Bayesian ML: project topics

Julyan Arbel

January 29, 2021

1 Nature of the project

Students should do the project either individually or by pair, 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.

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/9lhk-mash-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 February 8**.

3 Format of the deliverable

You can use either Python or R for the programming part, if you wish to include one. 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 this address[1] **no later than March 2**. There will be no deadline extension.

[1] if the above link is broken, this means: julyan.arbel@inria.fr

4 Proposed papers

Part 1: Bayesian nonparametrics

[BNP1] 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.

[BNP2] Federico Camerlenghi, Stefano Favaro, Zacharie Naulet, and Francesca Panero. Optimal disclosure risk assessment. *Annals of Statistics*, 2021.

[BNP3] Federico Camerlenghi, Antonio Lijoi, Peter Orbanz, and Pr

[BNP4] François Caron and Emily B Fox. Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1295–1366, 2017.

[BNP5] 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.

[BNP6] Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230, 1973.

[BNP7] Jan Greve, Bettina Grün, Gertraud Malsiner-Walli, and Sylvia Frühwirth-Schnatter. Spying on the prior of the number of data clusters and the partition distribution in Bayesian cluster analysis, 2020.

[BNP8] 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.

Part 2: Bayesian deep learning

[BDL1] Y. Gal and Z. Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *International conference on machine learning*, pages 1050–1059, 2016.

[BDL2] 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.

[BDL3] 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.

[BDL4] 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.

[BDL5] 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.

[BDL6] 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.

[BDL7] 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.

[BDL8] Andrew Gordon Wilson and Pavel Izmailov. Bayesian deep learning and a probabilistic perspective of generalization. *arXiv preprint arXiv:2002.08791*, 2020.

Evaluation criteria

Slides quality

5: Structure is excellent

4: Structure is good, there are minor suggestions

3: Correct global organisation, but some details lack of clarity.

2: Navigating requires some effort. Lack of structure; some parts are confusing.

1: Student does not seem to spend more than 5 mins on that

0: No slides

Clarity of the explanation/speech

5: Everything is as clear as blue sky, excellent

4: Clear overall, minor problems

3: Lack of clarity in details, but the idea explanation is understandable

2: Following requires some effort, some parts are confusing

1: Student doesn't seem to know what they is talking about

0: Doesn't know what to say. At all.

The paper understanding, answers to questions

5: Outstanding depth of analysis and validation process, excellent answers.

4: Very good depth in analysis, but some points remain to be investigated with more care.

3: Satisfying analysis, some potential mistakes or methodological errors in application, validation or interpretation.

2: Weaknesses in analysis, with some lack of depth.

1: Some poor understanding of the topic.

0: Total misunderstanding.