Introduction

Semester 1, 2021

Kris Ehinger

Outline

- What is machine learning?
- Welcome to COMP30027
- Overview of machine learning

What is machine learning?

Demo

Watch closely!

Demo

What did you see?

Problem

- Lots of data!
- How to use it?
  - What parts are meaningful?
  - What parts are noise?
  - How do you separate these?

Definition of machine learning

- Automatic extraction of valid, novel, useful, and comprehensible knowledge (rules, regularities, patterns, constraints, models, …) from arbitrary sets of data

Machine learning tasks

- Classification
- Clustering
- Regression
- Probability estimation
- Sequence discovery
- Association rule mining
- Model fitting
- …

ML vs. data mining / data science?

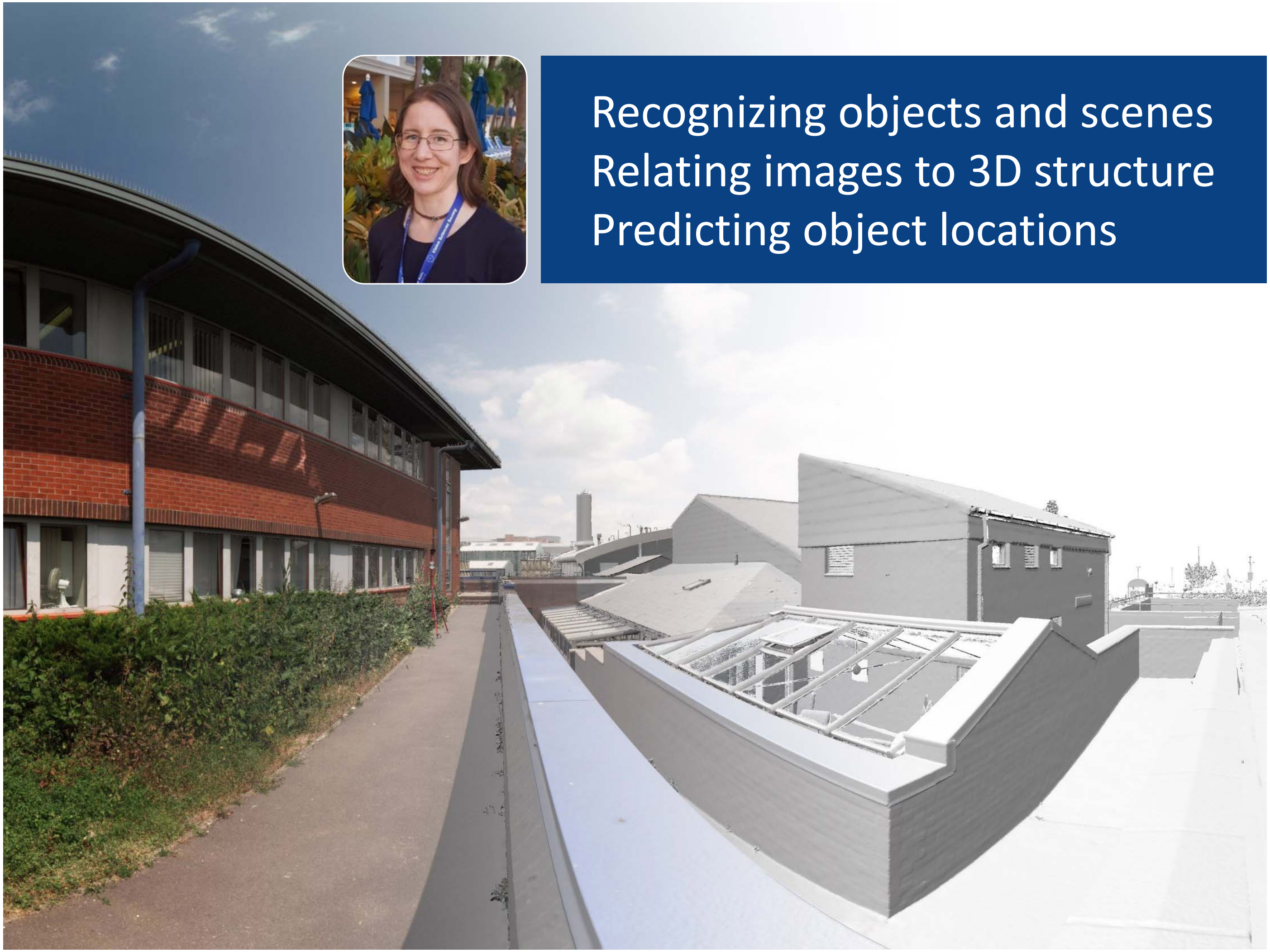
- Machine learning tends to:
  - Focus on theory more than application
  - Ignore problems of run time/complexity
- Data mining tends to:
  - Focus more on applications than theory
  - Worry about run time and scalability
- Data science tends to:
  - Combine elements of both
  - Focus on interpreting and communicating data insights
- …but there's a lot of overlap between all three

Welcome to COMP30027

Lecturers

- Kris Ehinger (subject co-ordinator)
- kris.ehinger@unimelb.edu.au
- Consultation: 1pm Fridays (on Zoom)

- Ling Luo
- ling.luo@unimelb.edu.au
- Consultation: TBD



Recognizing objects and scenes

Relating images to 3D structure

Predicting object locations

Tutors

- Hasti Samadi (head tutor)

- Kazi Adnan
- Shreyasi Datta
- Yujing Jiang
- Masoud Khorasani
- Sadia Nawaz
- Ali Qadar
- Shima Rashidi
- Aref Rekavandi
- Amila Silva
- Justin Tan
- Yifei Wang
- Hanming Zheng

Student representatives

- Two (2) volunteers needed!
- Responsibilities
  - Collect feedback from classmates
  - Attend a staff-student liaison committee meeting
- Benefits
  - Public speaking experience
  - Get to know CIS staff
  - Put it on your CV!
- Email lecturers if interested in volunteering

Contacting us

- General inquiries: Ed forum on LMS
  - We encourage all students to join in discussions – answering other students' questions is one of the best ways to improve your own understanding
  - Please do not post sections of your code or reports publicly! If you must include these, private-message the instructors
- Personal/private concerns: Email the instructors
  - If you email us about a general inquiry, we may ask you to re-post your question in the forum
- Please include COMP30027 in email subject

Dual delivery

- Lectures are online only
- Exams are online only
- Choice of online or in-person workshops

Lectures

- Mondays and Thursdays, 5.15-6.15pm
- Online via Zoom – links on Canvas
- Lecture recordings will be posted on Canvas the following day

Tutorials (starting week 2)

Day	Start	End	Location
Monday	9:00	10:00	
Monday	9:00	10:00	
Monday	9:00	10:00	PAR-100 Leicester Street-106
Monday	14:15	15:15	
Tuesday	11:00	12:00	PAR-100 Leicester Street-106
Tuesday	16:15	17:15	
Tuesday	17:15	18:15	
Tuesday	17:15	18:15	PAR-100 Leicester Street-106
Wednesday	14:15	15:15	
Wednesday	15:15	16:15	PAR-100 Leicester Street-106
Wednesday	17:15	18:15	
Wednesday	17:15	18:15	PAR-100 Leicester Street-107
Thursday	16:15	17:15	
Thursday	18:15	19:15	
Friday	12:00	13:00	
Friday	14:15	15:15	PAR-100 Leicester Street-106
Friday	17:15	18:15	
Friday	17:15	18:15	PAR-100 Leicester Street-106

Practicals (starting week 2)

Day	Start	End	Location
Tuesday	9:00	10:00	PAR-100 Leicester Street-107
Tuesday	10:00	11:00	PAR-100 Leicester Street-107
Tuesday	11:00	12:00	
Tuesday	14:15	15:15	
Tuesday	15:15	16:15	PAR-100 Leicester Street-107
Tuesday	16:15	17:15	
Wednesday	11:00	12:00	PAR-100 Leicester Street-107
Wednesday	13:00	14:00	
Wednesday	14:15	15:15	
Wednesday	15:15	16:15	
Wednesday	16:15	17:15	PAR-100 Leicester Street-107
Thursday	9:00	10:00	PAR-100 Leicester Street-107
Thursday	11:00	12:00	
Thursday	12:00	13:00	
Thursday	12:00	13:00	
Friday	11:00	12:00	PAR-100 Leicester Street-107
Friday	14:15	15:15	
Friday	15:15	16:15	PAR-100 Leicester Street-107

COVIDSafe Campus

Stay COVIDSafe on campus

If you haven't already, you must complete the COVIDSafe module and health declaration immediately:

students.unimelb.edu.au/COVIDSafe

WHAT WE DO NOW BECOMES WHAT HAPPENS NEXT

Completing this module is a requirement for being on campus, and you must follow all COVIDSafe guidelines while you're here.

THE UNIVERSITY OF MELBOURNE

unimelb.edu.au/coronavirus

V2 Feb 2021

COVIDSafe Campus

Stay COVIDSafe on campus

If you start to feel unwell while on campus, leave immediately.
Call the University Health Services COVID-19 Hotline at 8344 6905. You may be able to get a COVID-19 test before going home.

We encourage everyone to scan QR codes
in all areas you visit to assist with contact tracing if needed.

Where possible, keep at least 1.5 metres between yourself and others.

Remember to sneeze or cough into a tissue or your elbow, followed by hand hygiene.

Wash or sanitise your hands often.

Carry a face mask at all times and wear one as required.
Guidelines are changing often, so please check the latest requirements.

Leave classrooms and surrounding areas promptly.
This assists in maintaining physical distancing during changeover periods between classes.

WHAT WE DO NOW BECOMES WHAT HAPPENS NEXT

STAY SAFE. KEEP EVERYONE SAFE.

COVID-19 hotline: 8344 6905 | services.unimelb.edu.au/health

Stay informed about the latest COVID-19 public health advice, cases and exposure sites at www.coronavirus.vic.gov.au/coronavirus-covid-19-victoria

THE UNIVERSITY OF MELBOURNE

unimelb.edu.au/coronavirus

V2 Feb 2021

Tutorial vs. practical?

- Tutorials will focus on revising theoretical concepts and methods covered in class
  - Work through numerical examples
  - Understand how and why the algorithms work
  - Similar to the type of questions you'll see on final exam
- Practical sessions will give you hands-on experience applying machine learning methods
  - Build familiarity with Python tools like `scikit-learn`
  - Experiment and see results on small data sets
  - Helpful for the assignments

Subject material

- LMS is the primary portal for the subject
  - Lecture schedule, tutorial/practical schedule
  - Content page for each week
- Lecture content
  - Handouts will be posted before lecture
  - Slides and lecture capture available after lecture
- Tutorials/practicals
  - Cover content from previous week's lecture
  - Handouts posted before the first tutorial/practical
  - Solutions posted after the last tutorial/practical

Assessment

- Assignment 1 (20%, week 5)
  - Build a machine learning algorithm, experiment on provided data sets, and answer questions
  - Work in groups of 1-2
- Assignment 2 (20%, week 11)
  - Design a method to solve an open-ended classification problem, present algorithm and experiments in a written report
  - Work in groups of 1-2
- Final exam (60%, during exam period)

Prerequistites

- Programming skills
  - Practicals and assignments are in Python
  - Libraries: `numpy`, `scipy`, `scikit-learn`
- Mathematical skills
  - Basic familiarity with probability, statistics, geometry, linear algebra, and differential calculus
- Data mining skills
  - Reading, writing, sorting, partitioning, cleaning, and visualizing multidimensional data sets
  - Basic clustering / dimensionality reduction methods

How much math?

- Probability

$$L = \sum_i \log \sum_j P(C_j)P(X_i|C_j)$$

- Linear algebra

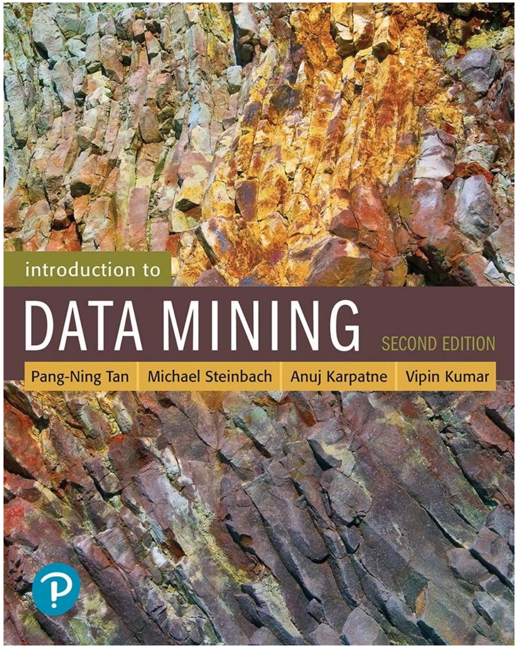
$$a_i = f\left(\left[\sum_j w_j x_{ij}\right] + b\right) = f(\mathbf{w} \cdot \mathbf{x}_i + b)$$

Textbooks

- Suggested links and readings will be posted on LMS each week
- Readings are not required – optional links to expand your knowledge of the week's topics if you are interested

Textbooks

- Pang-Ning Tan, Michael Steinbach and Vipin Kumar (2005) *Introduction to Data Mining*, Addison-Wesley.



Textbooks

- M. Mitchell (1997) *Machine Learning*, WCB/McGraw-Hill.
- Trevor Hastie, Robert Tibshirani and Jerome Friedman (2009) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer.
- Ian Witten, Eibe Frank, and Mark A. Hall (2011) *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann.
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville (2016) *Deep Learning*, MIT Press.

To do (this week!)

- Join the Ed forum using invite in your email

- Install Jupyter Notebook

- Complete COVIDSafe module and health declaration (if you plan to come to campus)

Overview of machine learning

What is learning?

- Why do we learn?
- What does it mean to “learn” something?

What is learning?

What is learning?

- Learning is fitting a function to data, which allows a mapping from every possible input to an output:
  - output = f(input)
- Examples:
  - Learning multiplication
  - Learning how to ride a bike
- Learning makes it possible to **generalise**: produce an output for any input, even inputs you've never seen before

Why generalise?

- When *don't* you need to learn?
- If the input set is finite, and you can memorize all the mappings from input->output, there is no need to generalise
  - Requires a small input set, or large memory space
- If there is no rule at all that could relate input to output, the only solution is to memorize
- Does memorization work for real-world problems?
  - Generally no, because the input is continuous (infinite values) or because we want to predict future events

Learning in practice

- Machine learning tasks
  - Classification – predict discrete class labels based on features
  - Regression – predict continuous outcomes based on features
  - Predict relationships between features / outcomes (sequence learning, association rule mining)
  - Understand and reconstruct the processes that produced the features / outcomes (model fitting, probability estimation)

Subject content

- Specific machine learning methods
  - Naïve Bayes classifiers
  - Decision trees
  - Support vector machines
  - Linear regression
  - Logistic regression
  - Gaussian mixture models
  - Hidden Markov models
  - Perceptron
  - Deep neural networks

Subject content

- Machine learning competence
  - Training models
  - Evaluating models
  - Interpreting model performance
  - Choosing the right model for a task

Subject objectives

- Recognize real-world problems amenable to machine learning
- Apply machine learning techniques correctly to realistic problems
- Interpret the results of machine learning methods on real data
- Compare benefits/drawbacks of various models and techniques
- Understand the statistical principles behind machine learning methods

Overlap with other subjects

- Points of contact between machine learning and:
  - Statistics
  - Artificial intelligence
  - Information theory / computational theory / complexity
  - Many applied fields (business, finance, health, government, earth sciences, biology, neuroscience, etc.)