Bayesian ML: project topics

Julyan Arbel, Rémi Bardenet

February 23, 2021

1 Nature of the project

Students should form groups of two to three, each group undertaking one project. We suggest in Section 4 a few scientific papers that can each lead to a project, but you can choose another paper, subject to our approval. There will be a few bonus points for groups of two and for groups that choose a journal paper (compare to a conference paper like NeurIPS or ICML, which are shorter).

For the paper your group will have chosen, you should: (1) explain the theoretical, computational and/or empirical methods, (2) emphasize the main 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. The whole point is to read the paper with a critical mind.

2 Assignment of papers

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

https://lite.framacalc.org/9ll3-bml

with the title of the paper, a link to it (if available), and the composition of the group. We ask that you fill in the form **before March 5**.

3 Format of the deliverable

You can use either Python or R for the programming part. 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 both teachers[1] **no later than March 22**. There will be no deadline extension.

4 Proposed papers

Lectures on Bayesics and foundations

[A1] 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.

[A2] 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.

[A3] 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.

[A4] Romain Lopez, Pierre Boyeau, Nir Yosef, Michael I Jordan, and Jeffrey Regier. Decision-making with auto-encoding variational Bayes. *arXiv preprint arXiv:2002.07217*, 2020.

[A5] Nicholas G Polson, Steven L Scott, et al. Data augmentation for support vector machines. *Bayesian Analysis*, 6(1):1–23, 2011.

[A6] P. Rigollet and A. Tsybakov. Exponential screening and optimal rates of sparse estimation. *The Annals of Statistics*, 39(2):731–771, 2011.

Lecture on MCMC

[B1] B. Calderhead. A general construction for parallelizing Metropolis-Hastings algorithms. *Proceedings of the National Academy of Sciences*, 111(49):17408–17413, 2014.

[B2] 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.

[B3] 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.

[B4] 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.

[1] if the above link is broken, this means: julyan.arbel@inria.fr, and remi.bardenet@gmail.com

Lecture on variational inference

[C1] G. P. Dehaene and S. Barthelmé. Bounding errors of expectation-propagation. In *Advances in Neural Information Processing Systems (NIPS)*, pages 244–252, 2015.

[C2] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *Journal of Machine Learning Research (JMLR)*, 14(1):1303–1347, May 2013.

[C3] M. Rabinovich, E. Angelino, and M. Jordan. Variational consensus Monte Carlo. In *Advances in Neural Information Processing Systems*, pages 1207–1215, 2015.

[C4] 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.

Lecture on Bayesian nonparametrics

[E1] Fadhel Ayed, Juho Lee, and Francois Caron. Beyond the chinese restaurant and Pitman-Yor processes: Statistical models with double power-law behavior. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 395–404. PMLR, 09–15 Jun 2019.

[E2] Federico Camerlenghi, Stefano Favaro, Zacharie Naulet, and Francesca Panero. Optimal disclosure risk assessment. *Annals of Statistics*, 2021.

[E3] Federico Camerlenghi, Antonio Lijoi, Peter Orbanz, and Igor Prünster. Distribution theory for hierarchical processes. *Ann. Statist.*, 47(1):67–92, 02 2019.

[E4] Pierpaolo De Blasi, Ramsés H Mena, and Igor Prünster. Asymptotic behavior of the number of distinct values in a sample from the geometric stick-breaking process. *arXiv preprint arXiv:2101.07607*, 2021.

[E5] Michael Lavine. Some aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, pages 1222–1235, 1992.

[E6] Jeffrey W Miller and Matthew T Harrison. Inconsistency of Pitman-Yor process mixtures for the number of components. *The Journal of Machine Learning Research*, 15(1):3333–3370, 2014.

[E7] Jim Pitman and Marc Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 25(2):855–900, 1997.

Lecture on Bayesian deep learning

[F1] Soufiane Hayou, Arnaud Doucet, and Judith Rousseau. On the impact of the activation function on deep neural networks training. In *International Conference on Machine Learning*, pages 2672–2680. PMLR, 2019.

[F2] Mohammad Emtiyaz Khan, Didrik Nielsen, Voot Tangkaratt, Wu Lin, Yarin Gal, and Akash Srivastava. Fast and scalable Bayesian deep learning by weight-perturbation in ADAM. *arXiv preprint arXiv:1806.04854*, 2018.

[F3] Mohammad Emtiyaz E Khan, Alexander Immer, Ehsan Abedi, and Maciej Korzepa. Approximate Inference Turns Deep Networks into Gaussian Processes. In *Advances in Neural Information Processing Systems*, pages 3088–3098, 2019.

[F4] J. Lee, J. Sohl-Dickstein, Jeffrey Pennington, Roman Novak, Sam Schoenholz, and Yasaman Bahri. Deep neural networks as Gaussian processes. In *International Conference on Machine Learning*, 2018.

[F5] A. Matthews, M. Rowland, J. Hron, R. Turner, and Z. Ghahramani. Gaussian process behaviour in wide deep neural networks. In *International Conference on Learning Representations*, volume 1804.11271, 2018.

[F6] Max Welling and Yee 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.

[F7] Andrew Gordon Wilson and Pavel Izmailov. Bayesian deep learning and a probabilistic perspective of generalization. *arXiv preprint arXiv:2002.08791*, 2020.