Chris Mingard

DPhil Student in Machine Learning, University of Oxford

(+44) 7887 376914 | christopher.mingard@queens.ox.ac.uk | C1510 | chris-mingard | chrismingard | Chris Mingard

Third year DPhil student at Oxford University (funded by an iCASE grant from IBM) with 4 years' experience in machine learning. I led work on two large-scale papers during my undergraduate degree, focusing on generalisation in neural networks, with the Louis Group, one of which has been published in JMLR. My PhD focuses on generalisation in neural networks and their applications to self-assembling systems, and I have recently completed an internship with Meta working on Graph Neural Networks.

Education

University of Oxford — *Oxford, U.K.*

DPhil in Machine Learning, Dept. Theoretical Physics/Dept. Theoretical Chemistry — *Oct. 2020 - Sept. 2023*

Masters in Mathematical and Theoretical Physics. Grade: Distinction (Rank 15th, 79%) — *Oct. 2019 - Jun. 2020*

B.A. in Physics. Grade: 1st — *Oct. 2016 - Jun. 2019*

The Judd School — *Tonbridge, U.K.*

A levels - Maths A*, Further Maths A*, Physics A*, Chemistry A*, French B. STEP II & III - 1 & 1. GCSEs - 10 A*. — *Sept. 2009 - Jun. 2016*

Publications & Preprints

Is SGD a Bayesian Sampler? Well almost.

Journal of Machine Learning Research 22(79) [Also my masters' thesis] — *Submitted Jun. 2020, Accepted Feb. 2021*

- First author. We find that the probability that learning algorithms commonly used in Deep Learning (such as Adam, SGD etc.), trained on (wide-layer) neural networks find a function f is very similar to the probability that kernel methods such as NTK and Gaussian Processes (with corresponding architectures) find f. This relationship can be tuned to near-perfect agreement, and is strong evidence that SGD optimised Deep Neural Networks can be understood with a Bayesian perspective. See also my blog post in Towards Data Science.

Feature Learning and Signal Propagation in Deep Neural Networks

ICML 2022, 14248-14282 — *June. 2022*

- This paper focuses on understanding representation learning in neural networks improves generalisation. We show that the layers that achieve some balance between forward and backward information loss are the ones with the highest alignment to data labels. Our experiments demonstrate an excellent match with the theoretical predictions.

Neural networks are biased towards low entropy functions.

Arxiv:1909.11522 [Result of undergraduate project funded by Hertford College] — *Jun. 2019*

- First author on a paper written with the Louis group. We proved several results focusing on biases at initialisation in perceptrons and small neural networks; and used large-scale simulations to test whether intuitions gained from the proofs scaled to large neural networks and real world datasets. One of our main proofs concerned the perceptron with no bias term and weights randomly initialised from a distribution symmetric about coordinate planes, acting as binary classifier on $\{0,1\}^n$. The probability that t elements in $\{0,1\}^n$ are classified as 1 is uniform in t. See also my blog post in Towards Data Science.

Experience

Meta — *Rathbone Square, London, UK*

Research Scientist Intern — *August - December 2022*

- I worked on improving the efficiency of training Graph Neural Networks (GNNs) with the AI Integrity Team. This project involves studying current neighbourhood samplers with the aim if improving them for efficient training of very large GNNs. I also worked on their internal codebase to improve training efficiency.

G-Research — *Whittington House, 19-30 Alfred Pl, London WC1E 7EA*

Quant ML research intern — *June-September 2023 (expected)*

- I will be working on time-series data.

Projects

NNGPs & NNs
In writing the above papers, I collaborated with Guillermo Valle-Perez to build a large modular codebase in running extensive experiments comparing neural networks (NNs) (including SOTA architectures and LSTMs) to their Neural Network Gaussian Process counterparts (NNGPs). Written in Keras/Tensorflow, GPy (and some PyTorch). Primarily used to sample the posteriors of arbitrary combinations of architectures, datasets, and hyperparameters, both for NNs and NNGPs. Also designed to be used with extremely wide frozen layers ($\mathcal{O}(10^9)$) parameters per layer), required for our latest experiments.

Boundary Machines
I have collaborated with Yoonsoo Nam to build a codebase in pytorch for a layer-wise training method (due primarily to theoretical interest) with custom layers. We use these Layerwise Feature Maximisers (LFMs) to investigate feature learning in neural networks. Implements multiple novel layer-wise training methods, and achieves competitive performance with SGD for multiple architectures/datasets including medium-scale CNNs/cifar10.

Rapid Image Classification
I created a small program that automatically selects potentially interesting features from an (electron/light) microscope image, containing potentially thousands of small features of interest, and presents them (auto-brightness and contrast adjusted for maximum ease of classification) one by one to a researcher for classification with a single key press. Includes a continue where you left off and infinite undo function. Sped up manual classification by 100x in several projects, one of which is being undertaken at NPL, Teddington.

Towards Data Science
I wrote three blog posts on medium (published in Towards Data Science) about my publications and surrounding theory. They have been viewed over 12000 views. See them at chrismingard.medium.com. I participated in a discussion around this at LessWrong (see here).

Graph Transformers
As part of my DPhil, I have implemented a spatio-temporal Transformer-based Graph Neural Network, with the goal of predicting how DNA origami self-assembles as it is cooled down. Written in pytorch, using pytorch-geometric for the graph layers. Tested on small DNA origami datasets, and I am currently generating larger DNA origami datasets using scadnano.

Particle Classification
I experimented with GANs, Normalising Flows, VAEs and BDTs on Monte Carlo generated data and background to act as a background remover for B meson decay.

Bias in Deep Q Learning
With a bursary from Hertford College, Oxford, I investigated biases in the priors of simple reinforcement learning systems. This required designing custom reinforcement learning agents with pytorch, and custom environments.

Other Projects
I created a generalised connect 4 environment (for any grid size in n dimensions and m in a row for any valid m) and a Deep Q learning agent for it. Tested on 8×8 and $6 \times 6 \times 6$ grids. I have also implemented Style transfer and a Google Deep Dream-like algorithm.

Skills

Python3 Keras, Tensorflow, Pytorch, Jax and GPy, and other standard python libraries for data science, mathematics and machine learning (e.g. numpy, pandas, matplotlib).

Other I have run large scale experiments on linux based clusters including machines on Google Cloud, and the Oxford clusters HYDRA and JADE. Basic Git, basic Bash. LaTeX

Courses Theories of Deep Learning, Differentiable Manifolds, Representation Theory, Quantum Field Theory, Topological Quantum Field Theory, General Relativity, Quantum Information Theory

Other achievements & Interests

- Oxford Blue in cricket (2021 and 2022). Vice Captain 2023.
- Grade 7 Piano with Merit
- I was highly commended for my Master's dissertation by The Examiners
- I was awarded two bursaries by Hertford College for undergraduate research (in 2018 and 2019).
- I am very interested in philosophy of science & rationality – both as tools for improving my research and ways of improving my understanding of the world.
- I am very interested in abstract maths, and followed the Oxford Maths 1st year course during my first year (2016) and informally sat the 1st year exams. I also play chess regularly.