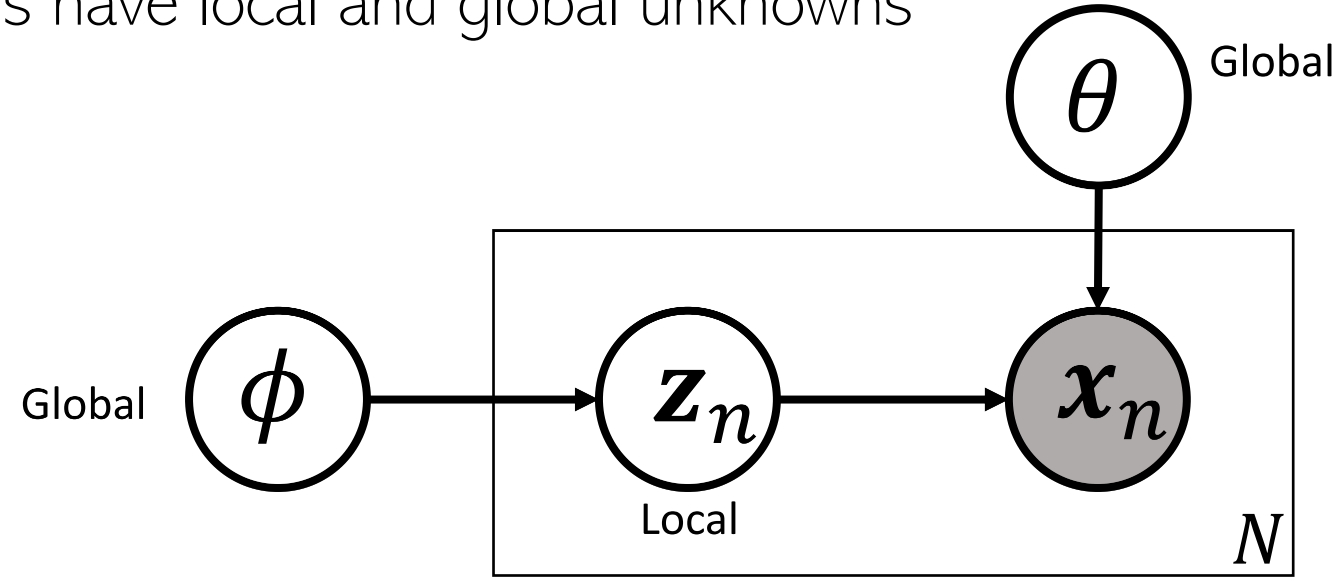
Mean-Field VI for Models with Local and Global Variables

CS698X: Topics in Probabilistic Modeling and Inference

Piyush Rai

LVMs with Local and Global Unknowns

- Many LVMs have local and global unknowns

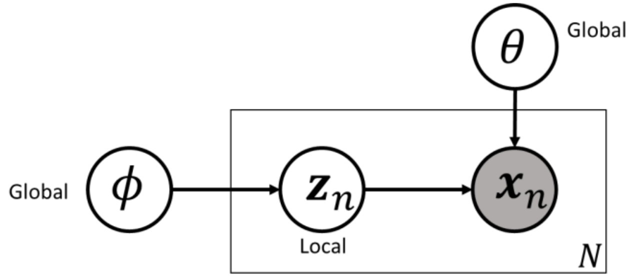


- Examples: Gaussian Mixture Model, Prob. PCA, Variational Autoencoder (VAE), etc
- Denote all local unknowns $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_N\}$ as $\boldsymbol{Z}$ and global unknown as $\boldsymbol{\beta} = (\theta, \phi)$
- The goal is to infer the posterior $p(\boldsymbol{Z}, \boldsymbol{\beta} | \boldsymbol{X})$ which is intractable in general
- Mean-field VI will approximating this posterior as $p(\boldsymbol{Z}, \boldsymbol{\beta} | \boldsymbol{X}) \approx q(\boldsymbol{Z}, \boldsymbol{\beta}) \approx q(\boldsymbol{Z}) q(\boldsymbol{\beta})$

LVMs with Local and Global Unknowns

- Assuming independence, the joint distribution of data $\boldsymbol{X}$ and unknowns $\boldsymbol{\beta} = (\theta, \phi)$

$$p(\mathbf{X}, \mathbf{Z}, \beta) = p(\beta) \prod_{n=1}^{N} p(\mathbf{x}_n | \mathbf{z}_n, \beta) p(\mathbf{z}_n | \beta) = p(\beta) \prod_{n=1}^{N} p(\mathbf{x}_n, \mathbf{z}_n | \beta)$$



- Assume the joint dist. of $\boldsymbol{x}_n$ and $\boldsymbol{z}_n$ to be an exp-fam dist with natural params $\boldsymbol{\beta}$

Sufficient statistics

$$p(\mathbf{x}_n, \mathbf{z}_n | \beta) = h(\mathbf{x}_n, \mathbf{z}_n) \exp \left[\beta^\top t(\mathbf{x}_n, \mathbf{z}_n) - A(\beta) \right]$$

- Assume a prior on $\boldsymbol{\beta}$, that is conjugate to the above exp-fam dist

$$p(\beta | \boldsymbol{\alpha}) = h(\beta) \exp \left[\boldsymbol{\alpha}^\top [\beta, -A(\beta)] - A(\boldsymbol{\alpha}) \right]$$

where $\alpha = [\alpha_1, \alpha_2]^\top$ are the the hyperparamers of the prior $p(\beta)$ and $[\beta, -A(\beta)]$ is the sufficient statistics vector for this exp-family distribution

Mean-Field VI for LVMs with Local and Global Vars

- Recall that mean-field VB can be obtained using CP of each unknown

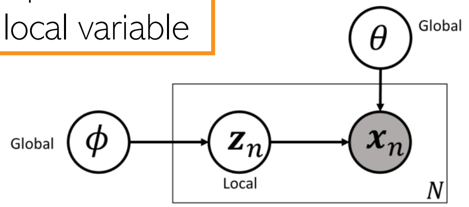
- Optimal variational distribution for each unknown requires expec. of the log of its CP

$$p(\boldsymbol{\beta}|\boldsymbol{\alpha}) = h(\boldsymbol{\beta}) \exp\left[\boldsymbol{\alpha}^\top[\boldsymbol{\beta}, -A(\boldsymbol{\beta})] - A(\boldsymbol{\alpha})\right]$$

- Due to conj, CP of global vars $\boldsymbol{\beta} = (\boldsymbol{\theta}, \boldsymbol{\phi})$, will be in same distr as prior $\boldsymbol{p(\beta|\alpha)}$. Thus

Updates to the natural parameters requires a summing suff-stats over all data and local variable

$$p(\boldsymbol{\beta}|\mathbf{X}, \mathbf{Z}) = p(\boldsymbol{\beta}|\hat{\boldsymbol{\alpha}}) \quad \text{where} \quad \hat{\boldsymbol{\alpha}} = \left[\alpha_1 + \sum_{n=1}^{N} t(\mathbf{x}_n, \mathbf{z}_n), \alpha_2 + N\right]$$



- Likewise, CP of each local variable $\boldsymbol{z_n}$

Due to the independence structure

Assuming CP is an exp-fam distribution (will be the case if the prior $p(z_n|\phi)$ and likelihood $p(x_n|z_n, \theta)$ are exp-family and conjugate to each other)

$$p(\mathbf{z}_n|\mathbf{Z}_{-n}, \mathbf{X}, \boldsymbol{\beta}) = p(\mathbf{z}_n|\mathbf{x}_n, \boldsymbol{\beta}) = h(\mathbf{z}_n) \exp\left[\eta(\mathbf{x}_n, \boldsymbol{\beta})^\top \mathbf{z}_n - A(\eta(\mathbf{x}_n, \boldsymbol{\beta}))\right]$$

Nat. params depends on data x_n and global var β

- Having these CPs, we can compute the mean-field updates for $\boldsymbol{q(\beta)}$ and $\boldsymbol{q(z_n)}$

Mean-Field VI for LVMs with Local and Global Vars

- Let's assume our mean-field approximation to be of the form

$$q(\boldsymbol{\beta}, \mathbf{Z}) = q(\boldsymbol{\beta}|\lambda) \prod_{n=1}^{N} q(\boldsymbol{z}_n|\phi_n)$$

- CPs are exp-fam, so optimal $\boldsymbol{q}$'s depend on expected suff-stats of CP's nat. params

- The optimal variational dist. for local vars $\boldsymbol{z}_n$ will be $\boldsymbol{q}(\boldsymbol{z}_n|\boldsymbol{\phi}_n)$ with

Basically requires expectation over the $q(\beta|\lambda)$ distribution

$$\phi_n = \mathbb{E}_\lambda\left[\eta(\boldsymbol{x}_n, \boldsymbol{\beta})\right] \qquad \forall n$$

- The optimal variational dist. for global vars $\boldsymbol{\beta}$ will be $\boldsymbol{q}(\boldsymbol{\beta}|\boldsymbol{\lambda})$ with

Basically requires expectation over the $q(z_n|\phi_n)$ distribution

$$\lambda = \left[\alpha_1 + \sum_{n=1}^{N} \mathbb{E}_{\phi_n}[\boldsymbol{t}(\boldsymbol{x}_n, \boldsymbol{z}_n)], \alpha_2 + N\right]^\top$$

- Mean-Field updates alternate between estimating $\boldsymbol{\phi}_n$'s and $\boldsymbol{\lambda}$ until convergence
- Potential bottleneck: Updating $\boldsymbol{\lambda}$ requires waiting for all $\boldsymbol{\phi}_n$'s to be updated (thus slow)
  - Can be handled using online VI (stochastic VI); will see this later

Coming Up Next

- Some properties of VI