Variational Cavity Learning for Fed-GVI

Arqam Patel, supervised by Theo Damoulas

2024

1 Introduction

The cavity distribution $q_{\setminus m}^{(i)}$ of a client m in federated variational learning serves as an estimate of the counterfactual posterior if it were computed using only data from the other clients except m.

It serves as a prior for the variational computation of the local posterior at m:

$$q_m^{(i)} = \arg\min_{q \in \mathcal{Q}} \left\{ \mathbb{E}_q \left[L_m(\theta) \right] + D_{KL}[q || q_{\setminus m}^{(i)}] \right\}$$

$$q_m^{(i)} = \arg\min_{q \in \mathcal{Q}} \left\{ \mathbb{E}_q \left[L_m(\theta) + l_{\setminus m}(\theta) \right] + D_{KL}[q || \pi_0] \right\}$$

1.1 Cavity distribution in Partitioned Variational Inference

In PVI, we formulate the approximate posterior as the product of prior and approximate likelihoods

$$p(\theta|y) = \frac{1}{Z} \pi(\theta) \prod_{m=1}^{M} f(y_m|\theta) \approx \frac{1}{Z'} \pi(\theta) \prod_{m=1}^{M} t_m(\theta) = q_s(\theta)$$

At the client, we compute the mth cavity distribution by simply dividing out the mth approximate likelihood term:

$$q_{\setminus m}^{(i+1)}(\theta) \propto \pi(\theta) \prod_{k \neq m} t_k^{(i)}(\theta) \propto \frac{q_s^{(i)}(\theta)}{t_k^{(i)}(\theta)}$$

2 Cavity distribution in Fed GVI

In Fed-GVI, we treat the approximate log likelihood terms as loss functions under the GVI framework. Taking $l_k^{(i)} := -\log t_k^{(i)}$ and $l_s^{(i)}(\boldsymbol{\theta}) = \sum_{m=1}^{M} l_m^{(i)}(\theta)$

$$\tilde{q}_s^{(i)}(\theta) = \arg\min_{q \in \mathcal{Q}} \left\{ \mathbb{E} \left[l_s^{(i)}(\theta) \right] + D[q || \pi_0] \right\}$$

Under a variational framework, if we want to compute the mth cavity now, we just remove the mth loss term and repeat the variational optimisation.

$$q_{\setminus m}^{(i)} = \arg\min_{q \in \mathcal{Q}} \left\{ \mathbb{E}_q \left[\sum_{j \neq m} l_j^{(i)}(\theta) \right] + D[q || \pi_0] \right\} = \arg\min_{q \in \mathcal{Q}} \left\{ \mathbb{E}_q \left[l_s^{(i)}(\boldsymbol{\theta}) - l_m^{(i)}(\boldsymbol{\theta}) \right] + D[q || \pi_0] \right\}$$

2.1 Cavity loss for Gaussians

We now prove that if our variational family is the family of Gaussians, we can express $\mathbb{E}_q\left[\sum_{k=1}^{M} l_k^{(i)}(\theta)\right]$ as one single expected negative log likelihood type term $\mathbb{E}_q[l_s^{(i)}(\theta)]$ avoiding the need for transmission of the individual client approximate loss terms for computation of the cavity distribution.

Let us first consider how $l_m^{(i)}(\theta)$s are computed:

$$l_m^{(i)}(\boldsymbol{\theta}) := l_m^{(i-1)}(\boldsymbol{\theta}) - log\frac{q_m^{(i)}(\boldsymbol{\theta})}{q_s^{(i-1)}(\boldsymbol{\theta})}$$

Observe that in case of a Gaussian variational family $\mathcal{Q}$, $l_k^{(i)}(\theta)$ will be in a quadratic form. Thus, we can assume $l_k^{(i)}(\theta) = a_k\theta^2 + b_k\theta + c_k$.

We get

$$a_k^{(i)} = a_k^{(i-1)} - \left[-\frac{1}{2}\left(\frac{1}{\sigma_m^2} - \frac{1}{\sigma_s^2}\right)\right]$$

and

$$b_k^{(i)} = b_k^{(i-1)} - \left[-\frac{1}{2} \times (-2)\left(\frac{\mu_m}{\sigma_m^2} - \frac{\mu_s}{\sigma_s^2}\right)\right]$$

If $q(\theta)$ follows $\mathcal{N}(\mu, \sigma^2)$,

$$\begin{aligned}
\mathbb{E}_q[l_k^{(i)}(\theta)] &= \mathbb{E}_q\left[a_k\theta^2 + b_k\theta + c_k\right] \\
&= a_k\mathbb{E}_q[\theta^2] + b_k\mathbb{E}_q[\theta] + c_k\mathbb{E}_q[1] \\
&= a_k(\sigma^2 + \mu^2) + b_k\mu + c_k
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{E}_q\left[\sum_{k=1}^{M} l_k^{(i)}(\theta)\right] &= \sum_{k=1}^{M}\mathbb{E}_q\left[l_k^{(i)}(\theta)\right] \\
&= \sum_{k=1}^{M}\left(a_k(\sigma^2 + \mu^2) + b_k\mu + c_k\right) \\
&= \left(\sum_{k=1}^{M} a_k\right)(\sigma^2 + \mu^2) + \left(\sum_{k=1}^{M} b_k\right)\mu + \left(\sum_{k=1}^{M} c_k\right)
\end{aligned}$$

Thus, we only require the aggregated $a_s = \sum_{k=1}^{M} a_k$ and $b_s = \sum_{k=1}^{M} b_k$ and not each (a_k, b_k) to compute $\mathbb{E}_q\left[\sum_{k=1}^{M} l_k^{(i)}(\theta)\right]$ to variationally find the cavity distribution.

$$\begin{aligned}
q_{\backslash m}^{(i)} &= \arg\min_{q\in\mathcal{Q}}\left\{\mathbb{E}_q\left[\sum_{j\neq m} l_j^{(i)}(\theta)\right] + D[q||\pi_0]\right\} \\
&= \arg\min_{q\in\mathcal{Q}}\left\{\mathbb{E}_q\left[l_s^{(i)}(\boldsymbol{\theta}) - l_m^{(i)}(\boldsymbol{\theta})\right] + D[q||\pi_0]\right\} \\
&= \arg\min_{q\in\mathcal{Q}}\left\{\mathbb{E}_q\left[l_s^{(i)}(\boldsymbol{\theta})\right] - \mathbb{E}_q\left[l_m^{(i)}(\boldsymbol{\theta})\right] + D[q||\pi_0]\right\} \\
&= \arg\min_{q\in\mathcal{Q}}\left\{(a_s - a_m)(\sigma^2 + \mu^2) + (b_s - b_m)\mu + D[q||\pi_0]\right\}
\end{aligned}$$

2.2 Exponential family

Assume $q(\theta) \sim$ Exponential Family: $q(\theta) = h(\theta) \exp\left(\eta^\top T(\theta) - A(\eta)\right)$.
If $l_k^{(i)}(\theta)$ is the ratio of two members of this exponential family

$$\begin{aligned}
q_1(\theta) &= h_1(\theta) \exp\left(\eta_1^\top T(\theta) - A_1(\eta_1)\right), \quad q_2(\theta) = h_2(\theta) \exp\left(\eta_2^\top T(\theta) - A_2(\eta_2)\right), \\
l_k^{(i)}(\theta) &= \log \frac{q_1(\theta)}{q_2(\theta)} \\
&= \log h_1(\theta) - \log h_2(\theta) + \eta_1^\top T(\theta) - \eta_2^\top T(\theta) - A_1(\eta_1) + A_2(\eta_2) \\
&= (\eta_1 - \eta_2)^\top T(\theta) + (\log h_1(\theta) - \log h_2(\theta)) - (A_1(\eta_1) - A_2(\eta_2)).
\end{aligned}$$

Restricting ourselves to $\mathcal{Q}$ such that the base measure $h(\theta) = 1$

$$l_k^{(i)}(\theta) = \eta_k^\top T(\theta) + c_k$$

$$\begin{aligned}
\mathbb{E}_q[l_k^{(i)}(\theta)] &= \mathbb{E}_q[\eta_k^\top T(\theta)] + \mathbb{E}_q[c_k] \\
&= \eta_k^\top \mathbb{E}_q[T(\theta)] + c_k \\
&= \eta_k^\top \bar{T}_q + c_k, \quad \text{where } \bar{T}_q = \nabla_\eta A(\eta) \\
\mathbb{E}_q\left[\sum_{k=1}^M l_k^{(i)}(\theta)\right] &= \sum_{k=1}^M \mathbb{E}_q[l_k^{(i)}(\theta)] \\
&= \sum_{k=1}^M \left(\eta_k^\top \bar{T}_q + c_k\right) \\
&= \left(\sum_{k=1}^M \eta_k\right)^\top \bar{T}_q + \sum_{k=1}^M c_k.
\end{aligned}$$

Thus, we only require the aggregated $\sum_{k=1}^M \eta_k$ to optimise $\mathbb{E}_q\left[\sum_{k=1}^M l_k^{(i)}(\theta)\right]$ wrt q.