EM – algorithm 求取 $\alpha_j^{(t+1)}$

$$Q(\theta|\theta^{(t)}) \equiv E_y[\log P(x,y|\theta)\,|x,\theta^{(t)}] = \sum_i^n \sum_j^k \left[\log\left(\alpha_j \phi\left(x_i \middle| \mu_j,\sigma_j\right)\right) P\left(y_i = j \middle| x_i, \theta^{(t)}\right)\right]$$

給定限制式：$\sum_j^k \alpha_j = 1$

找α_j使得 $Q(\theta|\theta^{(t)})$最大

使用 Lagrange 乘數

$$L = \sum_i^n \sum_j^k \left[\log\left(\alpha_j \phi\left(x_i \middle| \mu_j,\sigma_j\right)\right) P\left(y_i = j \middle| x_i, \theta^{(t)}\right)\right] + \lambda\left(\sum_j^k \alpha_j - 1\right)$$

$$\frac{\partial}{\partial \alpha_j} L = \sum_i^n \frac{\phi\left(x_i \middle| \mu_j,\sigma_j\right)}{\alpha_j \phi\left(x_i \middle| \mu_j,\sigma_j\right)} P\left(y_i = j \middle| x_i, \theta^{(t)}\right) + \lambda$$

$$\frac{\partial}{\partial \lambda} L = \sum_j^k \alpha_j - 1$$

令 $\frac{\partial}{\partial \alpha_j} L$ 以及 $\frac{\partial}{\partial \lambda} L$ 為 0

$$\frac{\partial}{\partial \alpha_j} L = 0 \Rightarrow \lambda = -\sum_i^n \frac{\phi\left(x_i \middle| \mu_j,\sigma_j\right)}{\alpha_j \phi\left(x_i \middle| \mu_j,\sigma_j\right)} P\left(y_i = j \middle| x_i, \theta^{(t)}\right) \rightarrow \alpha_j = -\sum_i^n \frac{P\left(y_i = j \middle| x_i, \theta^{(t)}\right)}{\lambda} \quad \text{--- (1)}$$

$$\frac{\partial}{\partial \lambda} L = 0 \Rightarrow \sum_j^k \alpha_j = 1 \quad \text{--- (2)}$$

將(1)代入(2)

$$-\sum_j^k \sum_i^n \frac{P\left(y_i = j \middle| x_i, \theta^{(t)}\right)}{\lambda} = 1 \Rightarrow \lambda = -\sum_j^k \sum_i^n P\left(y_i = j \middle| x_i, \theta^{(t)}\right)$$

$$\because \sum_j^k P\left(y_i = j \middle| x_i, \theta^{(t)}\right) = 1 \ \therefore \lambda = -\sum_i^n 1 = -n \quad \text{--- (3)}$$

將(3)代回(1)得到 α_j

$$\alpha_j^{(t+1)} = -\sum_i^n \frac{P\left(y_i = j \middle| x_i, \theta^{(t)}\right)}{-n} = \frac{\sum_i^n P\left(y_i = j \middle| x_i, \theta^{(t)}\right)}{n}$$