

# COMPSCI 760: Machine Learning and Data Mining, Semester 1 2018

## 1 Course description

Machine learning techniques are widely used in many computing applications; for example, in web search engines, spam filtering, speech and image recognition, computer games, machine vision, credit card fraud detection, stock market analysis and product marketing applications. Machine learning implies that there is some improvement that results from the learning program having seen some data. The improvement can be in terms of some performance program (e.g., learning an expert system or improving the performance of a planning or scheduling program), in terms of finding an unknown relation in the data (e.g., data mining, pattern analysis), or in terms of customizing adaptive systems (e.g., adaptive user-interfaces or adaptive agents).

In Pat Riddle's part of the paper, we will provide an overview of the learning problem and the view of learning as search. We will study several techniques for learning such as Ensemble Learning, Bayesian Learning, Genetic Algorithms, Reinforcement Learning, and Neural Networks. In addition, we will provide an overview of the experimental methods necessary for understanding machine learning research.

In Joerg Wicker's part of the paper, we will cover further recent research topics in machine learning and data mining. Specifically, we will address Multi-Label and Multi-Target Learning, Recommender Systems and Matrix Factorization, Privacy, Online Learning, and Time Series.

In Ian Waton's part of the paper, we will cover case-based reasoning, recommender systems, explainable artificial intelligence (XAI), and recent case-studies of applied ML (time permitting).

## 2 Learning outcomes

The students will be able to:

- Discuss the idea that all machine learning algorithms have a basis and will be able to describe the basis of several algorithms

- Discuss the theory that for a particular dataset one algorithm will perform well and for another dataset a different algorithm will perform well. There is no one algorithm that performs well on all datasets.
- Describe a machine learning algorithm as a search algorithm through a space of hypotheses.
- Design a good set of experiments for determining the answer do some basic research question, such that they can show that the experiments actually support the question they are asking.

### 3 Teaching staff

Pat Riddle (Coordinator)

Room: 490, Computer Science Building (Building 303S)

Phone: 373-7599, Ext 87093 Email: [pat@cs.auckland.ac.nz](mailto:pat@cs.auckland.ac.nz)

Joerg Wicker

Room: 526, Computer Science Building (Building 303)

Phone: 373-7599, Ext 82184 Email: [j.wicker@auckland.ac.nz](mailto:j.wicker@auckland.ac.nz)

Ian Watson

Room 493, Computer Science Building (Building 303S)

Phone 373-7599, Ext 88976 Email: [ian@cs.auckland.ac.nz](mailto:ian@cs.auckland.ac.nz)

### 4 Lecture Times

Mon 9am, Room 315, 206 (Arts 1)

Tues 8am, Room G20, 302 (Science) (It is in Building 302 not Building 303!!)

Thurs 4pm, Room 220, 206 (Arts 1)

### 5 Assessments

Your final grade will consist of a number of internal marks worth 40% combined and an exam worth 60%. This is set up as a research based course. So the internal marks will be based on a research project, done in 3 member teams. There is a practical pass on this paper. So make sure you spend enough time on the internal assessments.

## 5.1 Internal Marks

You will have a research project, done in 3 member teams, with 5 due dates.

Topic area agreement 0% Due: Friday March 29th

Literature Survey 12% Due: Friday April 20th

Proposal area agreement 0% Due: Friday April 27th

Proposal 4% Due: Friday May 4th

Final Report 24% Due: Friday June 1st

The two 0% agreements require you to get approval (in writing in email) from Pat, Joerg, or Ian concerning your topic and proposal areas.

## 6 Proposed lecture schedules (subject to change)

Week 1 Ensembles, Neural Networks

Week 2 Bayesian Learning, Reinforcement Learning

Week 3 Experimental Design, Genetic Algorithms

Week 4 Recommender Systems, Matrix Factorization, Multi-Label, Multi-Target, Privacy

Week 5 Case-based reasoning, Recommender Systems

Week 6 Explainable Artificial Intelligence (XAI), recent case-studies of applied ML, Online Learning

Week 7 Time Series

Weeks 8-10 Project Meetings

Week 11 Peer Review

Week 12 Project Interviews

## 7 Seeking assistance

The primary source of assistance is the teaching staff. Please contact Pat, Joerg, or Ian with any questions or concerns about the course. Both are available via email.

For help with more generic study skills or literacy, the Student Learning Centre and Library both offer many courses designed to help students become more efficient at study.

## 7.1 Missed lectures

Overhead slides and recommended reading will be provided for the lectures on the CS Department website. Please review the material prior to seeing the teaching staff. If you know in advance that you will miss a lecture, please let the course coordinator know.

## 7.2 Exam

The final exam is worth 60% of your final mark. Please check Student Services Online for the exam time and date. The exam is closed book, calculators are not permitted. Provisional exam results can be obtained from Student Services Online.

## 7.3 Missed exam

If you miss the exam for any valid reason, or you sit the exam but believe that your performance was impaired for some reason, then you may be able to apply for an aegrotat, compassionate or special pass consideration. For more detailed information, refer to pages 44–45 of the University of Auckland's 2012 Calendar.

## 7.4 Policy on Cheating and Plagiarism

Cheating is viewed as a serious offence by the University of Auckland. Penalties are administered by the Discipline Committee of the Senate, and may include suspension or expulsion from the university. Do not copy anyone else's work, or allow anyone else to copy from you.

For more information on the University's policy on cheating, please refer to the web page: <http://www.auckland.ac.nz/uoa/home/about/teaching-learning/honesty>