COMP7220

Data Science and Machine Learning

Session 1, Weekday attendance, North Ryde 2020

Department of Computing

Contents

General Information	2
Learning Outcomes	3
General Assessment Information	3
Assessment Tasks	3
Delivery and Resources	3
Unit Schedule	4
Policies and Procedures	5
Changes from Previous Offering	7

Disclaimer

Macquarie University has taken all reasonable measures to ensure the information in this publication is accurate and up-to-date. However, the information may change or become out-dated as a result of change in University policies, procedures or rules. The University reserves the right to make changes to any information in this publication without notice. Users of this publication are advised to check the website version of this publication [or the relevant faculty or department] before acting on any information in this publication.

General Information

Unit convenor and teaching staff

Lecturer, Convenor
Mark Dras
mark.dras@mq.edu.au
Contact via by email
4RPD, room 208
by appointment

Lecturer
Rolf Schwitter
rolf.schwitter@mq.edu.au
Contact via by email
4RPD, room 359
by appointment

Credit points
10

Prerequisites
Admission to MRes

Corequisites

Co-badged status
COMP8220

Unit description
This unit begins with conventional machine learning techniques for constructing classifiers and regression models, including widely applicable standard techniques such as Naive Bayes, decision trees, logistic regression and support vector machines (SVMs); in this part, given required prior knowledge of machine learning, we focus on more advanced aspects. We then look in detail at deep learning and other state-of-the-art approaches. We discuss in detail the advantages and disadvantages of each method, in terms of computational requirements, ease of use, and performance, and we study the practical application of these methods in a number of use cases.

Important Academic Dates

Information about important academic dates including deadlines for withdrawing from units are available at https://www.mq.edu.au/study/calendar-of-dates

Learning Outcomes

On successful completion of this unit, you will be able to:

ULO1: Derive algorithms to solve machine learning problems based on an understanding of how machine learning and data science problems are mathematically formulated and analysed.

ULO3: Analyse real-world data science problems, identify which methods are appropriate, organise the data appropriately, apply one or more methods, and evaluate the quality of the solution.

ULO2: Create machine learning solutions to data science problems by identifying and applying appropriate algorithms and implementations.

ULO4: Evaluate one or more approaches to advanced topics in machine learning and data science and report the findings in oral and written form.

Assessment Tasks

Coronavirus (COVID-19) Update

Assessment details are no longer provided here as a result of changes due to the Coronavirus (COVID-19) pandemic.

Students should consult iLearn for revised unit information.

Find out more about the Coronavirus (COVID-19) and potential impacts on staff and students

General Assessment Information

Late Submission

No extensions will be granted without an approved application for Special Consideration. There will be a deduction of 20% of the total available marks made from the total awarded mark for each 24 hour period or part thereof that the submission is late. For example, 25 hours late in submission for an assignment worth 10 marks – 40% penalty or 4 marks deducted from the total. No submission will be accepted after solutions have been posted.

Delivery and Resources

Coronavirus (COVID-19) Update

Any references to on-campus delivery below may no longer be relevant due to COVID-19.

Please check here for updated delivery information: https://ask.mq.edu.au/account/pub/display/unit_status

- **Classes**: The first half of each class will have a seminar/lecture format that will introduce the material for the week, while the second half of the class will involve practical lab work applying the ideas and concepts introduced in the first half of the class. You should bring along your own device to the second half of the class.
- **Textbook**: The main textbook for the unit is Aurélien Géron (2019) "Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow" (2nd edition; September 2019). This is available through the MQ library (MQ has an arrangement with publisher O'Reilly: you can register at O'Reilly using your MQ email, and get access to the book there). The book comes with source code that is available from https://github.com/ageron/handson-ml2. A supplementary source of material for a deeper understanding of the theoretical material is Trevor Hastie, Robert Tibshirani and Jerome Friedman (2009; corrected 12th printing Jan 2017) "The Elements of Statistical Learning: Data Mining, Inference, and Prediction." A freely downloadable pdf is available at the first author's webpage.

Background Material

- The unit requires a sound background in programming, and particularly Python. If you feel you need a refresher on Python (or an introduction from scratch, as long as you're a quick and independent learner), there's a popular tutorial at http://learnpython.org/. This goes all the way from basic programming to the mathematical and data science libraries used by Python, like numpy and pandas. There's also the resources at the Python website at python.org, like the Beginner's Guide.
- For a refresher on linear algebra as it is relevant to machine learning, Jason Brownlee (2018) "Basics of Linear Algebra for Machine Learning" has useful material that's linked to Python data structures. There's a free downloadable pdf available.

Unit Webpage and Technology Used and Required

- **iLearn** is going to be used as a main web server for the unit.
- The programming language for the unit will be Python. The "conventional" machine learning section will use Python's scikit-learn, and the deep learning section will use TensorFlow and Keras.
- For the most part, programming will be done via Jupyter notebooks. We'll typically be running these notebooks on Google Colab.

Unit Schedule

Coronavirus (COVID-19) Update

The unit schedule/topics and any references to on-campus delivery below may no longer be relevant due to COVID-19. Please consult iLearn for latest details, and check here for updated delivery information: https://ask.mq.edu.au/account/pub/display/unit_status

Week	Topic	Readings (from Géron)
1	What is Machine Learning?	Ch 1
2	Workflow of a Machine Learning Project	Ch 2
3	Classification and Regression	Ch 3-4
4	Support Vector Machines and Decision Trees	Ch 5-6
5	Ensemble Learning, Random Forests, and Dimensionality Reduction	Ch 7-8
6	Handling Text Data	supplementary notes
7	public holiday	
8-9	Introduction to Artificial Neural Networks: • ANN basics • Multi-Layer Perceptrons • The Tensorflow and Keras frameworks	Ch 10-11
10-11	Deep Neural Networks • The structure of deep NNs • Convolutional NNs • Practical issues in training NNs	Ch 11-14, supplementary notes
12-13	NNs for sequences, and advanced topics: • Recurrent NNs • Autoencoders • Reinforcement Learning	Ch 15 and onwards

Policies and Procedures

Macquarie University policies and procedures are accessible from Policy Central (https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central). Students should be aware of the following policies in particular with regard to Learning and Teaching:

- Academic Appeals Policy
- Academic Integrity Policy
- Academic Progression Policy
- Assessment Policy

- Fitness to Practice Procedure
- Grade Appeal Policy
- Complaint Management Procedure for Students and Members of the Public
- Special Consideration Policy *(**Note:** The Special Consideration Policy is effective from 4 December 2017 and replaces the Disruption to Studies Policy.)*

Students seeking more policy resources can visit the Student Policy Gateway (https://students.mq.edu.au/support/study/student-policy-gateway). It is your one-stop-shop for the key policies you need to know about throughout your undergraduate student journey.

If you would like to see all the policies relevant to Learning and Teaching visit Policy Central (https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central).

Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: https://students.mq.edu.au/study/getting-started/student-conduct

Results

Results published on platform other than eStudent, (eg. iLearn, Coursera etc.) or released directly by your Unit Convenor, are not confirmed as they are subject to final approval by the University. Once approved, final results will be sent to your student email address and will be made available in eStudent. For more information visit ask.mq.edu.au or if you are a Global MBA student contact globalmba.support@mq.edu.au

Student Support

Macquarie University provides a range of support services for students. For details, visit http://students.mq.edu.au/support/

Learning Skills

Learning Skills (mq.edu.au/learningskills) provides academic writing resources and study strategies to help you improve your marks and take control of your study.

- Getting help with your assignment
- Workshops
- StudyWise
- Academic Integrity Module

The Library provides online and face to face support to help you find and use relevant information resources.

- Subject and Research Guides
- Ask a Librarian

Student Services and Support

Students with a disability are encouraged to contact the Disability Service who can provide appropriate help with any issues that arise during their studies.

Student Enquiries

For all student enquiries, visit Student Connect at ask.mq.edu.au

If you are a Global MBA student contact globalmba.support@mq.edu.au

IT Help

For help with University computer systems and technology, visit http://www.mq.edu.au/about_us/offices_and_units/information_technology/help/.

When using the University's IT, you must adhere to the Acceptable Use of IT Resources Policy. The policy applies to all who connect to the MQ network including students.

Changes from Previous Offering

The topics are broadly similar to 2019 (which were changed significantly from 2018 and earlier). The assessment, however, is different: in 2020 there is both a major project focussing on a predefined dataset, and an individual project requiring dataset selection.