MAST90083 Computational Statistics and Data Mining

Credit Points:	12.5
Level:	9 (Graduate/Postgraduate)
Dates & Locations:	2015, Parkville This subject commences in the following study period/s: Semester 2, Parkville - Taught on campus.
Time Commitment:	Contact Hours: Contact Hours: 36 hours comprising 2 one-hour lectures per week and 1 one-hour practice class per week. Total Time Commitment: Estimated Total Time Commitment - 170 hours

Prerequisites:

One of

Subject	Study Period Commencement:	Credit Points:
MAST30025 Linear Statistical Models	Semester 1	12.50
MAST30027 Modern Applied Statistics	Semester 2	12.50

Corequisites:	None

Recommended Background Knowledge:

Subject	Study Period Commencement:	Credit Points:
MAST90082 Mathematical Statistics	Semester 1	12.50

Non Allowed Subjects:	None
Core Participation Requirements:	For the purposes of considering request for Reasonable Adjustments under the Disability Standards for Education (Cwth 2005), and Student Support and Engagement Policy, academic requirements for this subject are articulated in the Subject Overview, Learning Outcomes, Assessment and Generic Skills sections of this entry. It is University policy to take all reasonable steps to minimise the impact of disability upon academic study, and reasonable adjustments will be made to enhance a student's participation in the University's programs. Students who feel their disability may impact on meeting the requirements of this subject are encouraged to discuss this matter with a Faculty Student Adviser and Student Equity and Disability Support: http://services.unimelb.edu.au/disability
Coordinator:	Dr Guoqi Qian
Contact:	Email: **qguoqi@unimelb.edu.au (mailto:qguoqi@unimelb.edu.au)**
Subject Overview:	Computing techniques and data mining methods are indispensible in modern statistical research and applications, where "Big Data" problems are often involved. This subject will introduce a number of recently developed statistical data mining methods that are scalable to large datasets and high-performance computing. These include regularized regression such as the Lasso; tree based methods such as bagging, boosting and random forests; and support vector machines. Important statistical computing algorithms and techniques used in data mining will be explained in detail. These include the bootstrap, cross-validation, the EM algorithm, and Markov chain Monte Carlo methods including the Gibbs sampler and Metropolis-Hastings algorithm.
Learning Outcomes:	After completing this subject students should gain: # an understanding of theory and computing of modern statistics and how they are implemented in applications; # the skills of using nonparametric and Monte Carlo methods in statistics; and

	# the ability to pursue further studies in this and related areas.
Assessment:	Up to 40 pages of written assignments (two assignments worth 10% each) due mid and late semester (20%) Three hour written examination (80%)
Prescribed Texts:	None
Breadth Options:	This subject is not available as a breadth subject.
Fees Information:	Subject EFTSL, Level, Discipline & Census Date, http://enrolment.unimelb.edu.au/fees
Generic Skills:	In addition to learning specific skills that will assist students in their future careers in science, they will have the opportunity to develop generic skills that will assist them in any future career path. These include: # problem-solving skills: the ability to engage with unfamiliar problems and identify relevant solution strategies; # analytical skills: the ability to construct and express logical arguments and to work in abstract or general terms to increase the clarity and efficiency of analysis; # collaborative skills: the ability to work in a team; # time-management skills: the ability to meet regular deadlines while balancing competing commitments
Related Course(s):	Master of Philosophy - Engineering Master of Science (Mathematics and Statistics) Ph.D.- Engineering
Related Majors/Minors/ Specialisations:	Mathematics and Statistics