta^t}(z \mid x_i)$$

Compute

$$F_\theta(q^{t+1}) = \mathbb{E}_{q^{t+1}(z)} \log \frac{p_\theta(x_i, z)}{q^{t+1}(z)}$$

$$\theta^{t+1} = \underset{\theta}{\operatorname{argmax}} \; F_\theta(q^{t+1})$$

E-step:

$$F_\theta(q) = \mathbb{E}_{q(z)} \log \frac{p_\theta(x_i, z)}{q(z)}$$

$$= \mathbb{E}_{q(z)} \left\{ \log p_\theta(x_i, z) - \log q(z) \right\}$$

$$= \mathbb{E}_{q(z)} \log p_\theta(x_i, z)$$

$$= \mathbb{E}_{p(z \mid x_i)} \log p_\theta(x_i, z)$$

$$= \sum_{j=1}^{m} p(z=j \mid x_i) \log p_\theta(x_i, z=j)$$

$$= \sum_{j=1}^{m} p(z=j \mid x_i) \left\{ \log p_\theta (x_i \mid z=j) + \log p(z=j) \right\}$$

$$= \sum_{j=1}^{m} \gamma_{ij} \left\{ \log N(x_i ; \mu_j, \Sigma_j) + \log \alpha_j \right\}$$

Optimize this wrt to Constraint s.t

$$\sum_{j=1}^{m} \alpha_j = 1$$

$$L = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} \log N(x_i ; \mu_j, \Sigma_j) + \sum_{j=1} \gamma_{ij} \log \alpha_j + \lambda \left\{ \sum_{j=1}^{m} \alpha_j - 1 \right\}$$

M-Step.

$$\frac{\partial L}{\partial \alpha_j} = 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{\alpha_j \, N(x_i ; \mu_j, \Sigma_j)}{\sum_{k=1}^{m} \alpha_k N(x_i ; \mu_k, \Sigma_k)} + \lambda \alpha_j = 0$$

$$\sum_{i=1}^{n} \gamma_{ij} + \lambda \alpha_j = 0$$

$$\alpha_j = \frac{\sum_{i=1}^{n} \gamma_{ij}}{-\lambda}$$

using the Constraint

$$\sum_{j=1}^{m} \alpha_j = 1$$

$$\frac{-1}{\lambda} \sum_{j=1}^{m} \sum_{i=1}^{n} \gamma_{ij} = 1$$

$$\frac{-1}{\lambda} \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} = 1$$

wkt $\sum_{j=1}^{m} \gamma_{ij} = 1$

$$\lambda = -n$$

$$\Rightarrow \quad \alpha_j = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ij}$$

now for μ_j

$$\frac{\partial L}{\partial \mu_j} = 0$$

$$- \sum_{i=1}^{n} \frac{\alpha_j \, N(x_i ; \mu_j, \Sigma_j)}{\sum_{k=1}^{m} \alpha_k \, N(x_i ; \mu_k, \Sigma_k)}$$

$$\Sigma_j \, (x_i - \mu_j) = 0$$

$$\sum_{i=1}^{n} \gamma_{ij} \, x_i - \sum_{i=1}^{n} \gamma_{ij} \, \mu_j = 0$$

$$\mu_j = \sum_{i=1}^{n} \gamma_{ij} \, x_i \Big/ \sum_{i=1}^{n} \gamma_{ij}$$

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^{n} \gamma_{ij} \, x_i$$

|||ly diff wrt Σ_j & equate to zero.

$$\Sigma_j = \frac{1}{n_j} \sum_{i=1}^{n} \gamma_{ij} \, (x_i - \mu_j)(x_i - \mu_j)^T$$

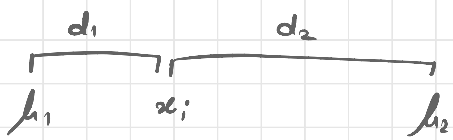
Proof of k-means a special case of GMM:

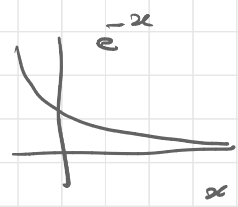
Assume $\Sigma_k = \sigma^2 I$

$$\gamma_{ik} = \frac{\alpha_k \, N(x_i ; \mu_k, \Sigma_k)}{\sum_{j=1}^{m} \alpha_j \, N(x_i ; \mu_j, \Sigma_j)}$$

$$= \frac{\alpha_k \exp\left\{ -\frac{1}{2\sigma^2} \| x_i - \mu_k \|^2 \right\}}{\sum_{j=1}^{m} \alpha_j \exp\left\{ -\frac{1}{2\sigma^2} \| x_i - \mu_j \|^2 \right\}}$$

as $\sigma^2 \to 0$





$$\exp\left\{\frac{-1}{\beta} d_1\right\} + \exp\left\{\frac{-1}{\beta} d_2\right\}$$

as $\sigma^2 \to 0$, the summation of denominator will be dominated by term with smallest $\|x_i - \mu_j\|^2$

for $\forall\, l \neq j$

$$\left.\begin{aligned} r_{il} &= 0 \\ r_{ij} &= 1 \end{aligned}\right\} \text{Hard assignment}$$