# 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

$$\mathcal{L}(q, \theta) = \sum_Z q(Z) ln\{\frac{p(X, Z|\theta)}{q(Z)}\}$$

$$KL(q||p) = -\sum_Z q(Z) ln\{\frac{p(Z|X, \theta)}{q(Z)}\}$$

Proof

$$p(X, Z|\theta) = p(Z|X, \theta)p(X|\theta)$$

$$\mathcal{L}(q, \theta) = \sum_Z q(Z) ln\{\frac{p(Z|X, \theta)p(X|\theta)}{q(Z)}\} =$$

$$= \sum_Z q(Z)(ln\{\frac{p(Z|X, \theta)}{q(Z)}\} + p(X|\theta)) =$$

$$= -KL(q, p) + p(X|\theta) \sum_z q(Z) = -KL(q, p) + p(X|\theta)$$