Introduction

Introduction to the course

Course description

Increasingly, organisations need to analyse enormous data sets to determine useful structures in them. In response to this, a range of statistical and machine learning methods have been developed in recent times. This course covers the key techniques in data mining and machine learning with theoretical background and applications. The topics include methods such as linear and logistic regression, neural networks, Bayesian neural networks, clustering and dimensionality reduction, ensemble learning, and an introduction to deep learning. Emerging machine learning tools and libraries are used to illustrate the methods in programming environments that includes Python and R.

- Week 0 Python and R Tutorials (no Lectures)
- Week 1 Data Processing and Introduction to Data mining
- Week 2 Logistic Regression and Evaluation
- Week 3 Intro to Neural Networks
- Week 4 Advances in Neural Networks
- Week 5 Bayesian Neural Networks
- Week 6 Break (no Lectures)
- Week 7 Trees and Forests
- Week 8 Ensemble Learning
- Week 9 Dimensionality Reduction
- Week 10 Emerging Topics in Data Mining

Course Coordinator and Lecturer

Dr Rohitash Chandra

Dr Rohitash Chandra is a Senior Lecturer in Data Science at the UNSW School of Mathematics and Statistics. Dr Chandra leads a program of research encircling methodologies and applications of artificial intelligence; particularly in areas of Bayesian deep learning, neuro-evolution, climate extremes, geoscientific models, and mineral exploration. Dr Chandra has developed novel methods for machine learning inspired by neural systems and learning behaviour that include transfer and multi-task learning, with the goal of modular deep learning. His current interest is in deep learning model development with application to language models, vaccine research, and COVID-19. More information Email: rohitash.chandra@unsw.edu.au

Ratneel Deo (Tutorial/Lab Support)

Mr Ratneel Deo is a PhD student in applied machine learning with a focus on reef and coastal development over thousands of years. His research focuses on unsupervised machine learning methods for the analysis of reef cores that represent over 10,000 years of climate history. In the past, Mr Deo has applied deep learning methods for cyclone track and intensity prediction. He has been an Assistant Lecturer in Computer Science at the University of the South Pacific (2018 - 2021) and comes with a strong programming background. More information

email: r.deo@unsw.edu.au

Arpit Kapoor (Tutorial/Lab Support)

Mr Arpit Kapoor is a PhD student in applied machine learning with a focus on hydrological models and climate change. More information

email: arpit.kapoor@unsw.edu.au

Lecture sessions

1. Monday: 2-4 pm, Ritchie Theatre (K-G19-LG02)
2. Wednesday: 1 -2 pm, Colombo Theatre C (K-B16-LG05)

Note **Week 1 Lecture will be held online** via BB Collaborate and **all lectures will have recordings available** via BB Collaborate (accessed from Moodle).

Tutors for lab sessions

1. Ratneel Deo: Wednesday 5 pm and Thursday 10 am lab session. Location: H13-E G13D (Weeks 2-5; 7-10)
2. Arpit Kapoor: Thursday 11 am and 12 pm lab session. Location: H13-E G13D (Weeks 2-5; 7-10)
3. Rohitash Chandra: Wednesday 3pm and 4pm lab session. Location: H13-E G13D (Weeks 2-5; 7-10): **Note that Wednesday 4 pm lab recording via BB Collaborate will be available only.**

Attendance is not mandatory in any lab sessions and none of the lab and lecture sessions will feature any assessments. We **highly recommend** that you attend **all lecture and lab sessions face-to-face**.

General Assessment Information

- **Assessment 1: Quiz (5 %).** You will engage in an online open-book quiz covering the first three weeks of the course content. This is a **30-minute** online quiz that is held **Tuesday at 6 pm Week 3 online via BB Collaborate.**
- **Assesment 2: Model building and evaluation (15 %).** The assignment provides an opportunity for you to apply existing machine learning-based model code on benchmark datasets and provides an opportunity to learn about the model evaluation and reporting of results. **This is due 10 pm Friday - Week 5.**
- **Assesment 3: Project (25%).** In this project, you will work in a small group to apply machine learning methods in real-world applications. It will provide the opportunity to learn about the model evaluation and reporting of results. It will also enhance skills in technical report writing that incorporate a literature review with a comprehensive presentation of results. The task

involves the submission of code and a technical report which is at least a thousand words long and includes at least 10 references. The report also includes the results from experiments presented as Tables and Figures that are discussed in detail. The report is accompanied by code and data that are submitted online. **This is due 10 pm Friday - Week 10.**

- **Assesment 4: Final Exam (55%).** The final exam is designed to test your learning and problem-solving skills on all topics delivered across the term. The exam consists of MCQ, short answer responses, and practical, i.e. – coding and problem-solving components. The details will be confirmed during the course. The examination will occur during the official university examination period. A sample final exam is available.

* **Requirements to pass the course: At least 50 % in the final exam.**

Consultation Session: Mondays - 4 - 5 p.m after Lecture - Dr Rohitash Chandra's office (Red Centre Building - Level 2)

Textbook
Print:

1. https://www.bookshop.unsw.edu.au/details.cgi?ITEMNO=9781492032649
2. https://www.bookshop.unsw.edu.au/details.cgi?ITEMNO=9781138492530
3. https://www.amazon.com.au/Hands-Machine-Learning-Scikit-Learn-TensorFlow/dp/1492032646

Digital:

1. https://unswbookshop.vitalsource.com/products/-v9781492032595
2. https://unswbookshop.vitalsource.com/products/-v9781000731071

This title is out of print but available in the Library

https://www.bookshop.unsw.edu.au/details.cgi?ITEMNO=9780070428072

Additional text:

Friedman, J. H. (2017). *The elements of statistical learning: Data mining, inference, and prediction*. springer open. https://web.stanford.edu/~hastie/Papers/ESLII.pdf (open book)