COMP783

Data Science and Machine Learning

S1 Day 2019

Dept of Computing

Contents

General Information

Unit convenor and teaching staff
Convenor, lecturer
Mark Dras
mark.dras@mq.edu.au
Contact via email
TBA
by appointment

Lecturer
Jia Wu
jia.wu@mq.edu.au
Contact via email
TBA
by appointment

Credit points
4

Prerequisites
Admission to MRes

Corequisites

Co-badged status
ITEC873

Unit description
This unit introduces basic machine learning techniques for constructing classifiers and regression models, focusing on widely applicable standard techniques such as Naive Bayes, decision trees, logistic regression and support vector machines (SVMs), and also including more general advanced frameworks such as graphical models. We discuss in detail the advantages and disadvantages of each method, both in terms of computational requirements, ease of use and performance, and study their 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:

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.

Assessment Tasks

Name	Weighting	Hurdle	Due
Project proposal presentation	10%	No	Week 6
Project proposal document	10%	No	Week 7
Project presentation	10%	No	Week 13
Project report	45%	Yes	Week 15
Practical exercises	25%	No	during the semester

Project proposal presentation

Due: **Week 6**
Weighting: **10%**

This is an in-class presentation. It should provide an initial overview of the project, including in summary form the same information as in the project proposal document. The intention is that you will get feedback on your presentation that can be incorporated into the project proposal document.

You have to submit the slides of your presentation prior to the presentation via iLearn.

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.

Project proposal document

Due: **Week 7**
Weighting: **10%**

This proposal should provide the following information about the project:

- What is the goal of the project?
- Is there any relevant prior work? (A reference to sections in the textbook might be appropriate.)
- What data set(s) will be used for the project, and where will they be obtained from?
- What method(s) will be used to analyse the data? Why are these methods used?

You have to submit the project proposal via iLearn.

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.

Project presentation

Due: **Week 13**
Weighting: **10%**

This in-class presentation should cover all aspects of the project, including results and

conclusion.

You have to submit the slides of your presentation prior to the presentation via iLearn.

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.

Project report

Due: **Week 15**
Weighting: **45%**
This is a hurdle assessment task (see assessment policy for more information on hurdle assessment tasks)

This report should describe all aspects of the research project. It should have the format of a short scientific paper (between 5-8 pages long, plus additional pages of data or graphs if required). It should contain the following sections:

- Introduction (including a statement of the problem)
- Related work (if relevant)
- Description of the data (including data source)
- Methods
- Results
- Conclusion

This assessment task has a **hurdle requirement**: you will need to obtain at least 40% for the project report in order to pass the unit. If you obtain between 30% and 40% marks for the project report, then you will be given a second (and final) attempt to submit your report.

Your have to submit the project report via iLearn.

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

Due: **during the semester**
Weighting: **25%**

We expect to assign 4 practical exercises during the semester.

You have to submit the solutions to the practical exercises via iLearn.

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.

Delivery and Resources

Note that this unit assumes some prior knowledge of machine learning or a related discipline (e.g. statistics). Please contact the unit convenor if you are in doubt.

CLASSES

- **Classes**: Except for the 2 weeks of student presentations, 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 focus on practical application of the ideas and concepts introduced in the first half of the class.
- **Textbook**: The main textbook for the unit is Aurélien Géron (2017) "Hands-On Machine Learning with Scikit-Learn and TensorFlow". This is available as an e-book through the MQ library. The book comes with source code that is available from https://github.com/a

geron/handson-ml. 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 assumes some 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.
- For the most part, programming will be done via Jupyter notebooks.

Unit Schedule

Week	**Topic**	**Readings** (from Géron)
Week 1	Welcome to unit • Overview of unit topics and assessment • Review of machine learning	Ch 1
Weeks 2-3	Revisiting conventional machine learning • Machine learning frameworks, evaluation, etc • Focus on selected conventional ML method(s)	Ch 2-6 (selected topics)

Weeks 4-5	Advanced conventional ML topics, e.g.: • Natural Language Processing • Handling unbalanced data • Ensembles • Dimensionality reduction	Ch 7-8
Week 6	Student presentations	
Weeks 7-8	Introduction to Artificial Neural Networks: • ANN basics • Multi-Layer Perceptrons • The Tensorflow framework	Ch 9-10
Weeks 9-10	Deep Neural Networks • The structure of deep NNs • Convolutional NNs	Ch 11, 13
Weeks 11-12	NNs for sequences, and advanced topics: • Recurrent NNs • Autoencoders • Reinforcement Learning	Ch 14-16
Week 13	Student presentations	

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.)*

Undergraduate 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 (http

s://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

Late Submissions

No extensions will be granted. Students who have not submitted the task by the deadline will be awarded a zero mark for the task, except for cases in which an application for special consideration is made and approved.

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 improve your marks and take control of your study.

- Workshops
- StudyWise
- Academic Integrity Module for Students
- Ask a Learning Adviser

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.

Graduate Capabilities

PG - Capable of Professional and Personal Judgment and Initiative

Our postgraduates will demonstrate a high standard of discernment and common sense in their professional and personal judgment. They will have the ability to make informed choices and decisions that reflect both the nature of their professional work and their personal perspectives.

This graduate capability is supported by:

Learning outcomes

- 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.

Assessment tasks

- Project proposal presentation
- Project proposal document
- Project presentation
- Project report
- Practical exercises

PG - Discipline Knowledge and Skills

Our postgraduates will be able to demonstrate a significantly enhanced depth and breadth of knowledge, scholarly understanding, and specific subject content knowledge in their chosen fields.

This graduate capability is supported by:

Learning outcomes

- 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.

Assessment tasks

- Project proposal presentation
- Project proposal document
- Project presentation
- Project report
- Practical exercises

PG - Critical, Analytical and Integrative Thinking

Our postgraduates will be capable of utilising and reflecting on prior knowledge and experience, of applying higher level critical thinking skills, and of integrating and synthesising learning and knowledge from a range of sources and environments. A characteristic of this form of thinking is the generation of new, professionally oriented knowledge through personal or group-based critique of practice and theory.

This graduate capability is supported by:

Learning outcomes

- 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.

Assessment tasks

- Project proposal presentation
- Project proposal document
- Project presentation
- Project report

- Practical exercises

PG - Research and Problem Solving Capability

Our postgraduates will be capable of systematic enquiry; able to use research skills to create new knowledge that can be applied to real world issues, or contribute to a field of study or practice to enhance society. They will be capable of creative questioning, problem finding and problem solving.

This graduate capability is supported by:

Learning outcomes

- 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.

Assessment tasks

- Project proposal presentation
- Project proposal document
- Project presentation
- Project report
- Practical exercises

PG - Effective Communication

Our postgraduates will be able to communicate effectively and convey their views to different social, cultural, and professional audiences. They will be able to use a variety of technologically supported media to communicate with empathy using a range of written, spoken or visual formats.

This graduate capability is supported by:

Learning outcome

- 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

- Project proposal presentation
- Project proposal document
- Project presentation
- Project report

PG - Engaged and Responsible, Active and Ethical Citizens

Our postgraduates will be ethically aware and capable of confident transformative action in relation to their professional responsibilities and the wider community. They will have a sense of connectedness with others and country and have a sense of mutual obligation. They will be able to appreciate the impact of their professional roles for social justice and inclusion related to national and global issues

This graduate capability is supported by:

Learning outcomes

- 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.

Assessment tasks

- Project proposal presentation
- Project proposal document
- Project presentation
- Project report
- Practical exercises

Changes from Previous Offering

There are substantial changes compared to the offerings of 2018 and earlier. The unit now assumes some detailed technical knowledge relevant to machine learning (e.g. prior study on machine learning, knowledge of appropriate statistics, or other relevant background). Conventional machine learning will be reviewed in the first few weeks of the unit, but the majority of the unit now focuses on deep learning.

Assessment standards

COMP783 will be graded according to the following general descriptions of the letter grades as

specified by Macquarie University.

• High Distinction (HD, 85-100): Provides consistent evidence of deep and critical understanding in relation to the learning outcomes. There is substantial originality, insight or creativity in identifying, generating and communicating competing arguments, perspectives or problem solving approaches; critical evaluation of problems, their solutions and their implications; creativity in application as appropriate to the program.

• Distinction (D, 75-84): Provides evidence of integration and evaluation of critical ideas, principles and theories, distinctive insight and ability in applying relevant skills and concepts in relation to learning outcomes. There is demonstration of frequent originality or creativity in defining and analysing issues or problems and providing solutions; and the use of means of communication appropriate to the program and the audience.

• Credit (Cr, 65-74): Provides evidence of learning that goes beyond replication of content knowledge or skills relevant to the learning outcomes. There is demonstration of substantial understanding of fundamental concepts in the field of study and the ability to apply these concepts in a variety of contexts; convincing argumentation with appropriate coherent justification; communication of ideas fluently and clearly in terms of the conventions of the program.

• Pass (P, 50-64): Provides sufficient evidence of the achievement of learning outcomes. There is demonstration of understanding and application of fundamental concepts of the program; routine argumentation with acceptable justification; communication of information and ideas adequately in terms of the conventions of the program. The learning attainment is considered satisfactory or adequate or competent or capable in relation to the specified outcomes.

• Fail (F, 0-49): Does not provide evidence of attainment of learning outcomes. There is missing or partial or superficial or faulty understanding and application of the fundamental concepts in the field of study; missing, undeveloped, inappropriate or confusing argumentation; incomplete, confusing or lacking communication of ideas in ways that give little attention to the conventions of the program.

• Fail Hurdle (FH, 49): Student has obtained a raw mark over 50, yet failed all available attempts of at least one hurdle assessment (as described within Schedule 2: Unit Assessment Requirements).

The standards of achievement that will be used to assess each of the assessment tasks with respect to the letter grades are as follows.

Learning outcomes 1, 2 and 3:

P	Can formulate and convey most important points that could be expected on the topic.
Cr / D	Can formulate and convey clearly all important points that could be expected on the topic.
HD	As for Cr or D and can come up with novel insightful points on the topic.

Learning Outcomes 4 and 5.

P	Be able to write a paper or document, or give a presentation, that would be acceptable at a conference.
Cr / D	Be able to write a paper or document, or give a presentation, that would be well received at a conference.
HD	Be able to write a paper or document, or give a presentation, that would be well received at a major international conference.

These assessment standards will be used to calculate a numeric mark for each assessed task during marking.

The total raw mark for the unit will be calculated by summing up the marks for all assessment tasks according to the percentage weightings shown in the assessment summary.

The project report has a **hurdle requirement** in this unit: you will need to obtain at least 40% for the project report in order to pass the unit. If you obtain between 30% and 40% for the project report, then you will be given a second (and final) attempt to submit your report.

In order to **pass** the unit, you need a total raw mark of at least 50%.