The EM algorithm

The goal is to maximize the likelihood function

$$p(X|\theta) = \sum_{Z} p(X,Z|\theta)$$

If we introduce a distribution q(Z) defined over the latent variables, for any choice of q(Z) we have the following decomposition:

$$ln(p(X|\theta)) = \mathcal{L}(q, \theta) + KL(q||p)$$

Where

$$\mathcal{L}(q, heta) = \sum_{Z} q(Z) ln\{rac{p(X, Z| heta)}{q(Z)}\}$$

$$KL(q||p) = -\sum_{Z} q(Z) ln\{rac{p(Z|X, heta)}{q(Z)}\}$$

Proof

$$egin{aligned} p(X,Z| heta) &= p(Z|X, heta)p(X| heta) \ & \mathcal{L}(q, heta) &= \sum_{Z} q(Z)ln\{rac{p(Z|X, heta)p(X| heta)}{q(Z)}\} = \ &= \sum_{Z} q(Z)(ln\{rac{p(Z|X, heta)}{q(Z)}\} + p(X| heta)) = \ &= -KL(q,p) + p(X| heta)\sum_{Z} q(Z) = -KL(q,p) + p(X| heta) \end{aligned}$$