Bayesian ML: project topics

Rémi Bardenet

December 20, 2021

1 Nature of the project

A group is made of one or two students. Each group is to pick a paper is the list below, and each paper can be chosen only once. The assignment in on a first-come first-served basis. You can also choose to work on another paper, subject to my explicit approval. There will be bonus points for working alone, and for picking a paper labeled below as long or difficult.

The whole point is to read your paper with a critical mind. For the paper your group will have chosen, you should: (1) explain the contents of the paper, (2) emphasize the strong and weak points of the paper, and (3) apply it to real data of your choice when applicable. Bonus points will be considered if you are creative and add something insightful that is not in the original paper: this can be a theoretical point, an illustrative experiment, etc. Be explicit in your introduction what is your creative contribution.

2 Assignment of papers

As a first step, we ask each group to fill the spreadsheet at

https://lite.framacalc.org/uv85svkjtp-9rih

with the title of the paper, a link to it (if available), and the composition of the group. Please fill the form **before January 5**.

3 Format of the deliverable

Please have each group send

- one report as a pdf (≤ 5 pages) in the NeurIPS template,
- the link to a GitHub or GitLab repository containing your code and a detailed readme file with instructions to (compile/install and) run the code.

to remi.bardenet@gmail.com **no later than January 18**. There will be no deadline extension.

4 Proposed papers

Short and "easy" papers

[Easy1] B. Calderhead. A general construction for parallelizing Metropolis-Hastings algorithms. *Proceedings of the National Academy of Sciences*, 111(49):17408–17413, 2014.

[Easy2] Carlos M Carvalho, Nicholas G Polson, and James G Scott. Handling sparsity via the horseshoe. In *Artificial Intelligence and Statistics*, pages 73–80. PMLR, 2009.

[Easy3] Lionel Cucala, Jean-Michel Marin, Christian P Robert, and D Michael Titterington. A Bayesian reassessment of nearest-neighbor classification. *Journal of the American Statistical Association*, 104(485):263–273, 2009.

[Easy4] Romain Lopez, Pierre Boyeau, Nir Yosef, Michael I Jordan, and Jeffrey Regier. Decision-making with auto-encoding variational Bayes. In *Advances in Neural Information Processing Systems*, 2020.

[Easy5] Nicholas G Polson and Steven L Scott. Data augmentation for support vector machines. *Bayesian Analysis*, 6(1):1–23, 2011.

[Easy6] M. Rabinovich, E. Angelino, and M. Jordan. Variational consensus Monte Carlo. In *Advances in Neural Information Processing Systems*, pages 1207–1215, 2015.

[Easy7] C. E. Rasmussen and Z. Ghahramani. Bayesian Monte Carlo. In *Advances in Neural Information Processing Systems*, 2003.

[Easy8] Y. W. Teh, D. Newman, and M. Welling. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In *Advances in neural information processing systems*, pages 1353–1360, 2007.

[Easy9] M. Welling and Y. W. Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688, 2011.

Long or difficult papers

[Hard1] R. Bardenet, A. Doucet, and C. Holmes. On Markov chain Monte Carlo methods for tall data. *Journal of Machine Learning Research (JMLR)*, 2017.

[Hard2] A. Belhadji, R. Bardenet, and P. Chainais. Kernel interpolation with continuous volume sampling. In *International Conference on Machine Learning (ICML)*, 2020.

[Hard3] P. Germain, F. Bach, A. Lacoste, and S. Lacoste-Julien. Pac-Bayesian theory meets Bayesian inference. In *Advances in Neural Information Processing Systems*, pages 1884–1892, 2016.

[Hard4] P. Grünwald and T. Van Ommen. Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Analysis*, 12(4):1069–1103, 2017.

[Hard5] Matthew D Hoffman and Andrew Gelman. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[Hard6] Pierre E Jacob, John O'Leary, and Yves F Atchadé. Unbiased Markov chain Monte Carlo methods with couplings. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3):543–600, 2020.

[Hard7] P. Rigollet and A. Tsybakov. Exponential screening and optimal rates of sparse estimation. *The Annals of Statistics*, 39(2):731–771, 2011.