A General Introduction to EM

Data: X (observed) + H(hidden) Parameter: θ

"Incomplete" likelihood: L(θ)= log p(X| θ)
"Complete" likelihood: Lc(θ)= log p(X,H| θ)

EM tries to iteratively maximize the incomplete likelihood:
Starting with an initial guess $\theta^{(0)}$,

1. E-step: compute the expectation of the complete likelihood

$$Q(\theta;\theta^{(n-1)}) = E_{\theta^{(n-1)}}[L_c(\theta) \mid X] = \sum_{h_i} p(H = h_i \mid X, \theta^{(n-1)}) \log P(X, h_i)$$

2. M-step: compute $\theta^{(n)}$ by maximizing the Q-function

$$\theta^{(n)} = \arg\max_\theta Q(\theta;\theta^{(n-1)}) = \arg\max_\theta \sum_{h_i} p(H = h_i \mid X, \theta^{(n-1)}) \log P(X, h_i)$$