

COMP473 (2018) - Special Topic: Introduction to Big Data Analysis

Prescription

Big Data refers to the large and often complex datasets generated in the modern world: data sources such as commercial customer records, internet transactions, environmental monitoring. This course provides an introduction to the theory and practice of working with Big Data. Students enrolling in this course should be familiar with the basics of statistical modelling and with programming.

Course learning objectives

Students who pass this course should be able to:

1. Understand the basic concepts of how and where Big Data arises.
2. Describe data structures and mechanisms for the capture, storage, processing, summary and visualisation of Big data.
3. Implement practical methods for data acquisition and management using appropriate software (R, Java, Python, Perl; Hadoop, Mapreduce; or Weka).
4. Understand basic methods of analysis of Big Data, including methods from machine learning for high dimensional data.

Course content

Section 1 Introduction to Big Data

- Introduction to Big Data, and the evolution of the science
- Global and local scale datasets, data generation, post processing and practical matters regarding interpretation.

Section 2 Regression, Clustering and other Techniques in Big Data

- Regression: ridge regression, local regression, lasso; curse of dimensionality
- Generalized additive models; case study on intelligible models in healthcare applications.
- Clustering and resampling methods.

Section 3 Machine learning for high-dimensional data

- Machine learning for high-dimensional data, dimensionality reduction and feature selection (and possibly missing data analysis).
- Wrapper, filter and embeded dimensionality reduction methods
- The techniques covered will include sequential forward selection, sequential backward selection, and other machine learning methods such as decision trees, random forest, support vector machines, genetic programming (and possibly particle swarm optimisation).

Section 4 Big Data Problems

- Big Data Examples, e.g. gene expression data, image analysis or text mining problems

Withdrawal from Course

Withdrawal dates and process:

<https://www.victoria.ac.nz/students/study/course-additions-withdrawals>

Lecturers

Mengjie Zhang (Coordinator)

Mengjie.Zhang@vuw.ac.nz 04 4635654

355 Cotton, Kelburn

Bing Xue

bing.xue@vuw.ac.nz 04 4635542

352 Cotton, Kelburn

Teaching Format

During the trimester there will be 2 hours lecture in a week. Tutorials and/or labs, and help desk will be announced.

Student feedback

Student feedback on University courses may be found at:

www.cad.vuw.ac.nz/feedback/feedback_display.php

Dates (trimester, teaching & break dates)

- Teaching: 05 March 2018 - 08 June 2018
- Break: 23 April 2018 - 27 April 2018
- Study period: 11 June 2018 - 14 June 2018
- Exam period: 15 June 2018 - 04 July 2018

Class Times and Room Numbers

05 March 2018 - 25 March 2018

- **Thursday** 10:00 - 12:00 – 202, New Kirk, Kelburn

05 March 2018 - 01 April 2018

- **Monday** 10:00 - 12:00 – 202, New Kirk, Kelburn

02 April 2018 - 15 April 2018

- **Thursday** 10:00 - 12:00 – 202, New Kirk, Kelburn

09 April 2018 - 15 April 2018

- **Monday** 10:00 - 12:00 – 202, New Kirk, Kelburn

30 April 2018 - 10 June 2018

- **Monday** 10:00 - 12:00 – 202, New Kirk, Kelburn
- **Thursday** 10:00 - 12:00 – 202, New Kirk, Kelburn

Set Texts and Recommended Readings

Required

There are no required texts for this offering.

Recommended

A Reading List is available via the course website.

Mandatory Course Requirements

In addition to achieving an overall pass mark of at least 50%, students must:

- submit reasonable attempts for at least two out of the three assignments

If you believe that exceptional circumstances may prevent you from meeting the mandatory course requirements, contact the Course Coordinator for advice as soon as possible.

Assessment

Assignment 1 (3 weeks)	Week 4	CLO: 1,2	20%
Assignment 2 (3 weeks)	Week 7	CLO: 2,3	20%
Assignment 3 (3 weeks)	Week 11	CLO: 3,4	20%
Final Exam	University exam period	CLO: 1,3,4	40%

Penalties

The penalty for assignments that are handed in late without prior arrangement is one grade reduction per day. Assignments that are more than one week late will not be marked.

Extensions

All assignments must be handed in on time unless you have made a prior arrangement with the lecturer or exceptional personal circumstances (for minor illnesses it is sufficient to discuss this with the lecturer).

Any assignment extension request with a special case should be made to the Course (co-)Coordinators Prof Mengjie Zhang and Dr Bing Xue before the submission deadline except for exceptional circumstances.

Submission & Return

There will be three assignments, each of which is worth 20% of the final grade. All the assignments will be submitted via <https://apps.ecs.vuw.ac.nz/submit/COMP473>

Marking Criteria

The criteria will be released during the course later.

Group Work

No group work is allowed.

Workload

In order to maintain satisfactory progress in COMP 473, you should plan to spend an average of at least 10 hours per week on this paper. A plausible and approximate breakdown for these hours would include:

- Lectures and tutorials: 3
- Readings: 2
- Assignments: 3

Teaching Plan

More information can be seen from the course website:
https://ecs.victoria.ac.nz/Courses/COMP473_2018T1/LectureSchedule

Communication of Additional Information

All online material for this course can be accessed at
https://ecs.victoria.ac.nz/Courses/COMP473_2018T1/

Links to General Course Information

- Academic Integrity and Plagiarism: <https://www.victoria.ac.nz/students/study/exams/integrity-plagiarism>
- Academic Progress: <https://www.victoria.ac.nz/students/study/progress/academic-progress> (including restrictions and non-engagement)
- Dates and deadlines: <https://www.victoria.ac.nz/students/study/dates>
- Grades: <https://www.victoria.ac.nz/students/study/progress/grades>
- Special passes: Refer to the Assessment Handbook, at <https://www.victoria.ac.nz/documents/policy/staff-policy/assessment-handbook.pdf>
- Statutes and policies, e.g. Student Conduct Statute: <https://www.victoria.ac.nz/about/governance/strategy>
- Student support: <https://www.victoria.ac.nz/students/support>
- Students with disabilities: https://www.victoria.ac.nz/st_services/disability/
- Student Charter: <https://www.victoria.ac.nz/learning-teaching/learning-partnerships/student-charter>
- Terms and Conditions: <https://www.victoria.ac.nz/study/apply-enrol/terms-conditions/student-contract>
- Turnitin: <http://www.cad.vuw.ac.nz/wiki/index.php/Turnitin>
- University structure: <https://www.victoria.ac.nz/about/governance/structure>
- VUWSA: <http://www.vuwsa.org.nz>

Offering CRN: [28126](#)

Points: 15

Prerequisites: STAT193 or ENGR123 or equivalent preparation in statistics; 30 approved 200/300 level points

Duration: 05 March 2018 - 04 July 2018

Starts: Trimester 1

Campus: Kelburn