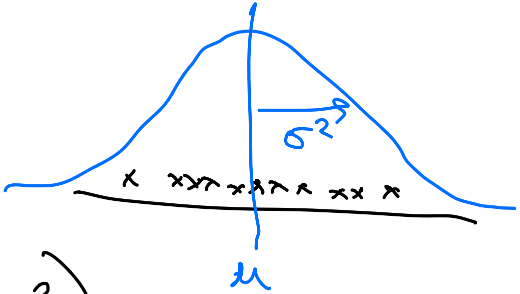
Maximum Likelihood



$$p(x \mid \theta) = \prod_{i=1}^{N} \mathcal{N}(x_i \mid \mu, \sigma^2)$$

$$\mathcal{L}(\underbrace{\theta}_{\sigma^2,\,\mu}) = \log p(x \mid \theta) = \sum_{i=1}^{N} \log \mathcal{N}(x_i \mid \mu, \sigma^2)$$

$$= -\frac{1}{2}\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$\frac{\partial}{\partial \mu}\mathcal{L} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i - \mu) = 0 \Longrightarrow \mu = \frac{1}{N}\sum x_i$$

$$\frac{\partial}{\partial \sigma^2}\mathcal{L} = \cdots \Longrightarrow \sigma^2 = \frac{1}{N}\sum (x_i - \mu)^2$$

Mixture model?

$$p(x|\theta,\pi) = \sum_k \pi_k \, p(x|\theta_k)$$

$$p(x) = \sum_{z \in Z} p(z)\,p(x|z) = \sum_k p(z=k)\,p(x|z=k,\theta_k)$$

$$\log p(x) = \sum_{i=1}^{N} \log \underbrace{p(x_i)}_{} = \sum_{i=1}^{N} \log \underbrace{\sum_k p(z_i=k|\pi)\,p(x_i|z_i=k,\theta_k)}_{}$$

x: obs.

z: „latent"

„complete data"

$$\log p(x,z) = \sum_{i=1}^{N} \overbrace{\log p(z_i)}^{\text{discrete} \to \log \pi_{ik}} + \overbrace{\log p(x_i|z_i)}^{\to \text{gaussian}}$$

$z \sim \text{Discrete}(\pi)$ ← $z \in \{1, \dots, K\} \equiv \mathcal{Z}$

$x | z \sim p(x | \theta_z)$

„Expectation Maximization" (EM) is M.L. in latent variable models.

$$\textcircled{1} \quad \underline{\log p(X|\theta)} = \mathcal{L}(q, \theta) + \overbrace{KL(q\|p)}^{\geq 0} \geq \underbrace{\mathcal{L}(q, \theta)}_{\text{lower bound on } \log p(X|\theta)}$$

$$\boxed{KL(q\|p)} = \sum_{\mathcal{Z}} q(z) \log \left[\frac{q(z)}{p(z|x, \theta)} \right]$$

↳ $\geq 0 \ \forall q$

↳ $= 0 \Leftrightarrow q = p$

$$\mathcal{L}(q, \theta) = \sum_{\mathcal{Z}} q(z) \log \left[\frac{p(X, Z|\theta)}{q(z)} \right]$$

↳ computable! ✓

verify ①!

only depend on $p(X, Z|\theta)$

$$\log p(x|\theta) = \underbrace{\mathcal{L}(q,\theta)}_{\text{max w.r.t. } \theta} + \underbrace{KL(q\|p)}_{\text{min w.r.t. } q}$$

init θ

$$\max_{\theta} \mathcal{L}(q,\theta) = \max_{\theta} \sum_{z} \underbrace{q(z)\log p(x,z|\theta)}_{\text{maximize}}$$

2

$$KL(q\|p) = 0 \iff q(z) = \underbrace{p(z|x,\theta)}_{\text{compute posterior}}$$

$$\mathcal{L} = \sum q(z) \log\left[\frac{q(z)}{p(z|x,\theta)}\right]$$

if $q(z)=p(z|x,\theta) \Rightarrow KL=0.$

2

$$Q(\theta, \theta^{old}) = \sum_{z} \underbrace{p(z \mid x, \theta^{old})}_{= q(z)} \log p(x, z \mid \theta)$$

est

$$\max_{\theta} \; Q(\theta, \theta^{old})$$

max

EM for GMM.

$$\Theta^{init} = \{\mu_k, \Sigma_k\}$$

$$p(z_n = k \mid x_n, \Theta^{init}) = \frac{p(z_n = k)\, p(x_n \mid z = k, \Theta_k^{init})}{\sum_{j=1}^{K} p(z_n = j)\, p(x_n \mid z_n = j, \Theta_j^{init})}$$

$$\underbrace{\gamma(z_{nk})}_{\text{„responsibilities"}} = \frac{\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}$$

$$Q(\Theta, \Theta^{old}) = \sum_{n=1}^{N} \sum_{k=1}^{K} \underbrace{\gamma(z_{nk})}_{\text{computed scalar}} \boxed{\log p(x_n, z_n | \Theta)}$$

$$\overline{Z}$$

$$N_k = \underbrace{\sum_{n=1} \gamma(z_{nk})}_{\text{„soft" cluster size}}$$

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n$$

$$\pi_k^{new} = \frac{N_k}{N}$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$