Unit Outline: KXA457 Machine Learning and Data Mining

Semester 1, 2005

Sandy Bay Campus, Hobart
Newnham Campus, Launceston

Prerequisites
None

Corequisites
None

Unit Weight
12.5% of one academic year

Unit Coordinator
Dr. Mike Cameron-Jones

Lecturing Staff
Sandy Bay Campus, Hobart: Dr. Mike Cameron-Jones
Newnham Campus, Launceston: Dr. Mike Cameron-Jones

Scheduled Teaching Sessions
Lectures: (See below)
Tutorials: (See below)
The Unit Timetable can be accessed from the Study Resources section of the School website. ([http://www.comp.utas.edu.au/app/studyresources.jsp](http://www.comp.utas.edu.au/app/studyresources.jsp)).

There will be 2 or 3 hours of lecture/tutorial per week depending upon the week. All sessions are delivered via video conference. For information about videoconferencing at UTAS and how to participate effectively, see the Students' guide to Videoconferencing available at: [http://www.utas.edu.au/itr/videoconf/StudentGuide2004.pdf](http://www.utas.edu.au/itr/videoconf/StudentGuide2004.pdf) or follow the Service desk link from the Current Students homepage>Videoconferencing.

Unit Website
The unit website is accessed from [http://vista.utas.edu.au/](http://vista.utas.edu.au/). You will need to use your university email pop account username and password to log on to the WebCT system. Once authenticated by the system your personalised MyWebCT area will be displayed. It contains links to the websites that you have permission to access - including the website for this unit.

This unit is Web Dependent: communication. This means that you will need to use the Web for this unit. The unit website contains unit information and resources.

If you are not able to access the unit website, please contact the University IT help desk:
Entrance Level, Morris Miller Library, Sandy Bay Campus;
Entrance Level, Launceston Campus Library, Newnham Campus.
Telephone: 6226 1818 and 1300 304 903.
The 1300 number is a local call from within Tas, with the exception of mobiles.
Email: servicedesk@utas.edu.au

University Website
Information and Resources for 'Current Students' are available on the university website at: [http://www.utas.edu.au/students/](http://www.utas.edu.au/students/)

Provider
School of Computing - Faculty of Science, Engineering, and Technology. [http://www.comp.utas.edu.au](http://www.comp.utas.edu.au)

OVERVIEW

Introduction
Machine Learning is commonly considered to be a sub-field of Artificial Intelligence, and can be seen as the study of computational approaches to finding patterns in data. Data Mining applies Machine Learning techniques to look for patterns in large data sets. This unit introduces the key current ideas and techniques in Machine Learning.

Prior Learning
Students entering the unit will be assumed to have a knowledge of algorithms, and of programming in Java, as covered in the BComp core.

Learning Outcomes
On successful completion of this unit, you will be able to:
Outcomes

- Understand the key current ideas and techniques in machine learning sufficiently to apply them to practical (data mining) problems and participate in research in the area.

Unit Content

Classifier Learning:
1. Decision Trees
2. Rule Sets
3. Naive Bayes
4. Instance Based Learning
5. Neural Networks
6. Combining Multiple Classifiers

Other Topics:
1. Numeric Prediction
2. Inductive Logic Programming
3. WEKA
4. Applications

For more information see the section titled 'Content' on the unit website.

Generic Skills

The university has defined a set of generic graduate attributes expected in its graduates.
Your course is designed to enable you to develop generic skills that are valued in, and expected of, graduates. These are skills that you will need to develop over time. Hence you are encouraged to look for opportunities, as you study each unit, to reflect on and improve these skills.

LEARNING AND TEACHING

Approach to Learning

You are expected to spend about 130 hrs studying in this unit - this includes attendance at scheduled teaching sessions. (For a 13 week semester this is, on average, 10 hr/wk.) This is the amount of study time that the 'typical' student will need to reach the level of competence and understanding required to fulfil the unit objectives.

You are expected to:

- attend all scheduled lectures and tutorials, unless otherwise notified by the unit coordinator
- prepare for, and actively participate in lectures and tutorials
- complete the assigned learning tasks
- review what has been learnt
- complete assessment items and submit them on time
- access and be familiar with the information and resources available on the unit website
- seek help from teaching staff if you have any questions or difficulties in studying this unit

You are encouraged to read the university's Code of Conduct for Teaching and Learning. Part A describes the 'Responsibility of the University to Students' and part B describes the 'Responsibilities of Students to the University'.

Schedule

See the 'Schedule' section on the unit website.

Teaching and Support Staff

Teaching Staff

Unit Coordinator:
Dr. Mike Cameron-Jones
E-Mail: Michael.CameronJones@utas.edu.au
Phone: (03) 6324 3395
Room: V171, Newnham Campus, Launceston

Lecturing Staff

Sandy Bay Campus, Hobart: Dr. Mike Cameron-Jones
Newnham Campus, Launceston: Dr. Mike Cameron-Jones

School Help Desk
Contact the School Help Desk if you have any queries or problems with accessing, using, or printing from the computers in the School of Computing labs.

- **Hobart**: the Help Desk is located near the School's reception desk and is open from 10am - 4pm Monday-Friday. The phone number is 6226 2960.
- **Launceston**: the Help Desk is located near the entrance to the computing labs and is open in the morning from 10am - 12pm, and in the afternoon from 2pm - 4:30pm, Monday-Thursday. On Fridays it is open from 10am - 12pm in the morning and 2pm - 4pm in the afternoon. The phone number is 6324 3447.
- **Burnie**: the computer labs at the NWC are maintained by ITS. Please contact the University Help Desk for assistance. The 6 Macs are maintained by the School of Computing. If you have a query or problem that is specific to the School of Computing please phone the School of Computing Help Desk in Launceston.

**University Services and Support**

The University has staff available to assist you, such as the:

- Learning Development Advisor
- Student Counselor
- Careers Advisor
- Disability Officer

For more information and contact details see the Services and Support section on the University 'Current Students' web page. [http://www.utas.edu.au/students/](http://www.utas.edu.au/students/)

**Resources**

**Unit Website**

The unit website contains unit information and resources. In order to provide comparable consultation opportunity to students on both campuses, a discussion forum on the unit web site will be the intended means by which students consult about matters that relate to the unit content.

**Prescribed Text**


Papers from the research literature will be prescribed during the unit.

**Software**

The software that you will need to access the unit website and to study this unit, including general purpose software such as word processors, is provided on the computers in the School's computing labs. If you intend to use software on other computers please check that the versions are compatible.

The unit will use the WEKA suite of machine learning programs (in Java).

**Computing Facilities**

The School has PC labs (Windows XP), Mac labs (Mac OS-X 10.3), and Networking labs at the Newnham and Sandy Bay campuses. It also maintains 6 Macs (Mac OS-X 10.3) at the NW Centre. Unix accounts can be accessed from all Macs and PCs.

If you have not used these facilities before please contact the School Help Desk to organise your account details. If you would like to access the facilities at the Newnham and Sandy Bay campuses after hours please contact the School Help Desk.

Please contact the School Help Desk if you have difficulty accessing or using these facilities.

**Use of Facilities**

Use of computing facilities provided by the School is subject to the School's Ethics Guidelines - [http://www.comp.utas.edu.au/app/ethics.jsp](http://www.comp.utas.edu.au/app/ethics.jsp). Copies of the guidelines are also available in all School labs. The School's facilities may only be used for study-related purposes, and may not be used for personal gain. The playing of games is strictly prohibited in all labs at all times. Before being granted access to the School's facilities, you will be required to sign a declaration that you have read and understand these guidelines, and that you will abide by them. Disciplinary action may be taken against students who violate the guidelines.
The university is committed to providing a safe and secure teaching and learning environment. For more information see [http://www.admin.utas.edu.au/hr/ohs/pol_proc/](http://www.admin.utas.edu.au/hr/ohs/pol_proc/)

**ASSESSMENT**

**Assessment Items**

**Item 1**
- **Title:** Paper Presentation and Discussion
- **Type:** In-Semester - individual assignment
- **Weighting:** 20%
- **Due:** During semester

Students will present and discuss prescribed papers from the research literature. The format and dates of the presentations and discussions, and the allocation of papers to students, will be specified during the unit. This activity may be organised around groups depending upon the number of students in the unit. (Note that missing a presentation or discussion will result in loss of marks.)

It is expected that these marks will be influenced by peer assessment.

**Item 2**
- **Title:** Assignment
- **Type:** In-Semester - individual assignment
- **Weighting:** 20%
- **Due:** 3pm 19th May (Thursday of Week 12)

The assignment will involve implementing or modifying some machine learning programs, experimenting with them and others, and writing a small paper reporting the results obtained.

**Item 3**
- **Title:** 2 hr Examination
- **Type:** Formal Examination
- **Weighting:** 60%
- **Due:** University Examination Period

It is expected that this year's exam will be similar to last year's exam in respect of the style of questions, and assessment balance amongst major topics. It is also expected that this year's exam will be similar to last year's exam in respect of the materials permitted. Further information regarding this year's exam will be given towards the end of the unit.

See the 'Assessment' section in unit website for more detailed information about assessment items.

**In-Semester Assessment**

Unless specifically stated in the specification of the assessment item provided on the unit website, it is required that:

- work submitted by a student is the work of that student alone OR
- where the assessment item is to be completed by a group of students, the work submitted by the group of students is the work of that group of students alone.

**Plagiarism**

Plagiarism is a form of cheating. It is taking and using someone else's thoughts, writings or inventions and representing them as your own, for example:

- using an author's words without putting them in quotation marks and citing the source;
- using an author's ideas without proper acknowledgment and citation; or
- copying another student's work.

If you have any doubts about how to refer to the work of others in your assignments, please consult your lecturer or tutor for relevant referencing guidelines, and the academic integrity resources on the web at [http://www.utas.edu.au/tl/supporting/academicintegrity/index.html](http://www.utas.edu.au/tl/supporting/academicintegrity/index.html).
The intentional copying of someone else's work as one's own is a serious offence punishable by penalties that may range from a fine or deduction/cancellation of marks and, in the most serious of cases, to exclusion from a unit, a course or the University. Details of penalties that can be imposed are available in the Ordinance of Student Discipline – Part 3 Academic Misconduct, see [http://www.utas.edu.au/policy/subject.html#students](http://www.utas.edu.au/policy/subject.html#students).

The University reserves the right to submit assignments to plagiarism detection software, and might then retain a copy of the assignment on its database for the purpose of future plagiarism checking.

**Referencing**

The university document on plagiarism contains information about referencing the work or ideas of others. (See [http://www.utas.edu.au/plagiarism/](http://www.utas.edu.au/plagiarism/)). The preferred text referencing systems for the School is the Harvard system (also referred to as the author-date system).

**Submissions**

The details of the submission method (paper, electronic or other) for each assignment will be supplied in a separate assignment specification sheet. All in-semester assignment submissions (including electronic submissions) are to include an Assignment Cover Sheet which includes a statement confirming that the submission is your own work. If this undertaking is not signed, the assignment will not be marked. The Assignment Cover Sheet is available from the School Help Desk in Launceston and Hobart, and on the School's website [http://www.comp.utas.edu.au/app/studyresources.jsp](http://www.comp.utas.edu.au/app/studyresources.jsp).

**Extensions**

Assessment items will not be accepted after the due date except under the conditions stated in the school policy on late assessment. [http://www.comp.utas.edu.au/app/late_assess.jsp](http://www.comp.utas.edu.au/app/late_assess.jsp)

**Formal Examination**

The formal examination is conducted by the University Registrar. The 'Current Students' section on the university website contains information about the conduct of, and timetable for, formal examinations.

**Final Grade**

Overall assessment will be based on the student's performance throughout the semester as well as in a formal examination. In order to achieve a pass (or better) result, a student must obtain:

1. at least 40% of the total mark for in-semester assessment items
2. at least 40% of the mark for the formal examination
3. at least 50% of the overall mark

Passing grades will be awarded based on the AVCC guidelines:

- PP at least 50% of the overall mark but less than 60%
- CR at least 60% of the overall mark but less than 70%
- DN at least 70% of the overall mark but less than 80%
- HD at least 80% of the overall mark

The maximum mark awarded to a student who fails the unit will be 44.

For more information, including other grades such as Supplementary and Terminating grades, see the School of Computing's guidelines for assessment - available at: [http://www.comp.utas.edu.au/app/assess.jsp](http://www.comp.utas.edu.au/app/assess.jsp)