COMP7220

Data Science and Machine Learning

Session 1, In person-scheduled-weekday, North Ryde 2022

School of Computing

Contents

General Information

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

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

Lecturer
Fred Amouzgar
fred.amouzgar@mq.edu.au
Contact via via email
by appointment

Tutor
David Warren
david.warren@mq.edu.au
Contact via via email
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.

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

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.

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

General Assessment Information

Late Submission

Late submissions **will not be accepted** without an approved Special Consideration request. Assessments submitted after the due date will receive a mark of **zero**.

Assessment Tasks

Name	Weighting	Hurdle	Due
Major Project	40%	No	Initial: end first week of break; final: week 13

Name	Weighting	Hurdle	Due
Individual Project	30%	No	Initial: week 6; final: week 13
Practical Exercises	30%	No	Throughout semester (see iLearn)

Major Project

Assessment Type [1]: Project
Indicative Time on Task [2]: 30 hours
Due: **Initial: end first week of break; final: week 13**
Weighting: **40%**

The student will apply knowledge of conventional machine learning and deep learning to design and implement a solution to a (classification or other) task on a defined dataset. The deliverables will be the implementation and a report describing this implementation.

On successful completion you will be able to:

- Derive algorithms to solve machine learning problems based on an understanding of how machine learning and data science problems are mathematically formulated and analysed.
- Create machine learning solutions to data science problems by identifying and applying appropriate algorithms and implementations.
- 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.
- Evaluate one or more approaches to advanced topics in machine learning and data science and report the findings in oral and written form.

Individual Project

Assessment Type [1]: Project
Indicative Time on Task [2]: 25 hours
Due: **Initial: week 6; final: week 13**
Weighting: **30%**

In contrast to the Major Project, in this one the student will select a dataset from an appropriate domain, and then design and implement a solution to a task on this chosen dataset. The deliverables will be the implementation and a report describing this implementation.

On successful completion you will be able to:

- Derive algorithms to solve machine learning problems based on an understanding of how machine learning and data science problems are mathematically formulated and

analysed.

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

- 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.

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

Practical Exercises

Assessment Type [1]: Problem set
Indicative Time on Task [2]: 30 hours
Due: **Throughout semester (see iLearn)**
Weighting: **30%**

These will consist of practical exercises set throughout the semester.

On successful completion you will be able to:

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

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

---

[1] If you need help with your assignment, please contact:

- the academic teaching staff in your unit for guidance in understanding or completing this type of assessment

- the Writing Centre for academic skills support.

[2] Indicative time-on-task is an estimate of the time required for completion of the assessment task and is subject to individual variation

Delivery and Resources

Classes

- **Classes**: There will be a two hour lecture each week, and additionally a small class that will focus on working through practical tasks.

- **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. (The book used to have a freely available pdf, but this seems to have disappeared. It is published by Machine Learning Mastery.)

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

Week	Topic	Readings (from Géron)
1	What is Machine Learning?	Ch 1
2	Workflow of a Machine Learning Project	Ch 2
3	Support Vector Machines and Decision Trees	Ch 3-6

Week	Topic	Readings (from Géron)
4	Ensemble Learning, Random Forests, and Dimensionality Reduction	Ch 7-8
5	Handling Text Data	supplementary notes
6-7	Introduction to Artificial Neural Networks: • ANN basics • Multi-Layer Perceptrons • The Tensorflow and Keras frameworks	Ch 10-11
8-9	Deep Neural Networks • The structure of deep NNs • Convolutional NNs • Practical issues in training NNs	Ch 11-14, supplementary notes
10	NNs for sequences, and advanced topics: • Recurrent NNs • Autoencoders	Ch 15 and onwards, supplementary notes
11-12	Reinforcement Learning	supplementary notes
13	Unit review	

Policies and Procedures

Macquarie University policies and procedures are accessible from Policy Central (https://policies.mq.edu.au). 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
- Assessment Procedure
- Complaints Resolution Procedure for Students and Members of the Public
- Special Consideration Policy

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

To find other policies relating to Teaching and Learning, visit Policy Central (https://policies.mq.e

du.au) and use the search tool.

Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: https://students.mq.edu.au/admin/other-resources/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

Academic Integrity

At Macquarie, we believe academic integrity – honesty, respect, trust, responsibility, fairness and courage – is at the core of learning, teaching and research. We recognise that meeting the expectations required to complete your assessments can be challenging. So, we offer you a range of resources and services to help you reach your potential, including free online writing and maths support, academic skills development and wellbeing consultations.

Student Support

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

The Writing Centre

The Writing Centre provides resources to develop your English language proficiency, academic writing, and communication skills.

- Workshops
- Chat with a WriteWISE peer writing leader
- Access StudyWISE
- Upload an assignment to Studiosity
- Complete the 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

Macquarie University offers a range of Student Support Services including:

- IT Support

- Accessibility and disability support with study

- Mental health support

- Safety support to respond to bullying, harassment, sexual harassment and sexual assault

- Social support including information about finances, tenancy and legal issues

Student Enquiries

Got a question? Ask us via AskMQ, or contact Service Connect.

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 late submission rule was changed to align with the new Faculty policy.