Department of Statistics

STATS 784: An Introduction to Data Mining

Course Information 2017

Lecturers:
First half of semester
Alan Lee
Department of Statistics
Room 367, Mathematics and Physics Building 303S
Telephone : 923-8749
email: lee@stat.auckland.ac.nz

Second half of semester
Thomas Yee
Department of Statistics
Room 267, Mathematics and Physics Building 303
Telephone : 923-8811
email: t.yee@auckland.ac.nz

Course web page: www.stat.auckland.ac.nz/~lee/784

Course Aims:
- To introduce you to the data mining process
- To teach you some useful data mining tools
- To give you some skills in data wrangling and handling large data sets

Course organization:
There will be two lectures per week (at 4pm Monday and 11am Friday), plus an in-class one-hour lab (Mon 5-6pm) where we will work through some practical applications of the techniques we have been studying. Monday class is in Rm 201E-508 in the Human Sciences building, and the Friday class is in Rm 303-G16 in the Maths/Stats building

Assessment:

Assignments (5)	20%
Mid-semester Test	20%
Final Exam	60%

Assignments:
There will be 5 assignments, at about 2-week intervals.

Collaboration:
It is our view that collaboration is an important part of the learning process and I encourage you to discuss the material you are studying with each other (and us!) However, you must not copy another person's assignment answers.

Computing:
The course will be R based, and a reasonable familiarity with R coding will be assumed. There will be an opportunity to use SAS for some of the homework in the second half of the course.

Course Content:
1. Introduction to data mining, the data mining process
2. Supervised learning (prediction, classification)
3. Unsupervised learning (clustering)
4. Visualization
5. Handling large data sets

Course Book:
Copies of the lecture slides will be handed out in class for the first half of the course. A course book covering the second half of the course will be also be available.

Web page:
The course web page is at www.stat.auckland.ac.nz/~lee/784 and we will post various resources (including copies of the lectures) on this page, plus administrative notices. Please check it regularly. Marks will be posted on CANVAS, but no resources will be posted there.

Office hours, First half:
Officially, Monday afternoon 2-4. Alan Lee works part time and will generally not be at the University on Tuesdays, Wednesdays and Thursdays. Other than that I am happy to have students drop in at any time on Monday or Friday.