Expectation Maximization: Derivation

- Consider general "imputation distributions" $\rho(z \mid x)$.
- We can construct a **tight lower bound** $LB(\theta, \rho)$ of the marginal loglikelihood function:

$$L(\theta) \geq LB(\theta,\ \rho), \quad \forall \rho, \qquad \text{and} \qquad L(\theta) = \max_{\rho} LB(\theta,\ \rho).$$

- Maximizing $L(\theta)$ is then equivalent to maximizing $LB(\theta, \rho)$.

$$\max_{\theta} L(\theta) = \max_{\theta,\rho} LB(\theta,\ \rho).$$

- Optimizing θ and ρ alternatively (coordinate descent) yields EM algorithm.

$$\max_{\theta} \sum_{i=1}^{n} \log P(x_i \mid \theta) \triangleq L(\theta)$$

$$L(\theta) = \sum_{i=1}^{n} \log \sum_{z_i} P(x_i\ z_i \mid \theta)$$

$$= \sum_{i=1}^{n} \log \sum_{z_i} \left(\frac{P(x_i, z_i \mid \theta)}{\rho(z_i \mid x_i)} \rho(z_i \mid x_i) \right)$$

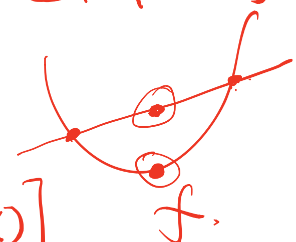
$$= \sum_{i=1}^{n} \log E_{z_i \sim \rho(\cdot \mid x_i)} \left[\frac{P(x_i\ z_i \mid \theta)}{\rho(z_i \mid x_i)} \right]$$

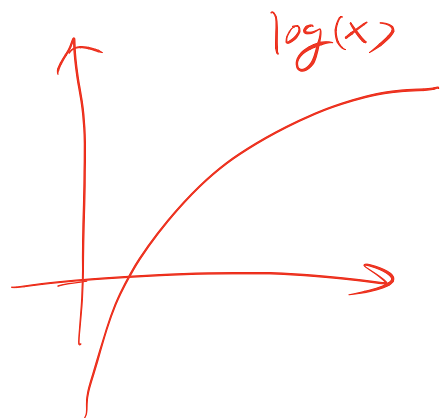
$$\geq \boxed{\sum_{i=1}^{n} E_{z_i \sim \rho(\cdot \mid x_i)} \log \left(\frac{P(x_i\ z_i \mid \theta)}{\rho(z_i \mid x_i)} \right) \triangleq LB(\theta,\ \rho)} \quad \forall \rho$$

Jensen's Inequality

f : convex

$f(E[X]) \leq E[f(X)]$





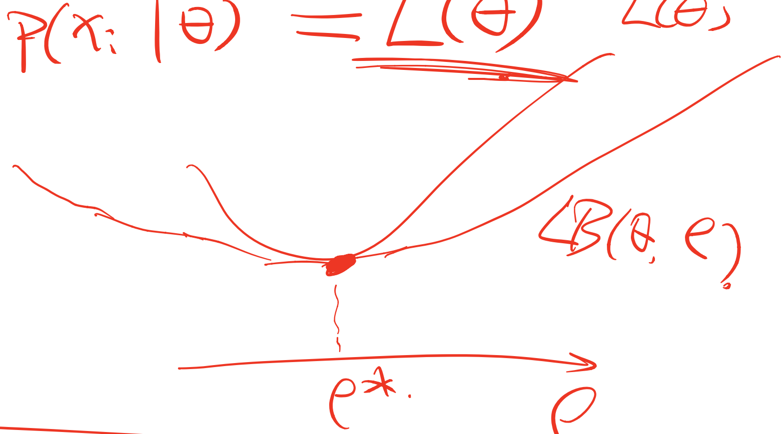
$-\log(x)$: convex

If $q^*(z|x,\theta) = P(z|x,\theta)$

$$\frac{P(x_i, z_i|\theta)}{q^*(z_i|x_i)} = \frac{P(x_i, z_i|\theta)}{P(z_i|x_i,\theta)} = P(x_i|\theta)$$

$$\Rightarrow LB(\theta, q^*) = \sum_{i=1}^{n} E_{z_i \sim q(\cdot|x_i)} \log P(x_i|\theta)$$

$$= \sum_{i=1}^{n} \log P(x_i|\theta) = L(\theta)$$



$\Rightarrow$

$$L(\theta) = \max_{q} LB(\theta, q)$$

$$\Rightarrow \max_{\theta, q} LB(\theta, q)$$

Initialize θ_0

Fix θ_t $\quad \boxed{q_{t+1} = \arg\max_{q} LB(\theta_t, q)}$ $\rightarrow P(z|x, \theta_t)$ E-step

Fix q_{t+1} $\quad \theta_{t+1} = \arg\max_{\theta} LB(\theta, q_{t+1})$ M-step

$$LB(\theta, q_{t+1}) = \sum_{i=1}^{n} E_{z_i \sim q(\cdot|x_i)} \left[\log P(x_i, z_i|\theta) - \log q(z_i|x_i)\right]$$

Const.

$$= \sum_{i=1}^{n} E_{z_i \sim Q(\cdot|x_i)}[\log P(x_i, z_i | \theta)] + \text{Const}$$

Expected joint likelihood

① Monotonically decrease $LB(\theta, Q)$

$L(\theta)$

② Converge to local optimal of $L(\theta)$