# Kenyon Ng

Statistician. Machine Learner.

Email: kenyon.ng@gmail.com
Phone: +61 4 6642 4353

LinkedIn: linkedin.com/in/kenyon-ng
GitHub: github.com/weiyaw
Webpage: kenyon.xyz

## Education

**Master of Philosophy**, The University of Western Australia **Aug 2017 – Jul 2019**

*Dean's list, among the top 5% of theses examined by external examiners*
*Major: Mathematical Statistics (supervised by Berwin Turlach & Kevin Murray)*
*Thesis: Shape-constrained regression for nonparametric longitudinal and flexible parametric modelling*

- Modelled growth curves (e.g. human growth) with Bayesian penalised splines
- Derived fast, efficient and parallelisable MCMC methods for model fitting; achieved 20 times improvement in computation time compare to fitting with Stan
- Derived novel optimisation algorithms to fit constrained regression models
- Builded open-source R packages to implement MCMC and optimisation algorithms

**Bachelor of Science (Honours)**, The University of Western Australia **Aug 2013 – Jul 2017**

First Class Honours, WAM: 85.88

Bachelor of Science (Majors: Engineering, Applied Maths & Statistics), WAM: 87.46

**Deep Learning Specialization, Course 1, 2 & 3** by deeplearning.ai on Coursera
**Algorithms Specialization, Course 1** by Stanford University on Coursera
**Programming Languages, Part A, B & C** by University of Washington on Coursera
**Introduction to Java Programming, Part 1 & 2** by HKUST on edX

## Work Experience

**Research Scientist**, Department of Primary Industries and Regional Development, Perth **Nov 2019 – Present**

- Research and implement semiparametric models to predict grain flowering dates with data collected from agronomy trials
- Analyse trial data for the DPIRD frost team with linear mixed-effect models
- Develop and maintain R packages for analysing agriculture data
- Develop R-shiny apps to present research output to stakeholders

**Lab Demonstrator**, The University of Western Australia **Aug 2016 – Present**

- Explain statistical concepts in plain English to students (class size 20~100) with limited mathematics and statistics background
- Perform administrative duties and organise replacement labs
- Review lecture slides and suggest improvements
- Mark assignments, tests and final exams

**Data Analyst**, Sustainable Platform Pty Ltd, Perth **Feb 2018 – Apr 2018**

- Researched NLP algorithms to extract appropriate features from text corpora
- Developed regression models to predict sustainability score of a given company
- Scraped data from webpages and annual reports, and performed data cleaning
- Analysed listed companies' profiles with proprietary algorithms to determine sustainability scores
- Optimised and maintained existing R and Python code

## Scholarships and Awards

| | |
|---|---|
| Dean's list (top 5% master thesis judged by external examiners) | 2019 |
| Australian Government Research Training Program Stipend Scholarship | 2017 – 2019 |
| Statistical Society of Australia (WA branch) Honours Scholarship | 2017 |
| Turnbull Foundation Scholarship | 2015 |
| Abraham Wald Prize in Probability and Mathematical Statistics | 2014 |

## Publications

Ng, K., Turlach, B.A. and Murray, K., 2019. A flexible sequential Monte Carlo algorithm for parametric constrained regression. Computational Statistics & Data Analysis, 138, pp.13-26.

## Conferences

“A flexible sequential Monte Carlo algorithm for shape-constrained regression”, at

1. BioInfoSummer 2018, Australian Mathematical Science Institute, Perth, Dec 2018;
2. Faculty Research Conference, The University of Western Australia, Perth, Oct 2018;
3. Royal Statistical Society 2018, Cardiff, Sep 2018;
4. May Meeting, Statistical Society of Australia (WA Branch), Perth, May 2018.

## Skills

**Computer and Programming Skills**

Languages : R, Python, C++, Stan, Java, SAS
Packages : ggplot2, R-Shiny, Rcpp, Tensorflow
Miscellaneous : Git, PostgreSQL, Bash, Openstack cloud

**Statistics and Machine Learning**

Regression : generalised linear regression, nonparametric models, constrained regression
Classification : k-nearest neighbours, logistic regression, discriminant analysis, neural networks
Clustering : k-means, Gaussian mixture models
Modelling : Bayesian statistics, time series, longitudinal modelling, PCA, penalised and robust regression
Diagnostic : k-fold cross-validation, diagnostic plots (e.g. QQ plots)
Optimisation : simulated annealing, gradient descent, Newton’s methods, EM algorithm
MCMC : Metropolis-Hastings, Gibbs, Hamiltonian Monte Carlo, sequential Monte Carlo

**Languages**

Mandarin (Native), English (Fluent), Malay (Intermediate)

## Referees

Available upon request