School of
Engineering and Computer Science
Te Kura Mātai Pūkaha, Pūrorohiko

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 machine learning, data mining, statistical modelling and with programming.

Course learning objectives

Students who pass this course should be able to:

1. Identify properties and challenges of very large data sets in order to determine appropriate analysis techniques to apply a specific Big Data task.
2. Explain the challenges in high-dimensional data and choose appropriate dimensionality reduction methods, from a software library such as KNIME, to solve high-dimensional problems.
3. Analyse regression and clustering data to choose appropriate analysis methods with good parameter settings from a software library such as R to generate data visualisations and to address regression and clustering problems.
4. Use their understanding of tools such as Hadoop MapReduce and Apache Spark to implement relevant algorithmic analysis of Big Data problems using appropriate machine learning libraries.

Course content

2022: The course is primarily offered in-person, and there are components such as tests, labs, tutorials, and marking sessions which require in-person attendance. There will be remote alternatives for all the components of the course, but these are only available to students studying from outside the Wellington region. The remote option for tests will use a Zoom-based system for online supervision of the tests.

Students taking this course remotely must have access to a computer with camera and microphone and a reliable high speed internet connection that will support real-time video plus audio connections and screen sharing. Students must be able to use Zoom; other communication applications may also be used. A mobile phone connection only is not considered sufficient. The comuputer must be adequate to support the programming required by the course: almost any modern windows, macintosh, or unix laptop or desktop computer will be sufficient, but an Android or IOS tablet will not.
==========

Section 1 Introduction to Big Data

- What is Big Data?
- Where does Big Data come from?
- What we can do and what we should do with Big Data?
- Typical examples of Big Data analysis in real word

Section 2 Machine learning for high-dimensional data

- Data Preprocessing and Introduction to Feature Manipulation
- Machine learning for high-dimensional data, dimensionality reduction and feature selection (and possibly missing data analysis) Wrapper, filter and embedded dimensionality reduction method

- 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 3 Regression, Clustering and other Techniques in Big Data

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

Section 4 Big Data Tools/Project

- Hadoop MapReduce
- Apache Spark
- Spark Machine Learning Libraries

Withdrawal from Course

Withdrawal dates and process:
https://www.wgtn.ac.nz/students/study/course-additions-withdrawals

Lecturers

Dr Qi Chen (Coordinator)

qi.chen@vuw.ac.nz 04 886 5631

CO 329 Cotton Building (All Blocks), Gate 7, Kelburn Parade, Kelburn

Dr Andrew Lensen

andrew.lensen@vuw.ac.nz 04 886 5336

CO 354 Cotton Building (All Blocks), Gate 7, Kelburn Parade, Kelburn

Teaching Format

This course will be offered in-person and online. For students in Wellington, there will be a combination of in-person components and web/internet based resources. It will also be possible to take the course entirely online for those who cannot attend on campus, with all the components provided in-person also made available online.

Two lectures per week, with associated assignments. Additional content may be provided through video resources.

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: 28 February 2022 - 03 June 2022
- Break: 11 April 2022 - 24 April 2022
- Study period: 06 June 2022 - 09 June 2022
- Exam period: 10 June 2022 - 25 June 2022

Class Times and Room Numbers

28 February 2022 - 10 April 2022

- **Monday** 16:10 - 17:00 – LT220, Murphy, Kelburn
- **Tuesday** 15:10 - 16:00 – LT101, Murphy, Kelburn

25 April 2022 - 05 June 2022

- **Monday** 16:10 - 17:00 – LT220, Murphy, Kelburn
- **Tuesday** 15:10 - 16:00 – LT101, Murphy, Kelburn

Set Texts and Recommended Readings

Required

There are no required texts for this offering.

Mandatory Course Requirements

There are no mandatory course requirements for this course.

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

Assessment Item	Due Date or Test Date	CLO(s)	Percentage
Assignment 1 (25 hours)	Monday Week 5	CLO: 1,2,3	20%
Assignment 2 (25 hours)	Monday Week 9	CLO: 1,2,3	25%
Test (50 Minutes)	Monday Week 11	CLO: 1,2,3	25%
Assignment 3 (25 hours)	Tuesday Second Week of Assessment Period	CLO: 4	30%

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

Individual extensions will only be granted in exceptional personal circumstances, and should be negotiated with the course coordinator before the deadline whenever possible. Documentation (eg, medical certificate) may be required.

Submission & Return

All work should be submitted through the ECS submission system, accessible through the course web pages. Marks and comments will be returned through the ECS marking system.

The School normally has a goal of returning marks for all assessment items within two weeks of the submission deadline. This year, the course will aim to meet this goal, but we expect that sickness and self-isolation due to Covid will extend the time required to mark some assignments and tests.

Workload

In order to maintain satisfactory progress in AIML 427, 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: 2
- Readings: 2-4
- Assignments: 3-5

However, since this is multidisciplinary course, students with different background may need different amounts of time to work on different sections/assignments of the course, i.e. could be more or could be less.

Teaching Plan

See: https://ecs.wgtn.ac.nz/Courses/AIML427_2022T1/LectureSchedule

Communication of Additional Information

All online material for this course can be accessed at https://ecs.wgtn.ac.nz/Courses/AIML427_2022T1/

Links to General Course Information

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

Offering CRN: 33069

Points: 15
Prerequisites: one of (AIML 420, 421, COMP 307, 309, STAT 393, 394); one of (ENGR 123, STAT

193, MATH 177, QUAN 102) or comparable background in Statistics;
Restrictions: COMP 424, COMP 473 (2016-2018)
Duration: 28 February 2022 - 26 June 2022
Starts: Trimester 1
Campus: Kelburn