Fast Bayesian Coresets via Subsampling and Quasi-Newton Refinement



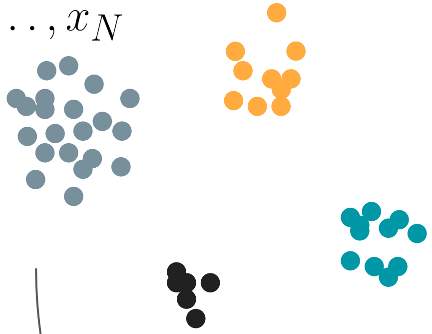
Cian Naik Judith Rousseau
University of Oxford

Trevor Campbell
University of British Columbia

Data Subsampling for Large Scale Inference

Large dataset (size N)

$x_1, \ldots, x_N$



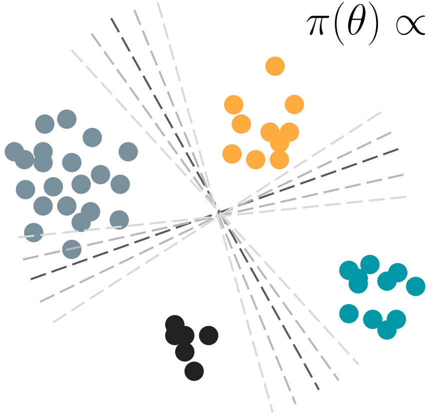
likelihood $\mathcal{L}_n(\theta) := \log p(x_n \mid \theta)$

prior $\pi_0(\theta)$

Inference (slow)

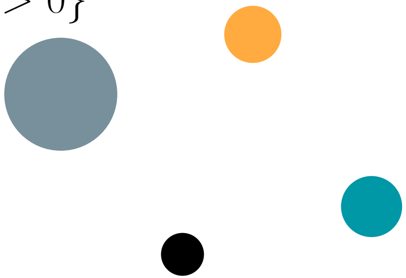
'Full' posterior

$\pi(\theta) \propto e^{\sum_{n=1}^{N} \mathcal{L}_n(\theta)} \pi_0(\theta)$



Small, weighted dataset (size M)

$\{(x_i, w_i) | w_i > 0\}$



$w \geq 0$

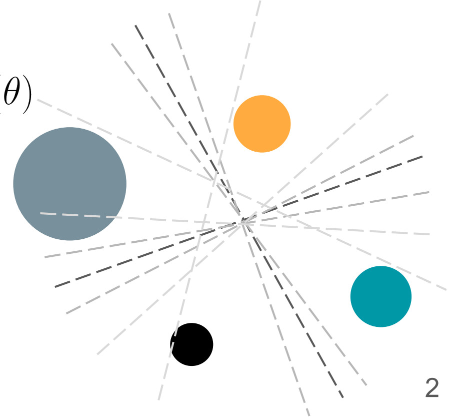
$||w||_0 = M \ll N$

Coreset construction (fast)

Inference (fast)

'Coreset' posterior

$\pi_w(\theta) \propto e^{\sum_{n=1}^{N} w_n \mathcal{L}_n(\theta)} \pi_0(\theta)$



Subsampling and Refinement

Goal: minimize $D_{KL}(\pi_w || \pi)$

Step 1: Uniformly sample $M \ll N$ points to include in the coreset.

Step 2: Use the exact form for the gradient, and an approximation of the Hessian to optimize the weights to minimize the KL via a Quasi-Newton algorithm.

Theoretical Results

Theorem 1 (Uniform Sampling): The optimal coreset posterior is close to the full posterior for large N.

For $M \gtrsim D(\log N + 1)$, with probability greater than $1 - o(1)$ and $N \to \infty$:

$$\min_{w \in \mathcal{W}_N} \mathrm{KL}(\pi_w || \pi) = o(1)$$

Theorem 2 (Quasi-Newton Refinement) : Our algorithm converges to the optimal coreset posterior.

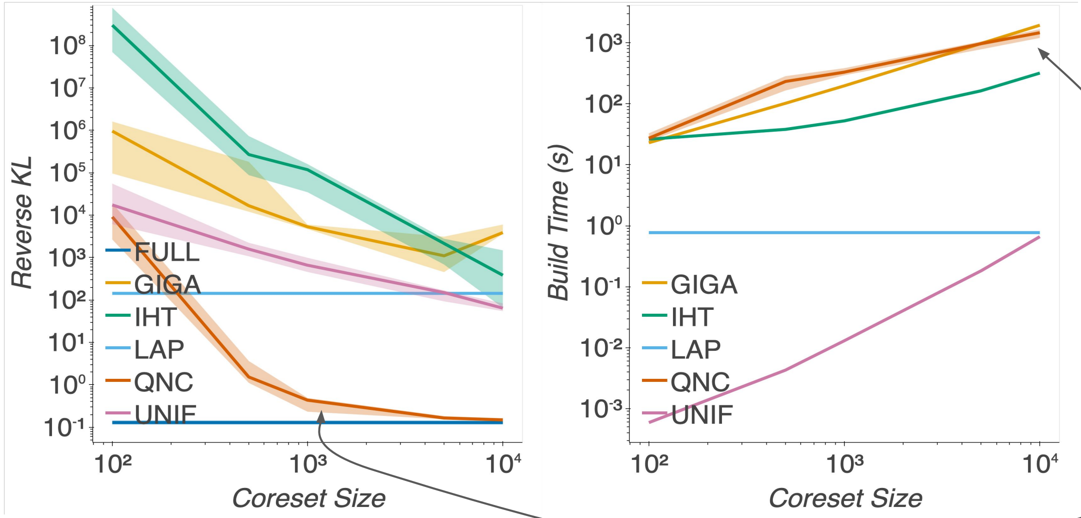
In a special (but widely applicable) case, where $w^\star$ is the optimal coreset:

$$||w_k - w^\star|| \leq \eta^k \, ||w_0 - w^\star||$$

Empirical Results

Experiment: Bayesian logistic regression

$$y_n \mid x_n, \theta \overset{\text{indep}}{\sim} \text{Bern}\left(\frac{1}{1+e^{-x_n^T\theta}}\right) \qquad \theta_i \overset{\text{i.i.d.}}{\sim} \text{Cauchy}(0,\sigma), \quad i = 1,\ldots,D$$



FULL - Full posterior

LAP - Laplace approximation

UNIF - Uniform subsampling

GIGA, IHT - Sparse regression coresets

QNC - our method

Thanks

For further questions:
ciannaik@gmail.com
trevor@stat.ubc.ca

References

[MXMC20]: Bayesian Pseudocoresets
Manousakas, Xu, Mascolo, Campbell
https://proceedings.neurips.cc/paper/2020/hash/ab452534c5ce28c4fbb0e102d4a4fb2e-Abstract.html

[CBe19]: Sparse Variational Inference: Bayesian Coresets from Scratch
Campbell, Beronov
https://proceedings.neurips.cc/paper/2019/hash/7bec7e63a493e2d61891b1e4051ef75a-Abstract.html

[CBr18]: Bayesian Coreset Construction via Greedy Iterative Geodesic Ascent
Campbell, Broderick
ICML 2018, https://arxiv.org/abs/1802.01737

[CBr17,19]: Automated Scalable Bayesian Inference via Hilbert Coresets
Campbell, Broderick
JMLR 20(15), 1-38, 2019, https://arxiv.org/abs/1710.05053

[HCB16]: Coresets for scalable Bayesian logistic regression
Huggins, Campbell, Broderick
NeurIPS 2016, https://arxiv.org/abs/1605.06423