



The University of Southern Queensland

Course specification

This version produced 11 Jan 2008.

The current and official versions of the course specifications are available on the web at
<<http://www.usq.edu.au/coursespecification/current>>.

Please consult the web for updates that may occur during the year.

Description: Data Mining Discovers Knowledge

| Subject | Cat-nbr | Class | Term | Mode | Units | Campus |
|---------|---------|-------|---------|------|-------|-----------|
| CSC | 3417 | 62852 | 1, 2007 | ONC | 1.00 | Toowoomba |

| | |
|-----------------------------------|--------|
| Academic group: | FOSCI |
| Academic org: | FOS003 |
| Student contribution band: | 2 |
| ASCED code: | 029999 |

STAFFING

Moderator: Ron Addie

REQUISITES

Pre-requisite: (CSC1401 and MAT1101) or USQIT16

RATIONALE

Data mining is an interdisciplinary field which brings together techniques of machine learning, database information retrieval, mathematics and statistics. These techniques are used to find useful patterns in large datasets. Methods for such knowledge discovery in data bases are required owing to the size and complexity of data collection in administration, business and science.

SYNOPSIS

Data mining aims at finding useful regularities or patterns in large data sets generated in modern management and science. This course covers the main data mining methods, including clustering, classification, association rules mining, and recent techniques for data mining. The methods are developed and applied to various data sets.

OBJECTIVES

On completion of this course students will be able to:

1. program key algorithms in data mining applications (Assignment 1);
2. appropriately classify and develop knowledge discovery problems (All assessment items);
3. discuss the effectiveness and estimate the performance of data mining algorithms (All assessment items);
4. design data mining solutions to simple, real world problems (Assignment 1, Exam).

TOPICS

| | Description | Weighting (%) |
|----|---|---------------|
| 1. | Find association rules | 20.00 |
| | 1.1. Market basket analysis, rules, frequent itemsets, A Priori algorithm, scalability | |
| 2. | Determine clusters | 20.00 |
| | 2.1. data vectors, scaling, metric spaces, K-means algorithm, finds clusters, agglomerate into hierarchial clusters | |
| 3. | Grow and prune decision trees to classify | 20.00 |
| | 3.1. recursive decisions, Gini index and entropy, prune to avoid overfitting, scalability, cross validation | |
| 4. | Other commonly used data mining methods | 30.00 |
| | 4.1. Artificial neural networks | |
| | 4.2. Nearest neighbour classifier | |
| | 4.3. Bayesian inference | |
| | 4.4. Web search engine | |
| 5. | Linear regression and Hypersurface models | 10.00 |
| | 5.1. vector space, linear regression, hypersplines, decision trees decide knots | |

TEXT and MATERIALS required to be PURCHASED or ACCESSED

ALL textbooks and materials are available for purchase from USQ BOOKSHOP (unless otherwise stated). Orders may be placed via secure internet, free fax 1800642453, phone 07 46312742 (within Australia), or mail. Overseas students should fax +61 7 46311743, or phone +61 7 46312742. For costs, further details, and internet ordering, use the 'Textbook Search' facility at <http://bookshop.usq.edu.au> click 'Semester', then enter your 'Course Code' (no spaces).

2007, 'CSC3417 Data mining discovers knowledge: Introductory book' (Available: <http://www.sci.usq.edu.au/courses/csc3417>).

2007, 'CSC3417 Data mining discovers knowledge: Study book' (Available: <http://www.sci.usq.edu.au/courses/csc3417>).

Students will require access to computer or internet facilities for computer programming.

REFERENCE MATERIALS

Reference materials are materials that, if accessed by students, may improve their knowledge and understanding of the material in the course and enrich their learning experience.

Berry, M & Linoff, G 1997, *Data mining techniques: for marketing, sales and customer support*, Wiley, New York.

Han, J & Kamber, M 2001, *Data mining: Concepts and techniques*, Morgan Kaufmann, San Francisco.

(ISBN: 1-55860-489-8)

Hand, D, Mannila, H & Smyth, P 2001, *Principles of data mining*, MIT Press, Cambridge, Mass.

Hegland, M 2001, Data mining techniques, *Acta Numerica*, Vol 10, no. , pp313-355.

Kantardzic, M 2003, *Data mining - concepts, models, methods and algorithm*, IEEE Press & Wiley-InterScience,

Tan, P, Steinbach, M & Kumar, V 2005, *Introduction to data mining*, Addison Wesley,

(ISBN: 0321321367)

Witten, I & Frank, E 2005, *Data Mining: Practical machine learning tools and Techniques*, 2nd edn, Morgan, Kaufmann.

(ISBN: 0120884070)

STUDENT WORKLOAD REQUIREMENTS

| ACTIVITY | HOURS |
|---------------|-------|
| Assessment | 30.00 |
| Examinations | 2.00 |
| Laboratory | 12.00 |
| Lectures | 39.00 |
| Private Study | 85.00 |

ASSESSMENT DETAILS

| Description | Marks out of | Wtg(%) | Due date |
|----------------------------|--------------|--------|-------------|
| ASSIGNMENT 1 | 10.00 | 10.00 | 23 Mar 2007 |
| ASSIGNMENT 2 | 10.00 | 10.00 | 05 Apr 2007 |
| ASSIGNMENT 3 | 10.00 | 10.00 | 11 May 2007 |
| 2HR RESTRICTED EXAMINATION | 100.00 | 70.00 | END S1 |

IMPORTANT ASSESSMENT INFORMATION

1 Attendance requirements:

It is the students' responsibility to attend and participate appropriately in all activities (such as lectures, tutorials, laboratories and practical work) scheduled for them, and to study all material provided to them or required to be accessed by them to maximise their

- chance of meeting the objectives of the course and to be informed of course-related activities and administration.
- 2 Requirements for students to complete each assessment item satisfactorily:
To complete each of the assessment items satisfactorily, students must obtain at least 50% of the marks available for each assessment item.
 - 3 Penalties for late submission of required work:
If students submit assignments after the due date without prior approval then a penalty of 20% of the total marks gained by the student for the assignment will apply for each working day late.
 - 4 Requirements for student to be awarded a passing grade in the course:
To be assured of receiving a passing grade a student must submit all of the summative assessment items, achieve at least 50% in the examination and at least 50% of the available weighted marks for the summative assessment items.
 - 5 Method used to combine assessment results to attain final grade:
The final grades for students will be assigned on the basis of the aggregate of the weighted marks obtained for each of the summative assessment items in the course.
 - 6 Examination information:
Candidates are allowed access only to specific materials during a Restricted Examination. The only materials that candidates may use in the restricted examination for this course are: writing materials (non-electronic and free from material which could give the student an unfair advantage in the examination); calculators which cannot hold textual information (students must indicate on their examination paper the make and model of any calculator(s) they use during the examination).
 - 7 Examination period when Deferred/Supplementary examinations will be held:
Any Deferred or Supplementary examinations for this course will be held during the examination period at the end of the semester of the next offering of this course.
 - 8 University Regulations:
Students should read USQ Regulations 5.1 Definitions, 5.6. Assessment, and 5.10 Academic Misconduct for further information and to avoid actions which might contravene University Regulations. These regulations can be found at the URL <http://www.usq.edu.au/corporateservices/calendar/part5.htm> or in the current USQ Handbook.

ASSESSMENT NOTES

- 9 The due date for an assignment is the date by which a student must despatch the assignment to the USQ. The onus is on the student to provide proof of the despatch date, if requested by the Examiner.
- 10 Student may be required to provide a copy of assignments submitted for assessment purposes. Such copies should be despatched to the USQ within 24 hours of receipt of a request to do so.

OTHER REQUIREMENTS

- 1 Students will require access to e-mail and internet access for this course.
-