Learning objectives and outcomes

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.

By the end of the subject, students should 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
- 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

Staff

<table>
<thead>
<tr>
<th>Staff Name</th>
<th>Role</th>
<th>Email</th>
<th>Office</th>
<th>Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mike Bain</td>
<td>Lecturer &amp; Coordinator</td>
<td><a href="mailto:mike@cse.unsw.edu.au">mike@cse.unsw.edu.au</a></td>
<td>K17-401H</td>
<td>9385 6935</td>
</tr>
</tbody>
</table>

Contact course coordinator to arrange consultation.
Course Web Pages

http://www.cse.unsw.edu.au/~cs9417/

Course Information

- Units of credit – 6
- Parallel teaching – no, only COMP9417 students attend the class
  - The students include postgraduates and senior undergraduates

Assumed knowledge/prerequisites

Prerequisites are COMP9024 Data Structures and Algorithms or COMP1921 Data Structures and Algorithms (check with University handbook).

Waivers can be granted where applicable (arrange to see Course Coordinator).

Mathematical assumed knowledge is completion of basic university mathematics courses, such as the UNSW courses MATH1131 and MATH1231.

In practical terms, some knowledge of basic probability, statistics and logic will be the starting point for the course materials.

Ability to program and construct working software (preferably in Java) is assumed.

Course philosophy and teaching strategies

Learning is mainly focused on lectures and assignments. Assessment is by mid-session and final exams and two marked assignments, plus a short eLearning activity. 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.
### Syllabus

Topics will be chosen from:

**Module 1:** Fundamentals of machine learning and data mining.

**Module 2:** Computational and statistical foundations of machine learning.

**Module 3:** Advanced Machine Learning Concepts and Techniques.

### Assignments

Assignments will involve the process of applying and modifying or implementing machine learning software, using the tools and techniques described in lectures. The first assignment will involve the use of a standard open-source machine learning toolkit. The second assignment will be a more open-ended machine learning application project to solve a learning problem with a written report evaluating the methods used. The assignments contribute 40% of the overall mark for the course.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Description</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Machine learning toolkit</td>
<td>Week 5</td>
</tr>
<tr>
<td>2</td>
<td>Project</td>
<td>Week 12</td>
</tr>
</tbody>
</table>

Note that these dates are subject to change.
Exams

There will be two closed-book written exams: a mid-term exam and a final exam. The written exams contribute 55% of the overall mark for the course.

Exam timetable

<table>
<thead>
<tr>
<th>Exam</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid-term</td>
<td>Week 6 or 7</td>
</tr>
<tr>
<td>Final</td>
<td>Exam period</td>
</tr>
</tbody>
</table>

Note: dates may be subject to change.

eLearning

This year we will be running an eLearning tutorial-style activity which will become available towards the end of the course. This will contribute 5% of the overall mark for the course.

It will be an online activity which you complete in your own time.

More details closer to the release date.

Assessment

<table>
<thead>
<tr>
<th>Assessment component</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 assignment 1</td>
<td>10</td>
</tr>
<tr>
<td>2 mid-term exam</td>
<td>15</td>
</tr>
<tr>
<td>3 assignment 2 / project</td>
<td>30</td>
</tr>
<tr>
<td>4 eLearning</td>
<td>5</td>
</tr>
<tr>
<td>5 final exam</td>
<td>40</td>
</tr>
</tbody>
</table>

NOTE: course mark is total of component marks!

Relation of assessment to learning objectives

Course learning objectives relate to the assessment components as follows:

- set up a well-defined learning problem for a given task    \[2,3,5\]
- select and define a representation for data to be used as input to a machine learning algorithm \[2,3,5\]
- select and define a representation for the model to be output by a machine learning algorithm \[2,3,5\]
- compare different algorithms according to the properties of their inputs and outputs \[1,2,3,5\]
- compare different algorithms in terms of similarities and differences in the computational methods used \[1,2,3,4,5\]
- develop and describe algorithms to solve a learning problem in terms of the inputs, outputs and computational methods used \[2,3,4,5\]

\[\text{COMP9417: February 27, 2007 Course Intro: Slides 12-15}\]
• express key concepts from the foundations of computational and statistical learning theory and demonstrate their applicability \[3,4,5\]

• express knowledge of general capabilities and limitations of machine learning from computational and statistical theory \[3,4,5\]

• use or extend or invent algorithms in applications to real-world data sets and collect results to enable evaluation and comparison of their performance \[1,3\]

Plagiarism and Academic Honesty

All work submitted for assessment must be your own work. Assignments must be completed individually and will be checked. Copying is unacceptable. We regard copying of assignments, in whole or part, as a very serious offence. Penalties for copying range from receiving no marks, through receiving a mark of 00 FL for the course, to expulsion from UNSW (for repeat offenders). Allowing someone to copy your work counts as plagiarism, even if you can prove that it is your work.

• Submission of work derived from another person, without their consent, will result in automatic failure for the course with a mark of zero.

• Submission of work derived from another person with their knowledge, or jointly written with someone else, will result in zero marks for the submission.

Learning Strategies and Academic Honesty

We are aware that a lot of learning takes place in student conversations, and don’t wish to discourage those. However, it is important, for both those helping others and those being helped, not to provide/accept any programming language code in writing, as this is apt to be used exactly as is, and lead to plagiarism penalties for both the supplier and the copier of the codes. Write something on a piece of paper, by all means, but tear it up/take it away when the discussion is over.

If you are new to studying in Australia, be aware that attitudes to plagiarism at UNSW may be different from those in your home country. Make sure you are clear about the rules here at UNSW. In brief, and for the purposes of this course, plagiarism includes copying or obtaining all, or a substantial part, of the material for your assignment, whether programming language code, or written or graphical report material, without written acknowledgement in your assignment from:
1. a location on the internet;
2. a book, article or other written document (whether published or unpublished) whether electronic or on paper or other medium;
3. another student, whether in your class or another class;
4. a non-student (e.g. from someone who writes assignments for money)

Note that if you copy code or other material from another student or non-student with acknowledgement, you will not be penalised for plagiarism, but you are unlikely to get any marks for the copied material. If you use code found in a publication (on the internet or otherwise) then the marks you get for this will be at the marker’s discretion, and will reflect the marker’s perception of the amount of work you put into finding and/or adapting the code, and the degree to which you understand the code.

Note also that there is a big difference between being able to understand someone else’s code, and writing that code yourself from scratch. A computer programmer has to be able to write code from scratch. Assignment 2 provides opportunities for you to develop the skills necessary to write your own machine-learning-related code. Use these opportunities!

---

**Schedule**

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mar 11</td>
<td>Introduction; Concept learning</td>
</tr>
<tr>
<td>2</td>
<td>Mar 18</td>
<td>Decision tree learning</td>
</tr>
<tr>
<td>3</td>
<td>Mar 25</td>
<td>Rule learning</td>
</tr>
<tr>
<td>4</td>
<td>Apr  1</td>
<td>Learning for numerical prediction</td>
</tr>
<tr>
<td>5</td>
<td>Apr  8</td>
<td>Instance-based learning; Genetic algorithms</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Mid-session Break</em></td>
</tr>
<tr>
<td>6</td>
<td>Apr 22</td>
<td>Mid-session exam; Reinforcement learning</td>
</tr>
<tr>
<td>7</td>
<td>Apr 29</td>
<td>Logical and Relational Learning</td>
</tr>
<tr>
<td>8</td>
<td>May 6</td>
<td>Evaluation and hypothesis testing</td>
</tr>
<tr>
<td>9</td>
<td>May 13</td>
<td>Learning theory</td>
</tr>
<tr>
<td>10</td>
<td>May 20</td>
<td>Bayesian learning</td>
</tr>
<tr>
<td>11</td>
<td>May 27</td>
<td>Ensemble learning; Kernel methods</td>
</tr>
<tr>
<td>12</td>
<td>Jun  3</td>
<td>Unsupervised learning</td>
</tr>
</tbody>
</table>

*Note: this schedule may be subject to change.*
**Reference Books**

Textbook:
OR
Data Mining* (2nd ed.), Witten and Frank, (2005), Elsevier

Reference books:
Data Mining (2nd ed.), Han and Kamber, (2006), Morgan Kaufmann
Pattern Classification (2nd ed.), Duda, Hart and Stork, (2001), Wiley
Elements of Statistical Learning, Hastie, Tibshirani and Friedman, (2001), Springer
Pattern Recognition and Neural Networks, Ripley, (1996), Cambridge
Classification and Regression Trees, Breiman, Friedman, Olshen and Stone (1984), Kluwer
C4:5: programs for Machine Learning, Quinlan (1993), Morgan Kaufmann

**Journals and Conferences**

Here are the names of some of the more prominent Machine Learning and Data Mining journals and conferences:

- J: Machine Learning
- J: Journal of Machine Learning Research
- J: Data Mining and Knowledge Discovery
- J: ACM Transactions on Knowledge Discovery in Databases
- C: International Conference on Machine Learning (ICML)
- C: Knowledge Discovery in Databases (KDD)
- C: Neural Information Processing Systems (NIPS)
- C: many others!

**Software**

* WEKA machine learning toolkit in Java
http://www.cs.waikato.ac.nz/ml/weka/
RapidMiner machine learning toolkit in Java
http://rapid-i.com/

**Dictionary**

Bill Wilson's dictionary of machine learning (for COMP9414)
http://www.cse.unsw.edu.au/~billw/mldict.html
Reading email

Please check your CSE email account frequently in case of announcements relating to this course. We assume that you read e-mail sent to your CSE account by the next working day during teaching sessions.

How to forward your CSE E-mail to another e-mail address: http://www.cse.unsw.edu.au/~cs9444/csemailfwd.html

Continual course improvement

Student feedback on this course, and on the lecturing in this course, will be gathered via questionnaires held at or after the end of the course. Student feedback is taken seriously, and continual improvements are made to the course based in part on this feedback. The course questionnaire results go to the Head of the School of Computer Science and Engineering, who reads the results and follows up in cases where action is clearly needed.

Suggested improvements from last year that will be followed up this year are to continue the eLearning component and provide more revision materials.

Further information

- Students enrolled in COMP9417 are expected to attend all classes
- The use of School of Computer Science and Engineering computing laboratories is subject to rules described in the Yellow Form http://www.cse.unsw.edu.au/people/studentoffice/policies/yellowform.html which you acknowledge (electronic) receipt of when you receive your computing account. The Yellow Form also outlines what to do in case of illness or misadventure that affects your assessment, and supplementary examinations procedures within the School of Computer Science and Engineering
- Information on UNSW Occupational Health and Safety policies and expectations is at http://www.hr.unsw.edu.au/ohsuc/ohs/ohs_policies.html
- For information on Equity and Diversity issues see http://www.studentequity.unsw.edu.au