Advanced Machine Learning Module
Paper Replication Exercise

Richard E. Turner

Initial Plan Due: 12 noon end of second week of full term (30th January, email Rich Turner)
Posters Due: 12 noon last Monday of full term (11th March, submission via Moodle, Anne will arrange printing)
Poster Session: last Friday of full term (15th March), Dyson Seminar room 10am-12pm
Report Due: 12 noon Friday three weeks after poster session (5th April, submission via Moodle)

In this coursework you will attempt to replicate at least one key aspect of a fundamental machine learning research paper. You will present your findings at a poster session and in a final report. This coursework should be conducted in teams of three people. It should take roughly 10hrs per group member. The groups will be randomly assigned.

Some suggested papers are listed below. You can suggest alternative papers, but you will need to have these approved. The papers vary greatly in terms of their length, the level of technical content, the difficulty of algorithmic implementation, and the number of experiments performed. Please take these factors into consideration when choosing a paper and when targeting the aspect to reproduce. Discuss your choices in person if you have questions.

Here are some important factors to bear in mind:

- please acknowledge when you have used publicly available code to reproduce the experiments, referencing it and describing modifications that you made
- where possible and sensible, use the same data as used in the original papers, but if it is not available please seek alternatives
- please feel free to extend the experimental results e.g. to consider data and algorithms that were not part of the original study, but note that this is not necessary for the assignment
- ideally, each group should select a different paper to replicate (we will monitor choices, see Module administration and submission instructions below)
- students on the MLMI MPhil should form separate groups from the other students who are taking the class for credit

Assessment

Each group should submit an initial plan by email stating your assigned team number, the person in charge of submitting the poster and report, the title of the paper, and a brief plan of action (equivalent to roughly half a side of A4) by the end of the Lent Term (see Module administration and submission instructions below). This component carries no marks in itself, but must be completed to pass the module.

The poster evaluation will carry 35% of the marks. Each group should produce one A1-sized poster (see Module administration and submission instructions below). The poster should detail preliminary findings that they will present at a poster session to other students and to faculty. The faculty will grade the poster design, the presentation of the poster, and the progress the group has made.

The report will carry 65% of the marks. Each group should submit one report (see Module administration and submission instructions below). The report should describe the rationale for selecting the paper and the particular experiments for replication. It should summarise aspects of the paper relevant to the experiment

(both to set the experiment in context and to lay out relevant technical details). The report should detail the experimental simulations you ran, compare them to those in the original paper, and draw conclusions. You should end by critically evaluating your work and suggest what you would have done given more time. The report should be no longer than 5,000 words.

The recommended papers are:

1. Auto-Encoding Variational Bayes. DP Kingma, M Welling. Proceedings of the 2nd International Conference on Learning Representations, 2013[1]
2. Gaussian process latent variable models for visualisation of high dimensional data. ND Lawrence. Advances in neural information processing systems 16 (3), 329-336, 2004
3. A tutorial on particle filtering and smoothing: Fifteen years later. A Doucet, AM Johansen. Handbook of Nonlinear Filtering 12 (656-704), 3, 2009
4. Particle Markov chain Monte Carlo methods. C Andrieu, A Doucet, R Holenstein. Journal of the Royal Statistical Society: Series B Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2010
5. Variational learning of inducing variables in sparse Gaussian processes, MK Titsias, International Conference on Artificial Intelligence and Statistics, 567-574, 2009
6. Bayesian Learning via Stochastic Gradient Langevin Dynamics. M Welling, YW Teh. International Conference on Machine Learning (ICML), 2011
7. Elliptical slice sampling. I Murray, RP Adams, DJC MacKay. The Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), 541-548, 2010
8. Bayesian Learning for Neural Networks. RM Neal. Ph.D. Thesis, Dept. of Computer Science, University of Toronto, 1994 [Chapter 3]
9. Weight Uncertainty in Neural Networks. C Blundell J Cornebise K Kavukcuoglu D Wierstra. Proceedings of the 32nd International Conference on Machine Learning, 2015.
10. Practical Bayesian optimization of machine learning algorithms. J Snoek, H Larochelle, RP Adams. Advances in neural information processing systems, 2951-2959, 2012
11. Multiobjective optimization on a limited budget of evaluations using model-assisted mathcal S-metric selection. W Ponweiser, T Wagner, D Biermann, M Vincze. International Conference on Parallel Problem Solving from Nature. Springer Berlin Heidelberg, 2008.
12. Sequential Neural Models with Stochastic Layers. M Fraccaro, S Sønderby, U Paquet; O Winther. NIPS 2016.
13. Clamping Improves TRW and Mean Field Approximations. A Weller, J Domke, AISTATS 2016
14. Composing graphical models with neural networks for structured representations and fast inference. M Johnson, D Duvenaud, A Wiltschko, RP Adams and SR Datta In Advances in neural information processing systems, pp. 2946-2954, 2016.
15. Importance weighted autoencoders. Y Burda, R Grosse, and R Salakhutdinov. arXiv preprint arXiv:1509.00519, 2015.

[1] also see Weight Uncertainty in Neural Networks, Blundell et al. 2015

16. Scalable Variational Gaussian Process Classification. J Hensman, A Matthews, Z Ghahramani. Artificial Intelligence and Statistics, 9-12, 2015

17. Variational Continual Learning. CV Nguyen, Y Li, TD Bui, RE Turner, International Conference on Learning Representations, 2018

18. Variational Inference with Normalizing Flows. DJ Rezende and S Mohamed, Proceedings of the 32nd International Conference on Machine Learning, 2016

19. Towards a Neural Statistician, H Edwards, A Storkey, International Conference on Learning Representations, 2017

20. J Gilmer, S Justin, S Schoenholz, PF Riley, O Vinyals, and GE Dahl. Neural Message Passing for Quantum Chemistry. In International Conference on Machine Learning, 1263-1272. 2017.

21. DK Duvenaud, D Maclaurin, J Iparraguirre, R Bombarell, T Hirzel, A Aspuru-Guzik, and RP Adams. Convolutional networks on graphs for learning molecular fingerprints. In Advances in neural information processing systems 2224-2232, 2015.

22. Y Li, D Tarlow, M Brockschmidt, R Zemel. Gated Graph Sequence Neural Networks. International Conference on Learning Representations, 2016.

Module administration and submission instructions

The action points for submission are:

$\Longrightarrow$ The report submitter for each group should email Rich Turner (`ret26@cam.ac.uk`) by 12pm end of second week of full term with the following information. 1) the title of the paper that the group is going to replicate, and 2) include a brief plan of action (equivalent to half a page of A4).

$\Longrightarrow$ Each group must submit one poster by 12 noon on the last Monday of full term via Moodle. Anne will arrange printing. The poster session is in the Dyson Seminar Room on the last Friday of term 10am-12pm (the room will be open from 9.30am).

$\Longrightarrow$ Each group must submit one report by 12pm Friday three weeks after poster session via Moodle.

Here is a reminder for how to submit the files through Moodle:

1. Click on "Assessment Submission" on the Moodle Webpage
2. Click on the link entitled "Major Practical – Machine Learning"
3. Click on "Add Submission"
4. Drag the file to the "File Submission" field, wait until the files has uploaded (submit your report as a single pdf with a coursework coversheet at the front)
5. After the blue downloading status bar disappears click "Save Changes" button
6. Once you click on "confirm submission" you cannot resubmit or change anything