dr Daniel Matthew Steinberg

Personal

Citizenship:	Australian and New Zealand
homepage:	http://dsteinberg.github.io/

Education

July 2013	Ph.D., Australian Centre for Field Robotics, **The University of Sydney**, NSW Thesis: "An Unsupervised Approach to Modelling Visual Data" Advisor: Prof. Stefan B. Williams Co-advisor: Dr. Oscar Pizarro
November 2008	B.E. Mechatronic Engineering, **The University of Sydney**, NSW Honours class 1
November 2008	B.Com. Finance, **The University of Sydney**, NSW

Employment

Current	Senior Research Scientist/Team Leader, **CSIRO (Data61)**, Canberra
2019–2023	Principal Researcher, **Gradient Institute**, Canberra As a principal researcher at Gradient Institute I undertook observational causal studies, researched ethically aware machine learning algorithms, and implemented machine learning algorithms and systems.
2016–2019	Senior Research Engineer, **CSIRO (Data61)**, Sydney & Canberra I was a senior research engineer in the inference systems team, where I derived and applied new machine learning algorithms to spatial prediction problems involving large scale demographic, environmental and natural science data. New software tooling was required for a number of these projects, many of which have been released open source.
2013–2016	Researcher, **NICTA**, Sydney I was a researcher in the machine learning research group where I collaborated with geophysicists, physicists and ecologists to apply cutting edge machine learning techniques and practices to their research. This was as part of a three year Science and Industry Endowment Fund (SIEF) project, and it led to me developing novel Gaussian processes inference techniques for inverse problems.
2012–2013	Research Associate, **Australian Centre for Field Robotics**, Sydney As a research associate my duties included researching unsupervised methods for exploring and visualising sea-floor and water-column data gathered by autonomous underwater vehicles, ship-borne multi-beam sonar, towed video systems etc. These unsupervised algorithms aid scientists in exploring and categorising their data, and have also been used to aid survey planning.
2006–2010	Control and Software Engineer, **BioPower Systems**, Sydney I worked for this start-up renewable energy company as an undergraduate-intern and on a casual basis. I initially helped design and manufacture a control system for a small-scale tidal energy power generation prototype (BioStream). I also worked on sensor data retrieval, visualisation and logging systems for the full-scale prototype devices.

Research

A common thread to my research is in using, or developing, machine learning models to infer latent phenomena. This includes causal effects from observational data, unobservable or partially observable physical processes from related observable processes (e.g. inverse problems), and semantic structure in digital images.

One focus of research during my time at Gradient Institute has been observational causal inference. I completed a number of observational studies for an Australian school jurisdiction to understand how the school system and personal well-being effect educational outcomes for students. These studies had numerous confounding factors (hundreds of factors) and non-linear interactions. This justified the use of machine learning methods for

causal inference such as direct response surface modelling (S-learners) and double machine learning (R-learners). This work required new methods and software for reporting nonlinear causal effects. A publication on this work is in Scientific Reports [1], and open source software implementing the methodology was released (Causal Inspection). These studies also prompted a brief investigation into the effects of regularisation on estimation bias in the presence of collinearity, the results are summarised in this blog post.

I was involved in numerous large-scale spatial inference projects during my time in Data61. These included depth of cover and element concentration prediction projects with Geoscience Australia and CSIRO Minerals. I was also responsible for a number of national scale, high resolution, predictive soil mapping projects with CSIRO Agriculture and Food. These projects required development of scalable machine learning techniques (Gaussian processes, Bayesian neural nets, etc) and pipelines (for example Landshark below).

During my time at NICTA my research mainly focused on methods for variational inference in Bayesian models with highly non-linear, and sometimes non-differentiable, likelihood functions [6, 7]. These models have potential applications in Geoscience, i.e., performing geophysical inversions.

My PhD research focused on unsupervised scene understanding. The motivation for this research came from the ability of robotic agents to collect large volumes of visual seafloor data in novel environments at a rate that far out-paces human annotation of this data. I created hierarchical Bayesian models using variational Bayes to cluster images into like scenes [9, 10], and then extended these models to incorporate additional background context such as the photo album (or survey) an image appears in [13, ch. 4]. I then further extended these models to infer objects within images while utilising contextual information about the image's scene and album [8, 5].

Software Engineering Skills and Projects

Programming:	Python, C++/C, Matlab.
Frameworks:	Scikit Learn, Pandas, PyTorch, TensorFlow, Numpy/Scipy, Eigen.
Systems:	GNU Unix/Linux, Google Cloud/AWS, HPC systems, continuous integration, Git.

I have been involved in and led numerous open source machine learning projects, please see github.com/dsteinberg for a full list, highlights include:

Libcluster (⋆139, Y21) An extensible C++ library of hierarchical Bayesian models for clustering and scene understanding with variational Bayes inference.

Aboleth (⋆127, Y10) Aboleth provides a set of high performance and light weight components for building Bayesian neural nets and approximate (deep) Gaussian process computational graphs.

Landshark (⋆9, Y10) A set of python command line tools for supervised learning problems on large spatial raster datasets using distributed computing. It uses TensorFlow machine learning pipelines.

revrand (⋆57, Y13) A python library of scalable Bayesian generalised linear models with random approximate kernel features.

Causal Inspection A scikit-learn inspired python library for quantifying and visualising non-linear causal effects from S- and T-learners. This library allows various interpretable machine learning methods (such as permutation importance and partial dependence) to be run within a cross validation or bootstrap sampling pipeline. In this way we can quantify the uncertainty in non-linear and binary effects for arbitrary machine learning estimators.

twostageridge A simple python two-stage ridge regression model for causal inference. It allows for the confounding variables to be regularised, whilst minimising bias on the estimated treatment effect.

Select Publications and Statistics

Journal Articles

[1] D. Cárdenas, F. Lattimore, D. Steinberg, and K. J. Reynolds. "Youth well-being predicts later academic success". In: *Scientific reports* 12.1 (2022), pp. 1–13.

[2] J. S. Camac, R. Condit, R. G. FitzJohn, L. McCalman, D. Steinberg, M. Westoby, S. J. Wright, and D. S. Falster. "Partitioning mortality into growth-dependent and growth-independent hazards across 203 tropical tree species". In: *Proceedings of the National Academy of Sciences* 115.49 (2018), pp. 12459–12464.

[3] N. Butterworth, D. Steinberg, R. Müller, S. Williams, A. Merdith, and S. Hardy. "Tectonic environments of South American porphyry copper magmatism through time revealed by spatiotemporal data mining". In: *Tectonics* 35.12 (2016), pp. 2847–2862.

[4] S. B. Williams, O. Pizarro, D. M. Steinberg, A. Friedman, and M. Bryson. "Reflections on a decade of autonomous underwater vehicles operations for marine survey at the Australian Centre for Field Robotics". In: *Annual Reviews in Control* 42 (2016), pp. 158–165.

[5] D. M. Steinberg, O. Pizarro, and S. B. Williams. "Hierarchical Bayesian Models for Unsupervised Scene Understanding". In: *Journal of Computer Vision and Image Understanding (CVIU)* 131 (2015), pp. 128–144.

Conference Papers

[6] E. V. Bonilla, D. M. Steinberg, and A. Reid. "Extended and Unscented Random Kitchen Sinks". In: *International Conference on Machine Learning (ICML)*. 2016.

[7] D. M. Steinberg and E. V. Bonilla. "Extended and Unscented Gaussian Processes". In: *Advances in Neural Information Processing Systems (NeurIPS)*. (Awarded a spotlight presentation). Montreal, Canada, 2014.

[8] D. M. Steinberg, O. Pizarro, and S. B. Williams. "Synergistic Clustering of Image and Segment Descriptors for Unsupervised Scene Understanding". In: *International Conference on Computer Vision (ICCV)*. Darling Harbour, Sydney: IEEE, 2013.

[9] D. M. Steinberg, A. Friedman, O. Pizarro, and S. B. Williams. "A Bayesian Nonparametric Approach to Clustering Data from Underwater Robotic Surveys". In: *International Symposium on Robotics Research*. Flagstaff, AZ, 2011.

[10] D. M. Steinberg, S. B. Williams, O. Pizarro, and M. V. Jakuba. "Towards Autonomous Habitat Classification Using Gaussian Mixture Models". In: *Proceedings of Intelligent Robotics and Systems (IROS)*. IEEE/RSJ. Taipei, 2010.

[11] A. Bender, D. M. Steinberg, A. L. Friedman, and S. B. Williams. "Analysis of an autonomous underwater glider". In: *Proceedings of the Australasian conference on robotics and automation (ACRA)*. 2008, pp. 1–10.

Reports/Chapters

[12] H. Durrant-Whyte, L. McCalman, S. O'Callaghan, A. Reid, and D. Steinberg. "The Impact of Computerisation and Automation on Future Employment." In: *Australia's Future Workforce?* Committee for Economic Development of Australia (CEDA), June 2015. Chap. 1.4.

Thesis

[13] D. M. Steinberg. "An Unsupervised Approach to Modelling Visual Data". PhD thesis. Australian Centre for Field Robotics, The University of Sydney, 2013.

Publication Statistics

Recorded on 06/06/2023. Please see my google scholar page for more information.

Total number of (accepted) publications	30
Primary author publications	9
Citations	750
h-index	16
i10-index	18