

Expectation Maximization for the Gaussian Mixture

$Z \sim \text{Cat}(\underline{\pi})$
 $X \sim \mathcal{N}(\mu_z, \sigma_z^2)$

We have data
 $D = \{ (z, x), (z, x), (z, x), \dots \}$
What are $\underline{\pi}, \underline{\mu}, \underline{\sigma}$?

E-step $\beta_z^{c,i} = p(z=z | X=x^{c,i}; \underline{\theta}^{old})$
M-step $\underline{\theta}^{old} = \arg \max_{\underline{\theta}} \sum_{i=0}^{N-1} \sum_{z \in Z} p(z=z, x=x^{c,i}; \underline{\theta})$

joint of the Gaussian Mixture

$$p(z, x) = p(z) \cdot p(x|z)$$
$$= \text{Cat}(\underline{\pi}) \cdot \mathcal{N}(\mu_z, \sigma_z^2)$$
$$= \left(\prod_{d=0}^{D-1} \pi_d^{I(d=z)} \right) \cdot \left(\frac{1}{\sigma_z \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_z^2} \cdot (X - \mu_z)^2\right) \right)$$

① E-step

$$\beta_z^{c,i} = p(z=z | X=x^{c,i}; \underline{\theta}^{old})$$

We don't know the posterior $p(z)$

$\beta =$

\rightarrow Apply Bayes' Rule

$$p(z=z | X=x^{c,i}; \underline{\theta}^{old}) = \frac{p(X=x^{c,i} | z=z; \underline{\theta}^{old}) p(z=z; \underline{\theta}^{old})}{p(X=x^{c,i}; \underline{\theta}^{old})}$$

something tough

\rightarrow work with proportional and normalize later on

$$\propto p(z=z, X=x^{c,i}; \underline{\theta}^{old}) =: \tilde{\beta}_z^{c,i}$$

Params $\underline{\theta} = [\underline{\pi}^T, \underline{\mu}^T, \underline{\Sigma}^T]^T$ unnormalized responsibility

$$\tilde{\beta}_z^{c,i} = \pi_z \cdot \frac{1}{\sigma_z \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_z^2} \cdot (x^{c,i} - \mu_z)^2\right)$$

$$\beta_z^{c,i} = \frac{\tilde{\beta}_z^{c,i}}{\sum_{d=0}^{D-1} \tilde{\beta}_d^{c,i}} \quad (\text{row normalization})$$

② M-step

$$Q(D; \underline{\theta}^{old}, \underline{\theta}) = \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \beta_c^{c,i} \cdot \log(p(z=c, X=x^{c,i}; \underline{\theta}))$$
$$= \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \beta_c^{c,i} \cdot \left(\sum_{d=0}^{D-1} I(d=c) \log(\pi_d) + \log\left(\frac{1}{\sigma_c \sqrt{2\pi}}\right) - \frac{1}{2\sigma_c^2} (x^{c,i} - \mu_c)^2 \right)$$
$$+ \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \beta_c^{c,i} \cdot \left(\sum_{d=0}^{D-1} I(d=c) \log(\pi_d) - \log(\sigma_c) - \frac{1}{2} \log(2\pi) - \frac{1}{2\sigma_c^2} (x^{c,i} - \mu_c)^2 \right)$$

$$\underline{\pi}^*, \underline{\mu}^*, \underline{\Sigma}^* = \arg \max_{\substack{\underline{\pi} \in [0,1]^D \\ \underline{\mu} \in \mathbb{R}^D \\ \underline{\Sigma} \in \mathbb{R}_+^D \\ \text{constraint: } \sum_{d=0}^{D-1} \pi_d = 1}} (Q(D; \underline{\theta}^{old}, \underline{\theta}))$$

constrained optimization

\hookrightarrow Lagrange Multiplier
 $g(\underline{\pi}) = 1 - \sum_{d=0}^{D-1} \pi_d$

$$\hat{Q}(D; \underline{\theta}^{old}, \underline{\theta}, \lambda) = Q(D; \underline{\theta}^{old}, \underline{\theta}) + \lambda g(\underline{\pi})$$
$$= \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \left(\beta_c^{c,i} \left(\sum_{d=0}^{D-1} I(d=c) \log(\pi_d) \right) - \log(\sigma_c) - \frac{1}{2\sigma_c^2} (x^{c,i} - \mu_c)^2 \right) + \lambda \left(1 - \sum_{d=0}^{D-1} \pi_d \right)$$

\hookrightarrow how: unconstrained: take derivative and set to zero

① $\underline{\pi}$

$\pi_e \dots$ the e-th component of $\underline{\pi}$

$$\frac{\partial \hat{Q}}{\partial \pi_e} = \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \left(\beta_c^{c,i} I(e=c) \left(\frac{1}{\pi_e} \right) \right) - \lambda \stackrel{!}{=} 0$$
$$= \frac{1}{\pi_e} \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \left(\beta_c^{c,i} I(e=c) \right) - \lambda \stackrel{!}{=} 0$$
$$= \frac{1}{\pi_e} \sum_{i=0}^{N-1} \underbrace{\left(\beta_e^{c,i} \right)}_{\gamma_e^i} - \lambda \stackrel{!}{=} 0$$

$\gamma_e^i \dots$ the responsibility mass associated with the e-th class over all samples

$$\frac{\gamma_e^i}{\pi_e} - \lambda = 0 \quad \Leftrightarrow \quad \lambda = \frac{\gamma_e^i}{\pi_e}$$
$$\frac{\partial \hat{Q}}{\partial \lambda} = 1 - \sum_{d=0}^{D-1} \pi_d \stackrel{!}{=} 0 \quad \hookrightarrow \quad \sum_{d=0}^{D-1} \pi_d = 1$$

$\left[\lambda = \frac{\gamma_e^i}{\pi_e} = \text{const.} = \frac{\gamma_{e-1}^i}{\pi_{e-1}} = \frac{\gamma_{e+1}^i}{\pi_{e+1}} = \frac{\gamma_e^i}{\pi_e} \right]$

how do get rid of the λ ?

Strike

$$\lambda = \lambda = \lambda \cdot 1 = \lambda \cdot \sum_{d=0}^{D-1} \pi_d = \sum_{d=0}^{D-1} \lambda \cdot \pi_d = \sum_{d=0}^{D-1} \frac{\gamma_d^i}{\pi_d} \cdot \pi_d = \sum_{d=0}^{D-1} \gamma_d^i = N$$
$$N = \frac{\gamma_e^i}{\pi_e} \quad \Leftrightarrow \quad \boxed{\pi_e = \frac{\gamma_e^i}{N}}$$

② $\underline{\mu}$

$\mu_e \dots$ e-th component $\underline{\mu}$

$$\frac{\partial \hat{Q}}{\partial \mu_e} = \sum_{i=0}^{N-1} \beta_e^{c,i} \cdot \left(- \frac{x^{c,i} - \mu_e}{\sigma_e^2} \right) \stackrel{!}{=} 0 \quad | \cdot \sigma_e^2$$
$$= \sum_{i=0}^{N-1} \beta_e^{c,i} \cdot (x^{c,i} - \mu_e) \stackrel{!}{=} 0$$
$$= \sum_{i=0}^{N-1} \beta_e^{c,i} x^{c,i} - \sum_{i=0}^{N-1} \beta_e^{c,i} \mu_e \stackrel{!}{=} 0$$

γ_e^i

$$= \sum_{i=0}^{N-1} \beta_e^{c,i} x^{c,i} - \gamma_e^i \mu_e \stackrel{!}{=} 0$$
$$\Rightarrow \quad \boxed{\mu_e = \frac{1}{\gamma_e^i} \sum_{i=0}^{N-1} \beta_e^{c,i} x^{c,i}}$$

③ $\underline{\sigma}$

$\sigma_e \dots$ e-th component of $\underline{\sigma}$

$$\frac{\partial \hat{Q}}{\partial \sigma_e} = \sum_{i=0}^{N-1} \beta_e^{c,i} \cdot \left(-\frac{1}{\sigma_e} + \frac{(x^{c,i} - \mu_e)^2}{\sigma_e^3} \right) \stackrel{!}{=} 0$$
$$= \sum_{i=0}^{N-1} \beta_e^{c,i} \cdot \left(-\frac{1}{\sigma_e} + \frac{1}{\sigma_e^3} (x^{c,i} - \mu_e)^2 \right) \stackrel{!}{=} 0 \quad | \cdot \sigma_e^3$$
$$= \sum_{i=0}^{N-1} \beta_e^{c,i} \cdot (-\sigma_e^2 + (x^{c,i} - \mu_e)^2) \stackrel{!}{=} 0$$
$$= \sum_{i=0}^{N-1} \beta_e^{c,i} (x^{c,i} - \mu_e)^2 - \sum_{i=0}^{N-1} \beta_e^{c,i} \sigma_e^2 \stackrel{!}{=} 0$$

γ_e^i

$$\sigma_e = \sqrt{\frac{1}{\gamma_e^i} \sum_{i=0}^{N-1} \beta_e^{c,i} (x^{c,i} - \mu_e)^2}$$

Summary: EM for the GMM

E-step

- calculate unnormalized responsibilities
$$\tilde{\beta}_c^{c,i} = \pi_c \cdot \frac{1}{\sigma_c \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_c^2} \cdot (x^{c,i} - \mu_c)^2\right)$$
 - normalize responsibilities:
$$\beta_c^{c,i} = \frac{\tilde{\beta}_c^{c,i}}{\sum_{d=0}^{D-1} \tilde{\beta}_d^{c,i}}$$
 - calculate class responsibilities
$$\gamma_c^i = \sum_{i=0}^{N-1} \beta_c^{c,i}$$
- ### M-step
- update the class probabilities
$$\pi_c \leftarrow \frac{\gamma_c^i}{N}$$
 - update the mus
$$\mu_c \leftarrow \frac{\sum_{i=0}^{N-1} \beta_c^{c,i} x^{c,i}}{\gamma_c^i}$$
 - update the sigmas
$$\sigma_c \leftarrow \sqrt{\frac{1}{\gamma_c^i} \sum_{i=0}^{N-1} \beta_c^{c,i} (x^{c,i} - \mu_c)^2}$$