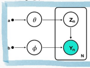
MODEL



$$\phi \sim p(\phi|b),$$
$$\theta \sim p(\theta|a),$$
$$Z_n|\theta \sim p(Z|\theta),$$
$$Y_n|Z_n,\phi \sim p(Y|Z,\phi),\; n = 1,\dots,N$$

Let $\psi = [\phi, \theta, Z_1 \dots Z_N$

Variational family $Q = \{q(\psi|\lambda) \mid \lambda \in \Delta\}$

Goal: $\lambda^* = \arg\min_{\lambda} D_{KL}[p(\psi|\text{Data}) \,\|\, q(\psi|\lambda)]$

Problem: we don't know $q(\psi|\lambda)$ and we can't evaluate it!

Solution: instead of min. D_{KL} we max. ELBO

$$\arg\max_{\lambda} \underbrace{\mathbb{E}_{\psi \sim q(\psi|\lambda)}\left[\log\left(\frac{p(\psi, \text{Data})}{q(\psi|\lambda)}\right)\right]}_{\text{ELBO}}$$

Why? ELBO involves the joint:

$p(\psi, \text{Data}) = p(\text{Data}|\psi)\, p(\psi)$

The joint is tractable while posterior is not.

Assumptions: mean-field

$$q(\psi|\lambda) = \prod_{i=1} q(\psi_i|\lambda_i)$$

How: right now we can only optimize fns analytically, so we do coordinate-ascent on ELBO, because this is easier!

iterate:

$$\max_{\lambda_i} \text{ELBO}(\psi_i|\psi_{-i})$$

Claim: $\mathbb{E}_{\psi \sim q(\psi|\lambda)}[\cdots] = \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}\left[\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}[\cdots]\right]$

$$\begin{aligned}
\mathbb{E}_{\psi \sim q(\psi|\lambda)}\left[\log\left(\frac{p(\psi|\text{Data})}{q(\psi|\lambda)}\right)\right] &= \int_{\psi}\left[\log\left(\frac{p(\psi, \text{Data})}{q(\psi|\lambda)}\right)\right] q(\psi|\lambda)\, d\psi \\
&= \int_{\psi_i}\int_{\psi_{-i}}\left[\log\left(\frac{p(\psi, \text{Data})}{q(\psi|\lambda)}\right)\right] q(\psi|\lambda)\, d\psi_i\, d\psi_{-i} \\
&= \int_{\psi_i}\int_{\psi_{-i}}\left[\log\left(\frac{p(\psi, \text{Data})}{q(\psi|\lambda)}\right)\right] q(\psi|\lambda)\, d\psi_i\, d\psi_{-i} \\
&= \int_{\psi_i}\int_{\psi_{-i}}\left[\log\left(\frac{p(\psi, \text{Data})}{q(\psi|\lambda)}\right)\right] q(\psi_i|\lambda_i)\, q(\psi_{-i}|\lambda_{-i})\, d\psi_{-i}\, d\psi_i \\
&= \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}\left[\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}\left[\log\left(\frac{p(\psi, \text{Data})}{q(\psi|\lambda)}\right)\right]\right]
\end{aligned}$$

In fact, with a bit more algebra, we can show that:

$$\max_{\lambda_i} \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}\left[\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}\left[\log\left(\frac{p(\psi, \text{Data})}{q(\psi|\lambda)}\right)\right]\right] \equiv \max_{\lambda_i} \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}\left[\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}\left[\log\left(\frac{p(\psi_i, \text{Data}|\psi_{-i})}{q(\psi|\lambda)}\right)\right]\right]$$

Claim: $\max_{\lambda_i} \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}[\cdots] \equiv \min_{\lambda_i} D_{KL}[\cdots]$

$$\begin{aligned}
\max_{\lambda_i} \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}\left[\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}\left[\log\left(\frac{p(\psi_i, \text{Data}|\psi_{-i})}{q(\psi|\lambda)}\right)\right]\right] &\overset{\text{algebra}}{=} \max_{\lambda_i} \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}\left[\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}\left[\log p(\psi_i, \text{Data}|\psi_{-i})\right] - \log q(\psi_i|\lambda_i)\right] \\
&= \max_{\lambda_i} \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}\left[\log\left(\exp\left\{\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}\left[\log p(\psi_i, \text{Data}|\psi_{-i})\right] - \log q(\psi_i|\lambda_i)\right\}\right)\right] \\
&\overset{\text{algebra}}{\equiv} \underbrace{\min_{\lambda_i} \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}\left[\log\left(\frac{q(\psi_i|\lambda_i)}{\exp\{\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}[\log p(\psi_i, \text{Data}|\psi_{-i})]\}}\right)\right]}_{\text{looks like } D_{KL} \text{ but denominator is not a pdf!}} \\
&= \underbrace{\min_{\lambda_i} \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}\left[\log\left(\frac{Z\, q(\psi_i|\lambda_i)}{Z\exp\{\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}[\log p(\psi_i, \text{Data}|\psi_{-i})]\}}\right)\right]}_{\text{make denominator into pdf with normalizing constant } Z!} \\
&= \min_{\lambda_i} \mathbb{E}_{\psi_i \sim q(\psi_i|\lambda_i)}\left[\log\left(\frac{q(\psi_i|\lambda_i)}{Z\exp\{\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}[\log p(\psi_i, \text{Data}|\psi_{-i})]\}}\right)\right] + \log(Z) \\
&\equiv \min_{\lambda_i} D_{KL}\left[q(\psi_i|\lambda_i) \,\|\, Z\exp\{\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}[\log p(\psi_i, \text{Data}|\psi_{-i})]\}\right]
\end{aligned}$$

Lesson: we solve for $\lambda^* = \arg\max_{\lambda} \underbrace{\mathbb{E}_{\psi \sim q(\psi|\lambda)}\left[\log\left(\frac{p(\psi, \text{Data})}{q(\psi|\lambda)}\right)\right]}_{\text{ELBO}}$ by coordinate ascent

each time updating λ_i by

$$\lambda_i^* = \arg\max_{\lambda_i} \text{ELBO}(\lambda_i, \lambda_{-i})$$

$$q(\psi_i|\lambda_i^*) \propto \exp\left\{\mathbb{E}_{\psi_{-i} \sim q(\psi_{-i}|\lambda_{-i})}\left[\log p(\psi_i, \text{Data}|\psi_{-i})\right]\right\}$$