Jack Baker

24 Kelsey Street, Lancaster, LA15DL
jackbaker92@mail.com +447989901918
https://github.com/jbaker92/
https://lancs.ac.uk/~bakerj1/

Data scientist educated to PhD level in computational statistics and machine learning. Experience developing machine learning models in Python, R and TensorFlow. Relevant internships which required working with SQL, Hadoop architecture and Python data science tools.

Technical Summary

Expert: statistics, machine learning, Python, R.

Proficient: TensorFlow, unix/shell, version control (git), scikit-learn.

Familiar: SQL, Hadoop, C, NLTK.

Experience

Travelling in Australia Jan 2019 – Apr 2019

PhD Student, Lancaster University & University of Washington Oct 2015 – Dec 2018

- PhD student in computational statistics and machine learning at STOR-i, a leading centre for doctoral training in statistics in the UK. PhD in collaboration with University of Washington.
- Submitted thesis consists of three (first author) publications accepted to leading journals in machine learning, computational statistics and statistical software (see overleaf).
- Worked on training Bayesian models at large scale. Developed complex models in Python/R; e.g. recommender systems, Bayesian neural networks, and Dirichlet processes (see GitHub).
- Developed an R package in TensorFlow: sgmcmc (see publications).
- Tutored data science & statistics students at Lancaster University and Edinburgh University.

Data Science Intern, Improve Digital July 2016 – Oct 2016

- Analysed customer advertising data to support data tech team.
- Included topic modelling, classification, forecasting and simulation modelling in Python.
- Data stored in a Hadoop architecture, accessed using Python, SQL and ibis.
- Analysis enabled tech team to roll out a new feature with confidence in performance.

Data Science Intern, Met Office Jun 2013 – Aug 2013

- Analysed the performance and usage of the 100TB per day MASS storage system.
- Developed a web interface for visualisation of key metrics. Used by the climate IT team.

Development Intern, Motorola Solutions Jul 2011 – Aug 2011

- Developed an efficient compression algorithm in C for the data on Motorola's wireless products.

Other Professional Experience

Contributor: Data science tutorials at Cambridge Spark (2018, see website).

Invited Speaker: University of Warwick (2018); University of Edinburgh (2017).

Invited Reviewer: Leading conferences on AI/Machine learning: NIPS (2018); UAI (2018).

Awarded Grants: Extended research visit to University of Washington (2017); expenses for industrial data science consulting week, Alan Turing Institute (2016); travel to present at international Bayesian conference, Switzerland (2016).

Chair: Talks and tutorials on tech for the STOR-i institute (2015–2018, see website).

IP Market Research: Voluntary project for 4 months, University Enterprise Centre (2016).

Outreach Lead: Various impact sessions for schools (2018, see website).

Contributed Speaker: Numerous statistics conferences (2015–2018, see website).

Keen Runner: run two marathons, a mountain marathon, and numerous races (2008–2018)

Education

PhD, Computational Statistics & Machine Learning Oct 2015 – Dec 2018
Lancaster University & University of Washington
Full scholarship and stipend. STOR-i Centre for Doctoral Training.

MRes, Statistics and Operational Research. Distinction (81%) Oct 2014 – Sep 2015
Lancaster University
Full scholarship and stipend. STOR-i Centre for Doctoral Training.

BSc, Mathematics and Statistics. First Class Honours (83%) Sep 2010 – July 2014
Edinburgh University

Publications

J. Baker, P. Fearnhead, E. B. Fox and C. Nemeth (2018), Large-Scale Stochastic Sampling from the Probability Simplex. Advances in Neural Information Processing Systems 31.

J. Baker, P. Fearnhead, E. B. Fox and C. Nemeth (2018), Control Variates for Stochastic Gradient MCMC. Statistics and Computing.

J. Baker, P. Fearnhead, E. B. Fox and C. Nemeth (2017), sgmcmc: An R Package for Stochastic Gradient Markov Chain Monte Carlo. Journal of Statistical Software (to appear).