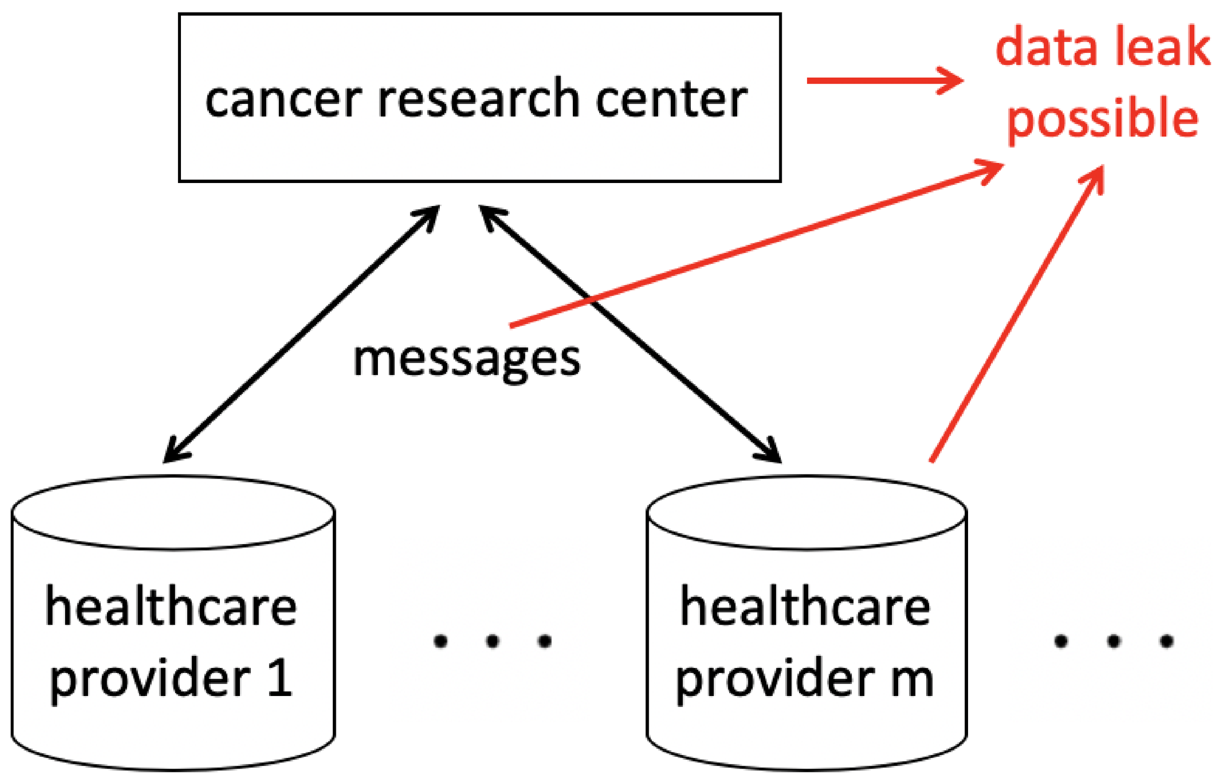
Distributed Variational Inference and Privacy

Xiping Liu (xl445@cam.ac.uk), supervised by Dr Richard Turner

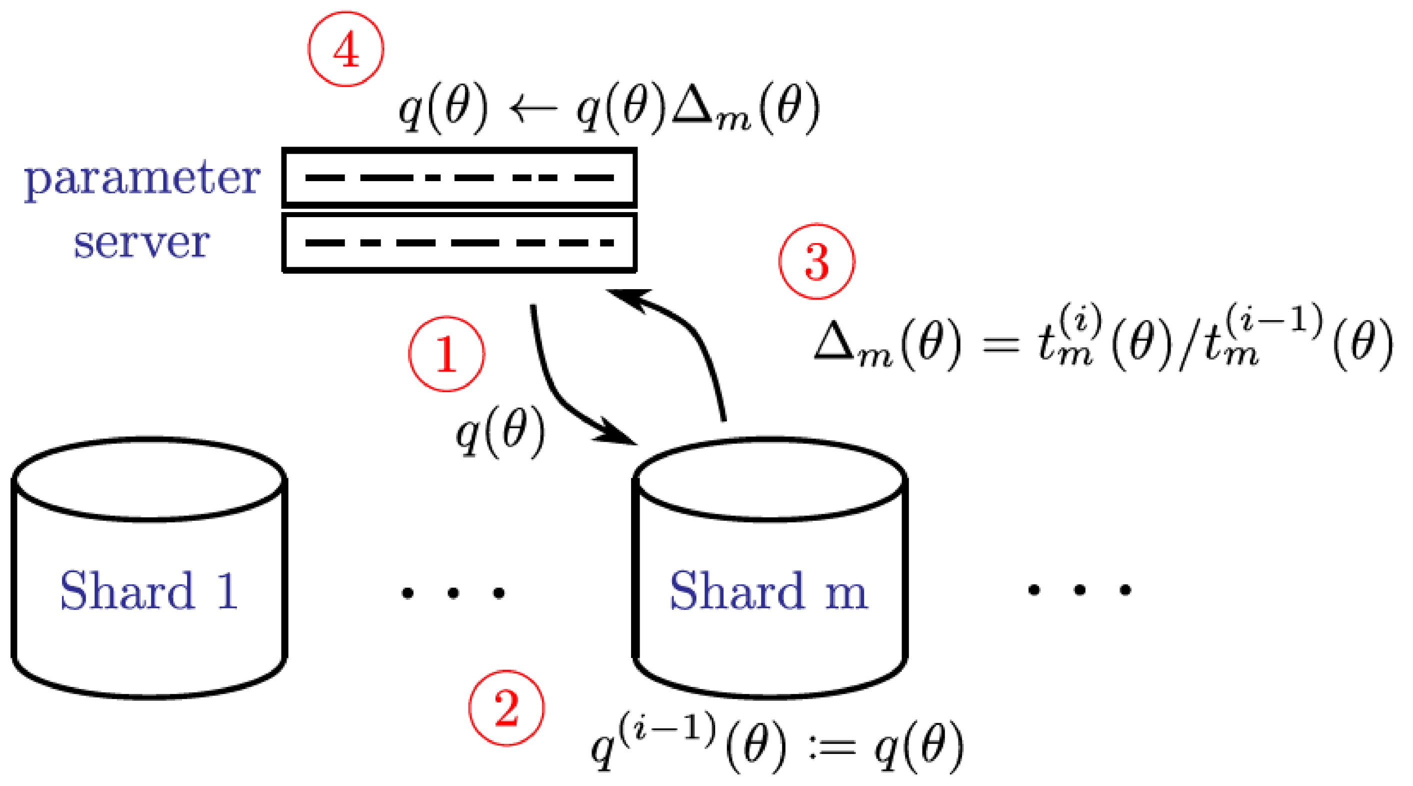
UNIVERSITY OF CAMBRIDGE

Introduction



- Aim to extend Partitioned Variational Inference (PVI) to support private federated machine learning using the concept of differential privacy (DP)

Partitioned Variational Inference (PVI)



$$q^{(i)}(\theta) = \underset{q(\theta)}{\operatorname{argmax}} \int \mathrm{d}\theta q(\theta) \log \frac{q^{(i-1)}(\theta)p(\mathbf{y}_m|\theta)}{q(\theta)t_m^{(i-1)}(\theta)}$$

$$t_m^{(i)}(\theta) = \frac{q^{(i)}(\theta)}{q^{(i-1)}(\theta)/t_m^{(i-1)}(\theta)}$$

Figure: Steps of the PVI algorithm when being used for federated learning [1]

Differential Privacy (DP)

- **Definition** (ϵ, δ)-Differential Privacy [2]: A randomized algorithm $\mathcal{A}$ is (ϵ, δ)-differentially private if for all pairs of adjacent data sets $(\mathcal{D}, \mathcal{D}')$ and for any subset of outputs $\mathcal{S}$:

$$\Pr(\mathcal{A}(\mathcal{D}) \in \mathcal{S}) \leq e^{\epsilon} \Pr(\mathcal{A}(\mathcal{D}') \in \mathcal{S}) + \delta$$

- Smaller ϵ and δ corresponds to stronger privacy guarantee
- DP is often achieved by clipping outputs and injecting Gaussian noise (Gaussian mechanism)
- The moments accountant keeps track of the total privacy guarantee composed by individual privacy guarantees and provides tight upper bounds on ϵ and δ [3]

Differentially Private PVI

1. Add DP to messages sent from workers to central server

- For each worker $m = 1, \ldots, M$ [4]:
  Compute new parameters for this worker:

$$\lambda_m = \arg\max_{\lambda} \int q(\theta|\lambda) \log \frac{q(\theta|\lambda^{(i-1)})p(\mathbf{y}_m|\theta)}{q(\theta|\lambda)t_m^{(i-1)}(\theta)} d\theta$$

$$\Delta\lambda_m = \lambda_m - \lambda^{(i-1)}$$

Clip and corrupt update:

$$\tilde{\Delta}\lambda_m = \alpha \left[\frac{\Delta\lambda_m}{\max(1, \|\Delta\lambda_m\|_2/C)} + \frac{\sigma C}{\sqrt{M}} z \right] \text{where } z \sim \mathcal{N}(0, 1)$$

$$\lambda_m = \lambda^{(i-1)} + \tilde{\Delta}\lambda_m$$

Update the approximate likelihood:

$$t_m^{(i)}(\theta) = \frac{q(\theta|\lambda_m)}{q(\theta|\lambda^{(i-1)})} t_m^{(i-1)}(\theta)$$

- For the central server, compute new global parameters [4]:

$$\lambda^{(i)} = \lambda^{(i-1)} + \sum_{m=1}^{M} \tilde{\Delta}\lambda_m$$

Differentially Private PVI (continued)

2. Add DP to every data point of a worker

- Achieved by optimizing local free energy using differentially private stochastic gradient descent [3]
- The worker is protected against all other parties since any external communication is differentially private

Future Experiments

- Test differentially private PVI on various models
  - 1-dimensional regression model
  - Multi-dimensional regression models
  - Non-linear models, like Bayesian neural networks
- Compare the above two ways of adding DP to PVI to see how privacy level and statistical performance trade off
- Investigate three different scheduling plans of messages: parallel, sequential, and asynchronous

References

[1] Thang D. Bui, Cuong V. Nguyen, Siddharth Swaroop, and Richard E. Turner. *Partitioned Variational Inference: A unified framework encompassing federated and continual learning*. arXiv preprint arXiv:1811.11206, 2018.

[2] Joonas Jälkö, Onur Dikmen, and Antti Honkela. *Differentially Private Variational Inference for Non-conjugate Models*. arXiv preprint arXiv:1610.08749, 2016.

[3] Martín Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. *Deep Learning with Differential Privacy*. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS 2016): 308-318.

[4] Mrinank Sharma. *Differential Privacy & Approximate Bayesian Inference*. MEng Thesis, Department of Engineering, University of Cambridge, 2019.