Course Outline

COMP9417
Machine Learning and Data Mining
Semester 1 2017

Contents

- Course Staff
- Course Details
- Course Summary
- Course Aims
- Student Learning Outcomes
- Assumed Knowledge
- Teaching Rationale
- Teaching Strategies
- Assessment
- Academic Honesty and Plagiarism
- Course Schedule
- Resources for Students
- Course Evaluation and Development

Course Staff

Contact Details

Staff Name	Role	Email	Phone
Mike Bain	Course Convenor	mike at cse.unsw.edu.au	56935
Edwin Bonilla	Lecturer	e.bonilla at unsw.edu.au	93856255

Consultations

Staff Name	Day/Time	Location
Mike Bain	Wednesday 1400-1500	K17-402

Course Details

Course Code: COMP9417
Course Title: Machine Learning and Data Mining
Units of Credit: 6
Course WebSite: http://www.cse.unsw.edu.au/~cs9417/
Handbook Entry: http://www.handbook.unsw.edu.au/postgraduate/courses/2017/COMP9417.html

Course Summary

This course explores machine learning as the algorithmic approach to learning from data. The course also covers key aspects of data mining, which is understood as the application of machine learning tools to obtain insight from data. Algorithms are placed in the context of their theoretical foundations in order to understand their derivation and correct application. Topics include: linear models for regression and classification, local methods (nearest neighbour), neural networks, tree learning, kernel machines, unsupervised learning, ensemble learning, computational and statistical learning theory, and Bayesian learning. To expand and extend the development of theory and algorithms presented in lectures practical applications will be given in tutorials, programming tasks and assignments.

Course Aims

This course aims to provide an introduction to and overview of machine learning and data mining. Theoretical foundations, algorithmic techniques and methods for practical application, evaluation and development of machine learning tools will be covered. As a result of successfully completing this course students will have a working knowledge of key topics in machine learning, and will be able to demonstrate their knowledge both by describing aspects of the topics and by solving problems related to the topics. They will have practical experience with the topics covered in the assignments undertaken.

Student Learning Outcomes

After completing this course, students will be able to:

- set up a well-defined learning problem for a given task
- select and define a representation for data to be used as input to a machine learning algorithm
- select and define a representation for the model to be output by a machine learning algorithm
- compare different algorithms according to the properties of their inputs and outputs
- compare different algorithms in terms of similarities and differences in the computational methods used s
- develop and describe algorithms to solve a learning problem in terms of the inputs, outputs and computational methods used
- express key concepts from the foundations of computational and statistical learning theory and demonstrate their applicability
- express knowledge of general capabilities and limitations of machine learning from computational and statistical theory
- use or extend or invent algorithms in applications to real-world data sets and collect results to enable evaluation and comparison of their performance

This course contributes to the development of the following graduate capabilities:

Graduate Capability	Acquired in
scholarship: understanding of their discipline in its interdisciplinary context	lectures
scholarship: capable of independent and collaborative enquiry	tutorials and assignments
scholarship: rigorous in their analysis, critique, and reflection	tutorials and assignments
scholarship: able to apply their knowledge and skills to solving problems	tutorials, assignments and exams
scholarship: ethical practitioners	lectures and assignments
scholarship: capable of effective communication	tutorials and assignments
scholarship: information literate	suggested readings, tutorials and assignments
scholarship: digitally literate	tutorials and assignments
leadership: enterprising, innovative and creative	assignment 2 (project)
leadership: capable of initiating as well as embracing change	assignment 2 (project)
leadership: collaborative team workers	assignment 2 (project)
professionalism: capable of independent, self-directed practice	lectures, tutorials and assignments
professionalism: capable of lifelong learning	lectures, tutorials and assignments
professionalism: capable of operating within an agreed Code of Practice	lectures, tutorials and assignments
global citizens: capable of applying their discipline in local, national and international contexts	lectures, tutorials and assignments, and student interactions
global citizens: culturally aware and capable of respecting diversity and acting in socially just/responsible ways	lectures, tutorials and assignments, and student interactions
global citizens: capable of environmental responsibility	lectures, tutorials and assignments

Assumed Knowledge

Before commencing this course, students should have completed the pre-requisite courses (or equivalent) and ensure they have acquired knowledge in the relevant areas:

- Prerequisite is COMP1927 Computing 2 or equivalent (since 2013). Waivers can be granted where applicable (see Course Coordinator)
- Mathematical assumed knowledge is completion of basic university mathematics courses, such as the UNSW courses MATH1131 and MATH1231
- Additionally, in practice, some knowledge of basic probability, statistics and logic will be the starting point for some course materials.
- Ability to program and construct working software in a general-purpose programming language (e.g., C, Java, Perl, Python, etc.) is assumed. An important part of practical data mining is Â“data wranglingÂ” Â– pre-processing, filtering, cleaning, etc. of data files Â– for this you need to have mastered Unix tools such as those taught in COMP2041 Software Construction, or equivalents such as can be found in R, Matlab/Octave, etc.

These are assumed to have been acquired in the courses ...

Teaching Rationale

This course is taught to emphasise that theory, algorithms and empirical work are essential inter-dependent components of machine learning. Teaching is mainly focused on lectures and assignments, with tutorials to expand and reinforce the lecture content. Assessment is by mid-session and final exams and two marked assignments. The assignments are aimed at giving students an opportunity for active learning in a structured way with submission deadlines. The purpose is to give students practical experience of machine learning and relate lecture material to real applications. The second assignment has a broad scope and should be treated as a

small-scale project with submission of software and a written report.

Teaching Strategies

- Lectures ... introduce concepts, show examples
- Tutorials ... reinforce concepts and provide additional examples
- Assignments .. allow students to solve significant problems

Assessment

Details of assessment items:

Assessment component	Marks
Assignment 1	10
Assignment 2	30
Mid-semester Exam	10
Final Exam	50

Assignment submission will be electronic via the CSE give system. Details of submission, deadlines and late penalties, etc. will be in the assignment specifications.

The overall course mark will be the sum of the marks for the course components.

The outcome of attaining a course mark in the range 45-49 will be decided on a case-by-case basis.

Academic Honesty and Plagiarism

Plagiarism is defined as "using the words or ideas of others and presenting them as your own". UNSW and CSE treat plagiarism as academic misconduct, which means that it carries penalties as severe as being excluded from further study at UNSW. There are several on-line sources to help you understand what plagiarism is and how it is dealt with at UNSW:

- Learning Centre: Plagiarism and Academic Integrity
- MyUNSW: Plagiarism and Academic Misconduct
- CSE: Addendum to UNSW Plagiarism Guidelines
- CSE: Yellow Form (whose terms you have agreed to)

Make sure that you read and understand these. Ignorance is not accepted as an excuse for plagiarism.

Course Schedule

Schedule of topics, organised by week. Note: this schedule may be subject to change !

Week	Lecture	Tutorial
1	Introduction to Machine Learning	-
2	Supervised Learning - Regression	Introduction
3	Supervised Learning - Classification	Regression (notebook)
4	Supervised Learning - Neural Networks	Regression & Classification
5	Supervised Learning - Tree Learning	Classification (notebook)
6	Supervised Learning - Kernel Machines	Neural Networks
7	Unsupervised Learning	Neural Networks (notebook - tbc)
8	Computational Learning Theory	Clustering
9	Ensemble Learning	Clustering (notebook)
10	Bayesian Inference - Supervised Learning	Ensembles
11	Bayesian Inference - Unsupervised Learning	Ensembles (notebook)
12	Applications and Guest Lecture (tbc)	Ensembles
13	-	Probabilistic models (notebook)

Resources for Students

Owing to the expansion of machine learning in recent years, and the availability of online materials, there is no single textbook for this course. However, we give a list of books (some freely available online) that can be consulted to back up and expand on the course content. If you plan to continue with machine learning, any of these are worth owning:

- Murphy, Kevin P. Machine learning: a probabilistic perspective. MIT press, 2012.
- Barber, David. Bayesian reasoning and machine learning. Cambridge University Press, 2012.

- Trevor Hastie, Robert Tibshirani and Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009.
- Bishop, Christopher M. Pattern recognition and Machine Learning. Springer, 2006.