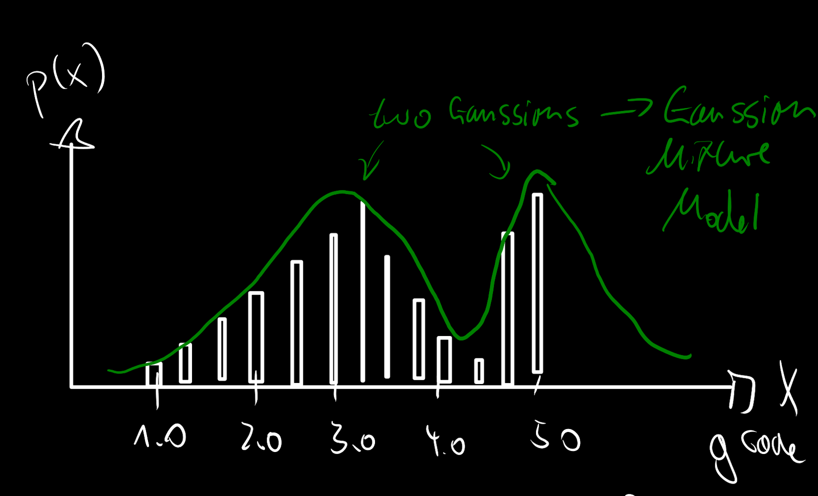
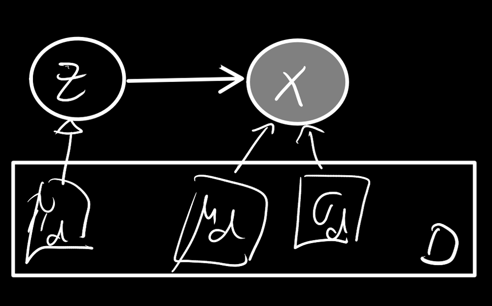
Expectation Maximization for the Gaussian Mixture





$Z \sim Cat(\underline{\pi})$
$X \sim \mathcal{N}(\mu_Z, \sigma_Z)$

E-Step $\rho_z^{[i]} = p(Z=z \mid X=x^{[i]}; \underline{\theta}^{[k]})$

M-Step $\underline{\theta}^{[k+1]} = \arg\max_{\underline{\theta}} \sum_{i=0}^{N-1} \sum_{z \in Z} \rho_z^{[i]} \log p(Z=z, X=x^{[i]}; \underline{\theta})$

we have data
$D = \{(?, 1.7), (?, 3.0), (?, 2.7), \dots\}$
what are $\underline{\pi}, \underline{\mu}$ & $\underline{\sigma}$?

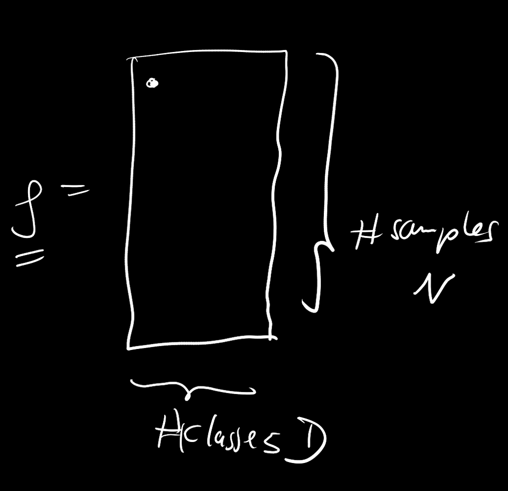
joint of the Gaussian Mixture

$$p(Z,X) = p(Z) \cdot p(X \mid Z)$$
$$= Cat(\underline{\pi})\, \mathcal{N}(\mu_Z, \sigma_Z)$$
$$= \left(\prod_{d=0}^{D-1} \pi_d^{\mathbb{I}(d=Z)}\right) \cdot \left(\frac{1}{\sigma_Z \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_Z^2} \cdot (X - \mu_Z)^2\right)\right)$$

(I) E-Step

$$\rho_z^{[i]} = p(Z=z \mid X=x^{[i]}; \underline{\theta}^{[k]})$$

we don't know the posterior (yet)



→ Apply Bayes' Rule

$$p(Z=z \mid X=x^{[i]}; \underline{\theta}^{[k]}) = \frac{p(X=x^{[i]} \mid Z=z; \underline{\theta}^{[k]})\, p(Z=z; \underline{\theta}^{[k]})}{p(X=x^{[i]}; \underline{\theta}^{[k]})}$$

something tough

→ work with proportional and normalize later on

$$\sim p(Z=z, X=x^{[i]}; \underline{\theta}^{[k]}) =: \tilde{\rho}_z^{[i]}$$

unnormalized responsibility

[Parameters $\underline{\theta} = [\underline{\pi}^T, \underline{\mu}^T, \underline{\sigma}^T]^T$]

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

$$\rho_z^{[i]} = \frac{\tilde{\rho}_z^{[i]}}{\sum_{d=0}^{D-1} \tilde{\rho}_d^{[i]}}$$ (row normalization)

(II) M-Step

$$Q(D; \underline{\theta}^{[k]}, \underline{\theta}) = \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \rho_c^{[i]} \cdot \log\left(p(Z=c, X=x^{[i]}; \underline{\theta})\right)$$

$$= \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \rho_c^{[i]} \cdot \left(\sum_{d=0}^{D-1} \mathbb{I}(d=c) \log(\pi_d) + \log\left(\frac{1}{\sigma_c \sqrt{2\pi}}\right) - \frac{1}{2\sigma_c^2}(x^{[i]} - \mu_c)^2\right)$$

$$\stackrel{+}{=} \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \rho_c^{[i]} \cdot \left(\sum_{d=0}^{D-1} \mathbb{I}(d=c) \log(\pi_d) - \log(\sigma_c) - \cancel{\frac{1}{2}\log(2\pi)} - \frac{1}{2\sigma_c^2}(x^{[i]} - \mu_c)^2\right)$$

$$\underline{\pi}^*, \underline{\mu}^*, \underline{\sigma}^* = \arg\max_{\substack{\underline{\pi} \in [0,1]^D \\ \underline{\mu} \in \mathbb{R}^D \\ \underline{\sigma} \in \mathbb{R}_+^D}} \left(Q(D; \underline{\theta}^{[k]}, \underline{\theta})\right)$$

constraint: $\sum_{d=0}^{D-1} \pi_d = 1$

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

$$\hat{Q}(D; \underline{\theta}^{[k]}, \underline{\theta}, \lambda) = Q(D; \underline{\theta}^{[k]}, \underline{\theta}) + \lambda g(\underline{\pi})$$

$$= \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \left(\rho_c^{[i]} \left(\sum_{d=0}^{D-1} \mathbb{I}(d=c) \log(\pi_d)\right) - \log(\sigma_c) - \frac{1}{2\sigma_c^2}(x^{[i]} - \mu_c)^2\right) + \lambda\left(1 - \sum_{d=0}^{D-1} \pi_d\right)$$

↳ now: unconstrained: take derivative and set to zero

(1) $\underline{\pi}$

π_e ... the e-th component of $\underline{\pi}$

$$\frac{\partial \hat{Q}}{\partial \pi_e} = \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \left(\rho_c^{[i]} \mathbb{I}(e=c) \frac{1}{\pi_e}\right) - \lambda \stackrel{!}{=} 0$$

$$= \frac{1}{\pi_e} \sum_{i=0}^{N-1} \sum_{c=0}^{D-1} \left(\rho_c^{[i]} \mathbb{I}(e=c)\right) - \lambda \stackrel{!}{=} 0$$

$$= \frac{1}{\pi_e} \underbrace{\sum_{i=0}^{N-1} \left(\rho_e^{[i]}\right)}_{\gamma_e} - \lambda \stackrel{!}{=} 0$$

γ_e ... the responsibility mass associated with the e-th class over all samples

$$\frac{\gamma_e}{\pi_e} - \lambda = 0 \quad \Leftrightarrow \quad \lambda = \frac{\gamma_e}{\pi_e}$$

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

[$\lambda = \frac{\gamma_e}{\pi_e} = \text{const.} = \frac{\gamma_{e-1}}{\pi_{e-1}} = \frac{\gamma_{e+1}}{\pi_{e+1}} = \frac{\gamma_d}{\pi_d}$]

how do get rid of the λ?

Shrick

$$\lambda = \lambda = \lambda \cdot 1 = \lambda \sum_{d=0}^{D-1} \pi_d = \sum_{d=0}^{D-1} \lambda \cdot \pi_d = \sum_{d=0}^{D-1} \frac{\gamma_d}{\cancel{\pi_d}} \cdot \cancel{\pi_d} = \sum_{d=0}^{D-1} \gamma_d = N$$

$$N = \frac{\gamma_e}{\pi_e}$$

$$\Leftrightarrow \boxed{\pi_e = \frac{\gamma_e}{N}}$$

(2) $\underline{\mu}$

μ_e ... e-th component $\underline{\mu}$

$$\frac{\partial \hat{Q}}{\partial \mu_e} = \sum_{i=0}^{N-1} \rho_e^{[i]} \cdot \left(\cancel{-} \frac{\cancel{2}(x^{[i]} - \mu_e) \cdot \cancel{(-1)}}{\cancel{2}\sigma_e^2}\right) \stackrel{!}{=} 0 \quad \Big| \cdot \sigma_e^2$$

$$= \sum_{i=0}^{N-1} \rho_e^{[i]} \cdot (x^{[i]} - \mu_e) \stackrel{!}{=} 0$$

$$= \sum_{i=0}^{N-1} \rho_e^{[i]} x^{[i]} - \underbrace{\sum_{i=0}^{N-1} \rho_e^{[i]}}_{\gamma_e} \mu_e \stackrel{!}{=} 0$$

$$= \sum_{i=0}^{N-1} \rho_e^{[i]} x^{[i]} - \gamma_e \mu_e \stackrel{!}{=} 0$$

$$\rightarrow \boxed{\mu_e = \frac{1}{\gamma_e} \sum_{i=0}^{N-1} \rho_e^{[i]} x^{[i]}}$$

(3) $\underline{\sigma}$

σ_e ... e-th component of $\underline{\sigma}$

$$\frac{\partial \hat{Q}}{\partial \sigma_e} = \sum_{i=0}^{N-1} \rho_e^{[i]} \cdot \left(-\frac{1}{\sigma_e} \stackrel{+}{\cancel{-}} \frac{\cancel{(-2)}}{\cancel{2}\sigma_e^3}(x^{[i]} - \mu_e)^2\right) \stackrel{!}{=} 0$$

$$= \sum_{i=0}^{N-1} \rho_e^{[i]} \cdot \left(-\frac{1}{\sigma_e} + \frac{1}{\sigma_e^3}(x^{[i]} - \mu_e)^2\right) \stackrel{!}{=} 0 \quad \Big| \cdot \sigma_e^3$$

$$= \sum_{i=0}^{N-1} \rho_e^{[i]} \cdot \left(-\sigma_e^2 + (x^{[i]} - \mu_e)^2\right) \stackrel{!}{=} 0$$

$$= \sum_{i=0}^{N-1} \rho_e^{[i]} (x^{[i]} - \mu_e)^2 - \underbrace{\sum_{i=0}^{N-1} \rho_e^{[i]}}_{\gamma_e} \sigma_e^2 \stackrel{!}{=} 0$$

$$\boxed{\sigma_e = +\sqrt{\frac{1}{\gamma_e} \sum_{i=0}^{N-1} \rho_e^{[i]} (x^{[i]} - \mu_e)^2}}$$

Summary: EM for the GMM

E-Step

(1) calculate unnormalized responsibilities

$$\tilde{\rho}_c^{[i]} = \pi_c \cdot \frac{1}{\sigma_c \sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2\sigma_c^2} \cdot (x^{[i]} - \mu_c)^2\right)$$

(2) normalize responsibilities:

$$\rho_c^{[i]} = \frac{\tilde{\rho}_c^{[i]}}{\sum_{d=0}^{D-1} \tilde{\rho}_d^{[i]}}$$

(3) calculate class responsibilities

$$\gamma_c = \sum_{i=0}^{N-1} \rho_c^{[i]}$$

M-Step

(1) update the class probabilities

$$\pi_c \leftarrow \frac{\gamma_c}{N}$$

(2) update the mus

$$\mu_c \leftarrow \frac{\sum_{i=0}^{N-1} \rho_c^{[i]} x^{[i]}}{\gamma_c}$$

(3) update the sigmas

$$\sigma_c \leftarrow \sqrt{\frac{1}{\gamma_c} \sum_{i=0}^{N-1} \rho_c^{[i]} (x^{[i]} - \mu_c)^2}$$