VEM + VAE

Wednesday, November 9, 2022 12:45 PM

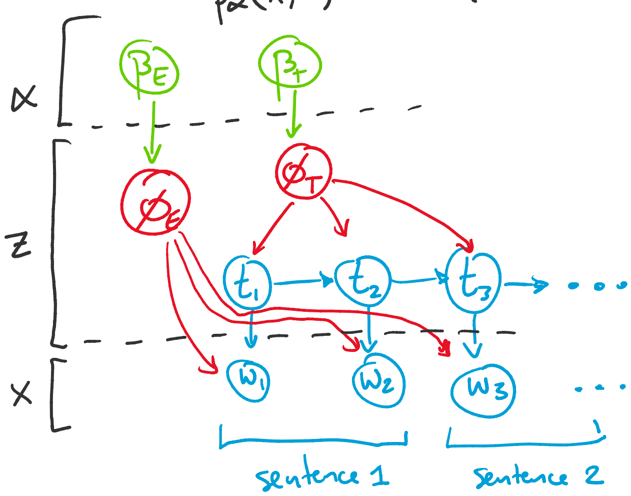
Ex: Unsupervised POS Tagging

Given: sentences only (concatenated together)

Goal: infer tags for unlabeled sentences

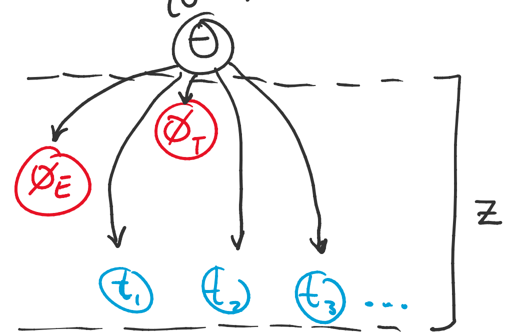
Model: Bayesian HMM

$p_\alpha(x,z)$



Mean Field Approx

$q_\theta(z)$



Training Objective

$$\log p(x|\alpha) = \log \sum_z p(x,z|\alpha) = \log \overbrace{\sum_{\vec t} \int_{\phi_E}\int_{\phi_T}}^{\Sigma_z} p(\overbrace{\vec w}^{x}, \overbrace{\vec t, \phi_E, \phi_T}^{z} \mid \overbrace{\beta_E, \beta_T}^{\alpha})$$

Using the ELBO:

$$\log p(x|\alpha) \geq ELBO(q_\theta) = E_{q_\theta}[\log p_\alpha(x,z)] - E_{q_\theta}[\log q_\theta(z)]$$

$$= E_{q_\theta}[\log p(w,t,\phi_E,\phi_T|\beta_E,\beta_T)] - E_{q_\theta}[\log q_\theta(t,\phi_E,\phi_T)]$$

Variational inference gives:

$$q_\theta(z) \approx p_\alpha(z|x) \Rightarrow q_\theta(\vec t) \approx p(\vec t|\beta) \quad q_\theta(\phi_E) \approx p(\phi_E|\beta) \quad q_\theta(\phi_T) \approx p(\phi_T|\beta)$$

Problem: How to estimate α?

☆New Idea☆: Jointly find p_α and q_θ to make ELBO as large as possible.

Two Approximations:

(1) Approximate Learning: choosing α for p / can't compute by assumption

Two Approximations:

① Approximate Learning: choosing α for p

- really want to maximize $\log p(x|\alpha)$ — can't compute by assumption
- instead maximize a variational lower bound

② Approximate Inference: choosing θ for q

- after reaching a (local) maximum of (p_α, q_θ) query q_θ approximately about z

  b/c directly querying p_α about z is intractable

Key Idea: jointly optimize ELBO as a function of both α and θ

$\boxed{\text{Variational EM}}$

Apply Block Coordinate Ascent to ELBO

Algo:

while not converged:

① Variational E-step:

adjust q_θ given current p_α

i.e. run V.I. to minimize $KL(q_\theta \| p_\alpha)$ for θ only

maximize $ELBO(q_\theta ; p_\alpha)$

$$\theta = \underset{\theta}{\operatorname{argmax}}\ ELBO(q_\theta ; p_\alpha) = \text{run-variational-inf}(q_\theta, p_\alpha)$$

② Variational M-step:

adjust p_α given current q_θ

i.e. improve $ELBO(q_\theta ; p_\alpha)$ for α only

only one term involves α: $E_{q_\theta}[\log p_\alpha(x,z)]$

$$\alpha = \underset{\alpha}{\operatorname{argmax}}\ ELBO(q_\theta ; p_\alpha) = \underset{\alpha}{\operatorname{argmax}}\ E_{q_\theta}[\log p_\alpha(x,z)]$$

Key Difference between Standard EM and Variational EM

We say that the "variational gap" in Standard EM is zero.

i.e. $q_\theta(z)$ perfectly estimates $p_\alpha(z|x)$

Standard EM is zero.

i.e. $q_\theta(z)$ perfectly estimates $p_\alpha(z|x)$

P