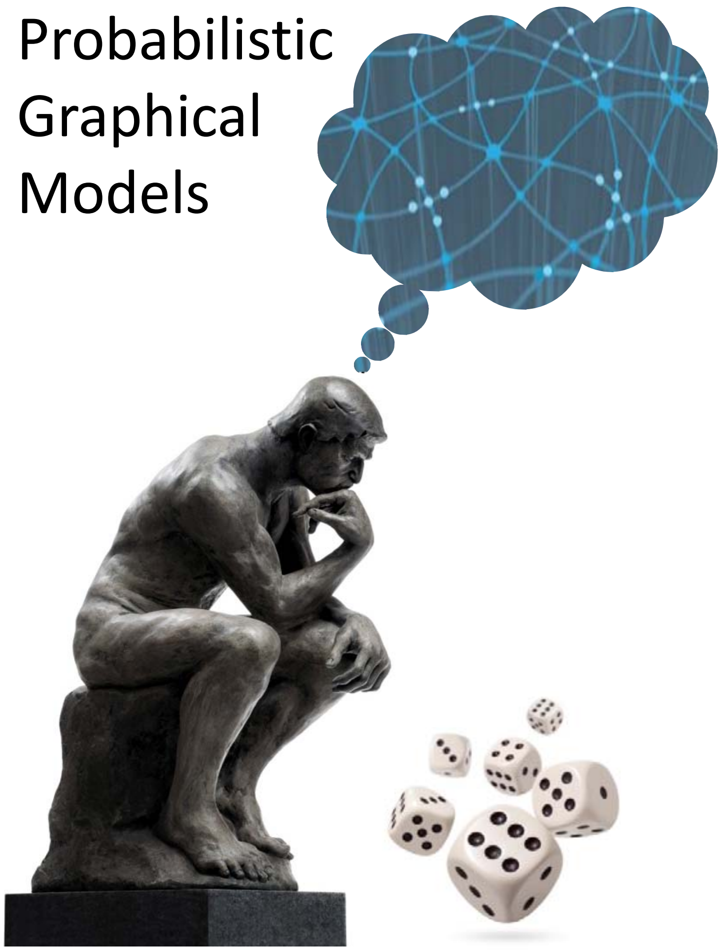
Probabilistic
Graphical
Models

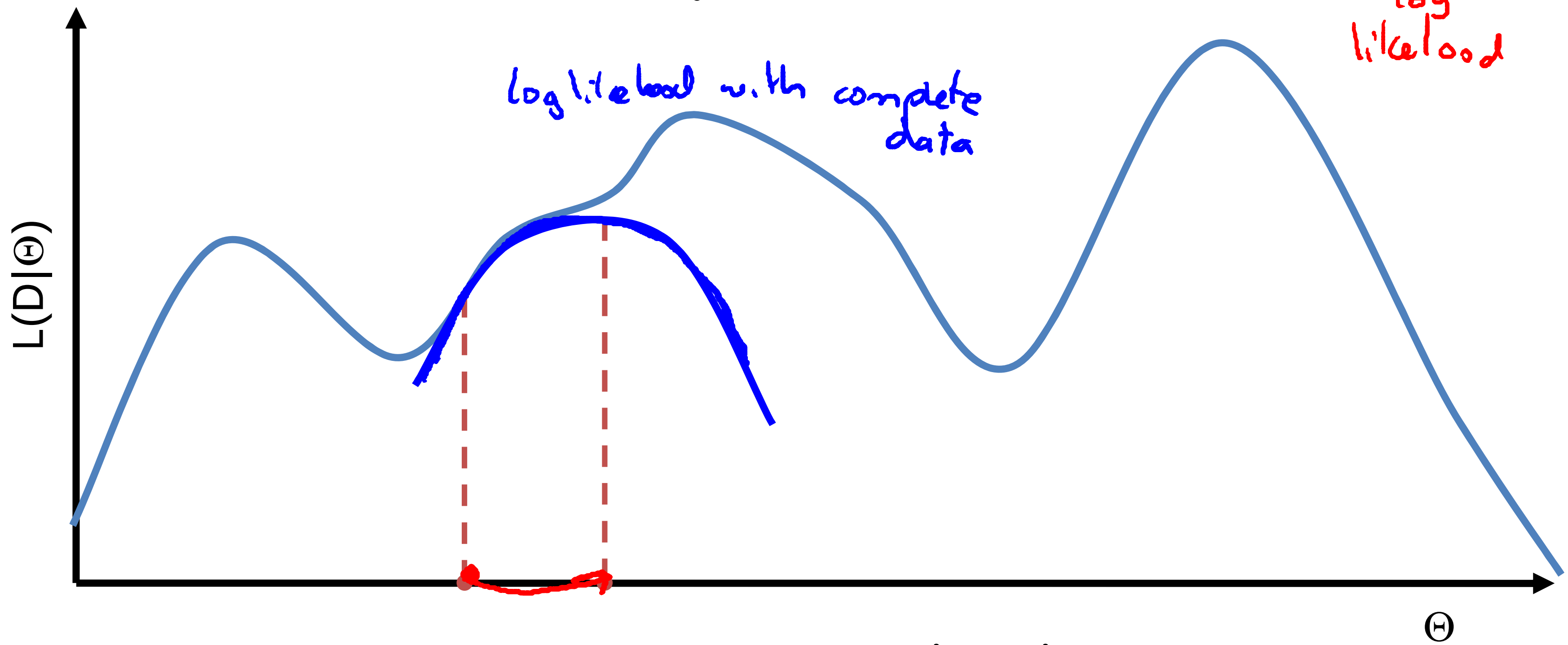


Learning

Incomplete Data

EM Analysis

More Formal Intuition



- Use current point to construct local approximation
- Maximize new function in closed form

Daphne Koller

More Formal Intuition

- **d**: observed data in instance
- **H**: hidden variables in instance
- Q(H): distribution over hidden variables

assignment to H

log parameter

$$\ell(\boldsymbol{\theta} : \langle \boldsymbol{d}, \boldsymbol{h} \rangle) = \sum_{i=1}^{n} \sum_{(\boldsymbol{x}_i, \boldsymbol{u}_i) \in Val(X_i, \mathbf{Pa}_{X_i})} \mathbf{1}_{\langle \boldsymbol{d}, \boldsymbol{h} \rangle}[\boldsymbol{x}_i, \boldsymbol{u}_i] \log \theta_{\boldsymbol{x}_i | \boldsymbol{u}_i}$$

$$\boldsymbol{E}_{Q(\boldsymbol{H})}[\ell(\boldsymbol{\theta} : \langle \boldsymbol{d}, \boldsymbol{H} \rangle)] = \sum_{i=1}^{n} \sum_{(x_i, \boldsymbol{u}_i) \in Val(X_i, \mathbf{Pa}_{X_i})} \boldsymbol{E}_{Q(\boldsymbol{H})}\left[\mathbf{1}_{\langle \boldsymbol{d}, \boldsymbol{H} \rangle}[x_i, \boldsymbol{u}_i]\right] \log \theta_{x_i | \boldsymbol{u}_i}$$

$$= \sum_{i=1}^{n} \sum_{(\boldsymbol{x}_i, \boldsymbol{u}_i) \in Val(X_i, \mathbf{Pa}_{X_i})} Q(\boldsymbol{x}_i, \boldsymbol{u}_i) \log \theta_{\boldsymbol{x}_i | \boldsymbol{u}_i}$$

Daphne Koller

More Formal Intuition

$$\boldsymbol{E}_{Q(\boldsymbol{H})}[\ell(\boldsymbol{\theta} : \langle \boldsymbol{d}, \boldsymbol{H} \rangle)] = \sum_{i=1}^{n} \sum_{(x_i, \boldsymbol{u}_i)} Q(x_i, \boldsymbol{u}_i) \log \theta_{x_i | \boldsymbol{u}_i}$$

$$Q^t_m(\boldsymbol{H}[m]) = P(\boldsymbol{H}[m] \mid \boldsymbol{d}[m], \boldsymbol{\theta}^t)$$

$$\sum_{m=1}^{M} \boldsymbol{E}_{Q^t_m(\boldsymbol{H}[m])}[\ell(\boldsymbol{\theta} : \langle \boldsymbol{d}[m], \boldsymbol{H}[m] \rangle)]$$

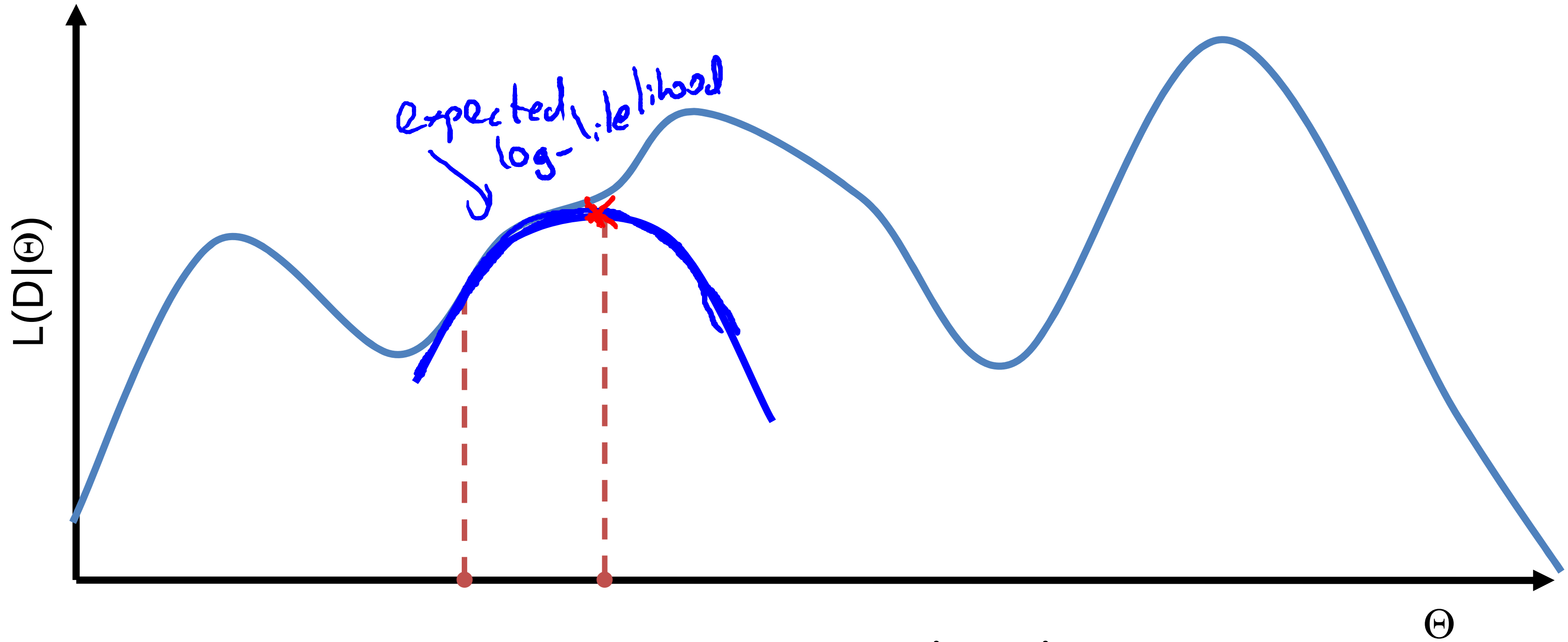
$$= \sum_{i=1}^{n} \sum_{(x_i, \boldsymbol{u}_i)} \left(\sum_{m=1}^{M} P(x_i, \boldsymbol{u}_i \mid \boldsymbol{d}[m], \boldsymbol{\theta}^t) \right) \log \theta_{x_i | \boldsymbol{u}_i}$$

expected suff stats

$$= \sum_{i=1}^{n} \sum_{(x_i, \boldsymbol{u}_i)} \underbrace{\bar{M}_{\boldsymbol{\theta}^t}[x_i, \boldsymbol{u}_i]}_{\text{ESS}} \log \theta_{x_i | \boldsymbol{u}_i}$$

log likelihood for complete data using ESS

More Formal Intuition



- Use current point to construct local approximation
- Maximize new function in closed form

Daphne Koller

EM Guarantees

- $L(D : \theta^{t+1}) \geq L(D : \theta^{t})$
  - Each iteration improves the likelihood

gradient is zero

- If $\theta^{t+1} = \theta^{t}$, then θ^{t} is a stationary point of $L(D : \theta)$
  - Usually, this means a local maximum