Differential privacy and Bayesian learning

Antti Honkela

Helsinki Institute for Information Technology HIIT,
Department of Mathematics and Statistics &
Department of Public Health,
University of Helsinki

Workshop on Advances in Approximate Bayesian Inference
NIPS 2017, 8 December 2017

The need for privacy: theory

No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.

—Universal Declaration of Human Rights, Article 12

The need for privacy: practice

- Lack of data is a major impediment to development of personalised medicine, but data cannot be shared without strong privacy
- The European General Data Protection Regulation (GDPR) coming to force in May 2018 sets strict conditions for use of private information with huge fines for breaches
- ML models retain a lot of private information, vulnerable to model inversion attacks
- Simple methods without strong theoretical guarantees cannot be relied upon in the long term

ϵ-differential privacy (DP; Dwork *et al.*, 2006)

Definition

An algorithm $\mathcal{M}$ operating on a data set $\mathcal{D}$ is said to be ϵ-differentially private (ϵ-DP) if for any two data sets $\mathcal{D}$ and $\mathcal{D}'$, differing only by one sample, the probabilities of obtaining any result S fulfil

$$\frac{\Pr(\mathcal{M}(\mathcal{D}) \in S)}{\Pr(\mathcal{M}(\mathcal{D}') \in S)} \leq e^{\epsilon}.$$

(ϵ, δ)-differential privacy

Definition

An algorithm $\mathcal{M}$ operating on a data set $\mathcal{D}$ is said to be (ϵ, δ)-differentially private ((ϵ, δ)-DP) if for any two data sets $\mathcal{D}$ and $\mathcal{D}'$, differing only by one sample, the probabilities of obtaining any result S fulfil

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

DP release of a function value $f(\mathcal{D})$

1. Evaluate the sensitivity of f

$$\Delta_p f = \max_{\|\mathcal{D}-\mathcal{D}'\|=1} \|f(\mathcal{D}) - f(\mathcal{D}')\|_p.$$

2. Compute

$$\mathcal{M}(\mathcal{D}) = f(\mathcal{D}) + \frac{\Delta_p f}{\epsilon}\eta$$

with

- **Laplace mechanism:** Pure ϵ-DP with $p = 1$ and

$$\eta \sim \text{Laplace}(0, 1)$$

- **Gaussian mechanism:** (ϵ, δ)-DP with $p = 2$ and

$$\eta \sim \mathcal{N}(0, 2\ln(1.25/\delta))$$

DP choice among alternatives

1. Evaluate the sensitivity of the utility $u(\mathcal{D}, r)$ over a choice between $r \in \mathcal{R}$

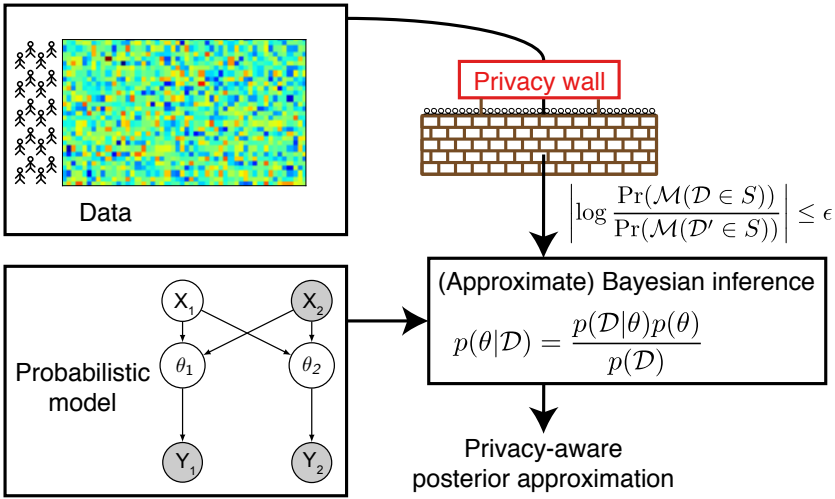
$$\Delta u = \max_{r \in (R)} \max_{\|\mathcal{D} - \mathcal{D}'\| = 1} |u(\mathcal{D}, r) - u(\mathcal{D}', r)|.$$

2. **Exponential mechanism**: choose r with probability proportional to

$$p(r) \propto \exp\left(\frac{\epsilon u(x, r)}{2\Delta u}\right).$$

This is ϵ-DP (McSherry & Talwar, FOCS 2007).

DP Bayesian inference



DP mechanisms for Bayesian inference

Three approaches to DP Bayesian inference:

1. Drawing single samples from the posterior with the exponential mechanism (Dimitrakakis *et al.*, ALT 2014; Wang *et al.*, ICML 2015; Geumlek *et al.*, NIPS 2017)
2. Sufficient statistic perturbation (SSP) for exponential family models with Laplace/Gaussian mechanism
   (e.g. Foulds *et al.*, UAI 2016; Honkela *et al.*, 2016, Park *et al.*, 2016)
3. Perturbation of gradients in SG-MCMC (Wang *et al.*, ICML 2015) or VI (Jälkö *et al.*, UAI 2017) with Laplace/Gaussian mechanism

Approach 1: Posterior sampling mechanisms

- Elegant and seemingly broadly applicable, but
- ... requires bounded or Lipschitz likelihood (proofs can be challenging)
- ... one sample tells very little, cost of additional samples linear in the number of samples
- ... privacy guarantee is conditional on exact sampling, which is infeasible for most models

Approach 2: Sufficient statistic perturbation (SSP)

For exponential family models, all information about the data $\mathcal{D} = \{x_1, \ldots, x_n\}$ is contained in the sum of sufficient statistics $\sum_i S(x_i)$.

This suggests a differentially private mechanism where we apply e.g. the Laplace mechanism on the sum to obtain perturbed sufficient statistics

$$\mathcal{M}(\mathcal{D}) = \sum_i S(x_i) + \xi,$$

with $\xi \sim \mathrm{Lap}(0, \Delta_1(S)/\epsilon)$, and then proceed with the inference as usual (Foulds *et al.*, UAI 2016; Honkela *et al.*, 2016).

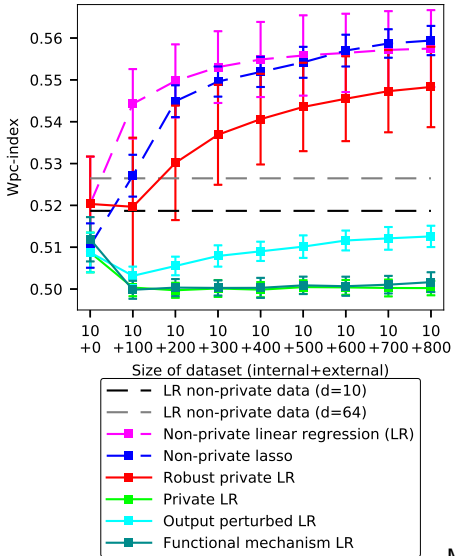
Consistency and efficiency

- Consistency: SSP DP estimates of posterior mean parameters converge to the corresponding non-private values as $n \to \infty$

$$\begin{aligned}\hat{\theta}_{\mathcal{M}} &= \frac{\tau + \mathcal{M}(\mathcal{D})}{n + n_0} = \frac{\tau + \sum_i S(x_i) + \xi}{n + n_0} \\ &= \frac{\tau + \sum_i S(x_i)}{n + n_0} + \frac{\xi}{n + n_0} \\ &\xrightarrow{p} \frac{\tau + \sum_i S(x_i)}{n + n_0} = \hat{\theta}_{NP}.\end{aligned}$$

- Convergence rate $\mathcal{O}(1/n)$ is optimal for any (ϵ, δ)-DP mechanism.

SSP results on drug sensitivity prediction

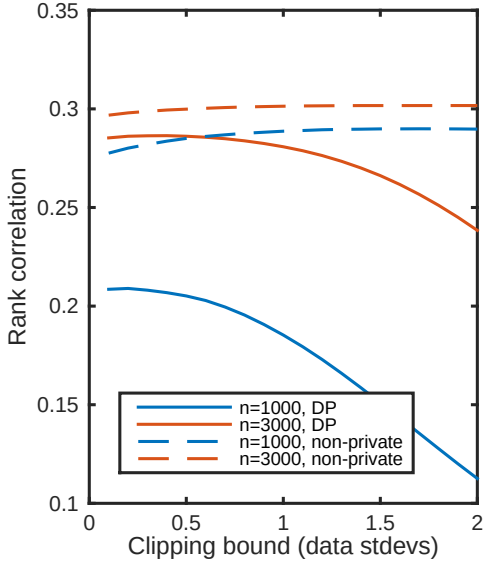


Mrinal Das, Arttu Nieminen

Practical challenges in efficient DP Bayesian learning

- Latent variables cannot be shared
- Asymptotic efficiency is insufficient to guarantee practical efficiency
- High dimensional data needs more DP noise
  - More aggressive dimensionality reduction than usual often needed
- Further: a single outlier can impose huge sensitivity on the data
  - Need to inject a lot of noise in DP to mask it
  - The useful contribution such points have in learning is at best minimal

Decreasing the sensitivity by clipping

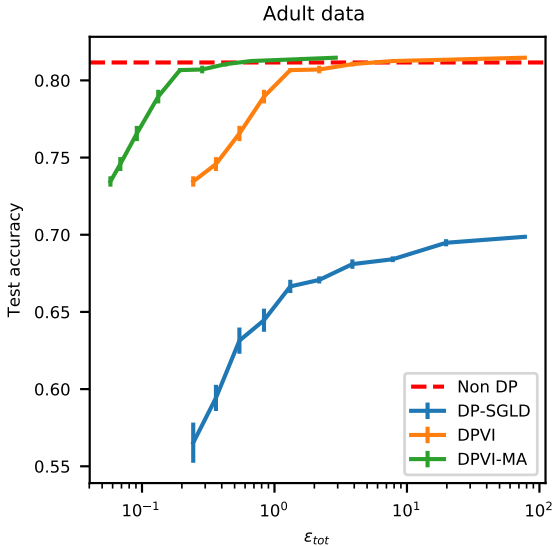


Approach 3: DP gradient perturbation methods

Assume target $\mathcal{L}(\theta, X) = \sum_i \mathcal{L}_i(\theta, x_i)$ (posterior or ELBO)

1. Each $g_i(\theta) = \nabla_\theta \mathcal{L}_i(\theta, x_i)$ is clipped s.t. $||g_i(\theta)||_2 \leq c_t$ in order to calculate *gradient sensitivity*
2. Subsampling x_i with frequency q in order to use the *privacy amplification theorem*
3. Gradient contributions from all data samples in the mini batch are summed and perturbed with Gaussian noise $\mathcal{N}(0, \sigma^2 \mathbf{I})$
4. Total privacy cost can be computed from composition theorems or using the *moments accountant* (Abadi *et al.*, CCS 2016)

DPVI logistic regression results on UCI Adult



Joonas Jälkö, Onur Dikmen

Open questions / research directions

- Practical methods for making inference aware of the injected DP noise
  - Cf. Williams & McSherry (NIPS 2010)
- Posterior uncertainty calibrated with the injected DP noise
  - Many current methods ignore this, running inference as if there was no extra noise
- Model engineering for DP: decreasing sensitivity
  - Robust models should in theory be better for privacy, but need to make sure our methods can take advantage of this

Conclusion

- DP as a strong privacy framework
- *Sufficient statistic perturbation* asymptotically consistent and efficient for exponential family models
- For finite data: *dimensionality reduction* and *clipping* the data are essential for obtaining better performance
- DPVI and DP-SG-MCMC applicable to more general models
- DP can be combined with encryption to run securely on distributed data (e.g. Heikkilä *et al.*, NIPS 2017)

Acknowledgements

Mrinal Das
Onur Dikmen
Mikko Heikkilä
Eemil Lagerspetz
Joonas Jälkö
Arttu Nieminen
Sasu Tarkoma
Kana Shimizu
Samuel Kaski

Funding: Academy of Finland