AUSVET

Investigating the mathematics behind statistical modelling

Vincent Tian

Prepared for Ausvet
09/07/2019

Contents

Summary

A statistical model is a mathematical model built upon a set of statistical assumptions regarding the generation of a data set. Linear models, generalised linear models (GLMs) and generalised additive models (GAMs) are common statistical models used to investigate potential relationships between covariates in a dataset. Constructed in R or Python, an understanding of the fundamental mathematics holding the statistical model is often not required as various libraries, methods and classes have been written to account for this. Thus, the main pre-requisites for the user includes a basic understanding of various statistical models, which methods to use, how to choose and compare models and how to understand the diagnostics and properly interpret statistical indicators before refining the final model.

The aim of this project is to understand the mathematics and statistics used to build these models and then to construct them in Python. Many methods have already been implemented in libraries such as finding the prediction coefficients to minimise the residual errors, r-squared and adjusted r-squared values, mean squared error, residual standard errors and prediction errors for coefficients. These will all be coded from scratch with minimal use of existing python libraries. The end result of the project will have a statistical model on a data set which returns all the required statistical indicators and also allows the user to modify the model for deeper analysis.

1 Supplier's details

Full Legal Organisation Name:	Ausvet Pty Ltd
Legal Status:	☐ Individual/Sole Trader ☐ Partnership X Company ☐ Sole Director Company ☐ Trust (see note below) ☐ Other (please state):
NOTE: If the Potential Supplier is trading as a trust, please provide details of the relevant trust (and trustee) including a copy of the relevant trust deed (including any variations to that deed) as an attachment to this Response.	
Australian Business Number (ABN):	64 613 142 939
Australian Company Number (ACN):	613 142 939
Australian Registered Body Number (ARBN):	
Registered Address:	5 Shuffrey St, Fremantle WA 6160 Australia
Web address:	www.ausvet.com.au
Is your organisation 50% or more Indigenous owned?	☐ Yes, see below. X No
If your organisation is 50% or more Indigenous owned, is your organisation registered on Supply Nation?	☐ Yes ☐ No – see note below X Not Applicable
Please provide a certificate or letter from a recognised Indigenous organisation such as Land Council, Indigenous Chamber of Commerce or Office of the Registrar of Indigenous Corporations verifying Indigenous ownership.	
Has your organisation ever had a judicial decision about employee entitlements or engaged in practices that have been found to be dishonest, unethical or unsafe?	☐ Yes, see below. X No
If yes, what was the date of discharge? *The Supplier acknowledges that the giving of false or misleading information to the Commonwealth is a serious offence under section 137.1 of the schedule to the Criminal Code Act 1995 (Cth).*	(dd-mm-yyyy) *Note: The Customer cannot enter a contract with a supplier who has an undischarged judicial decision relating to employee entitlements.*

Contact Officer

For matters relating to this Response contact:

Name:	Vincent Tian
Position Title:	R&D Intern
Mobile:	0426 803 733
Email Address:	vincent@ausvet.com.au
Postal Address:	5 Shuffrey St, Fremantle WA 6160 Australia

Address for Notices (if different from the Contact Officer)

Name:	As above
Position Title:	
Email Address:	
Postal Address:	

Contract Manager (if different from the Contact Officer)

For matters of a general nature, including acceptance and issuance of written notices contact:

Name:	Dr Rohan Sadler
Position Title:	Statistician
Mobile:	0491 257 446
Email Address:	rohan.sadler@ausvet.com.au
Postal Address:	5 Shuffrey St, Fremantle WA 6160 Australia

Appendix A CVs of key personnel

VINCENT TIAN

Mathematics and Computer Science Major

+61 426 803 733 | vincent.tian72@gmail.com | github.com/vtian72 | linkedin.com/in/vincentian

ABOUT ME

Programming Python, Java, C
Analytics R, SQL, SAS, Excel
Languages English, Mandarin

EDUCATION

University of Western Australia — *Perth, Australia*
Bachelor of Science: Mathematics and Statistics, Computer Science — Feb 2017 – Jun 2020
- **Grade:** Distinction

University College London — *London, UK*
Exchange Program: Statistics — Jan 2019 – Jun 2019
- **Grade:** First Class Honours

Rossmoyne Senior High School — *Perth, Australia*
High School — Feb 2012 – Nov 2016
- **ATAR:** 99.4
- WACE Certificate of Distinction

WORK EXPERIENCE

Ausvet — *Perth, Australia*
Upcoming Data Science Intern — June 2019 – Present
- 1-month internship in Ausvet's research and development team
- Developing data science ideas and technology stacks in veterinary and human epidemiology

Philight Software International — *Perth, Australia*
Consultant Intern — Aug 2018 – Dec 2018
- Implemented C and Python algorithms to automate various data sorting processes
- Developed data models through Sharperlight for data migration from Oracle to SQL server
- Documented and utilised Sharperlight to build SQL queries for online based reporting

EXTRA-CURRICULUM ACTIVITIES

Projects — Jun 2017 - Present
- ***Haemoglobin levels in Afghanistan Women***
  - Used scientific literature to perform initial exploratory analysis on dataset
  - Built and optimised linear models and generalised linear models using statistical diagnostics in R
  - Predicted the haemoglobin levels for an unknown set of Afghanistan women using the chosen model
- ***Statistics and Trading Indicators***
  - Developed Python scripts for Sharpe Ratio calculation and technical indicators such as Bollinger Bands
  - Worked with python libraries such as numpy, pandas, sklearn and matplotlib
  - Implemented various probability distributions, Spearman and Pearson correlation coefficients in Java
- ***Jarvis***
  - Created a bot in C to provide daily weather updates and worked with the Openweathermap API