

Lecture 1: Introduction and Overview

COMP90049

Introduction to Machine Learning

Semester 1, 2024

Hasti Samadi, Kris Ehinger and Ting Dang

Copyright @ University of Melbourne 2024. All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the author.

Acknowledgement: Lea Frermann and Lida Rashidi



This lecture

- Introduction and Warm-up
- Housekeeping COMP90049
- Machine Learning



Introduction & Warm-up

About Me

- Part of COMP90049 Since 2018
- Research interest area Teaching and Learning
- Background Software Engineering and Business Analysis
- 8 years in industry
- 8 years in academics

About you

Please go to: pollev.com/hs2024

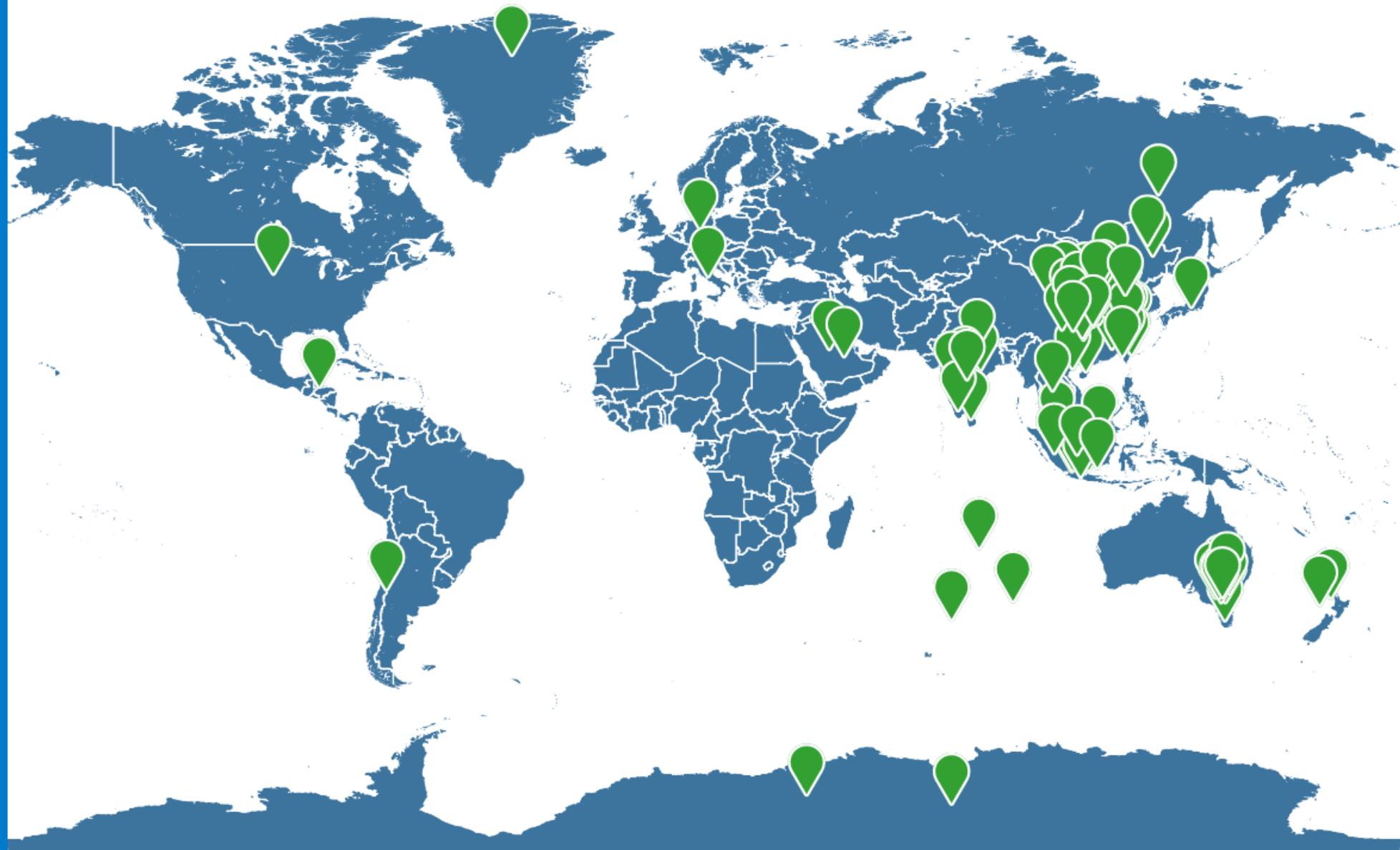


About you

110 surveys completed

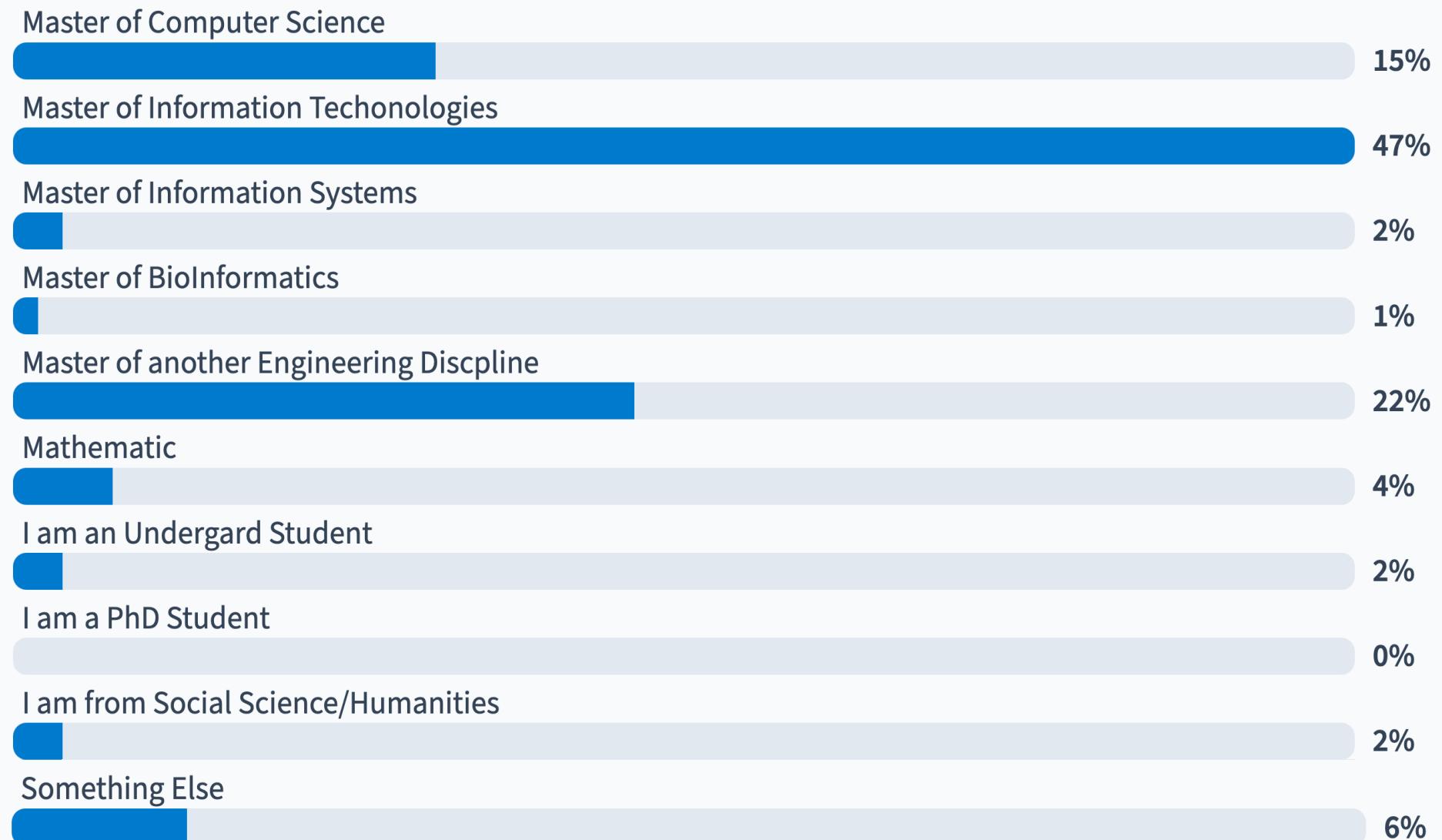
18 surveys underway

Where do you call home?



Start the presentation to see live content. For screen share software, share the entire screen. Get help at pollev.com/app

Which of the following describes your academic background best?



How do you rate your mathematic skills?

A little rusty — 1



With a little support, I will be okay - 2



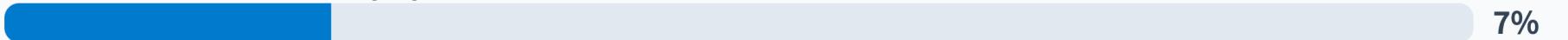
I think I am fine — 3



I am rather cofident with my math skills - 4



I solve derivatives with my eyes closed — 5



How do you rate your Python skills?

I'm still in the 'trial and error' phase, but I'm getting there.

20%

I'm confident enough to write some basic scripts, but I wouldn't trust me with anything too comp...

32%

I'm at an intermediate level – comfortable with most tasks but still learning the finer points.

31%

I'm pretty solid with Python. I can handle most projects without too much trouble.

13%

Python is my jam!

5%



What is Learning?



Please go to: pollev.com/hs2024



What is Learning?

Finding patterns

 0  0



Gaining knowledge

 0  0



Through gaining information and experience. Becoming more powerful

 0  0



To acquire new information which you do not know, to understand about something which is new to you

 0  0



Brainstorm / Discuss

What is Machine Learning?

Please go to: pollev.com/hs2024



What is Machine Learning?

The process of converting data into useful information

Like 2 Dislike 0



Blackbox to let computer do complex tasks

Like 2 Dislike 0



More pattern recognition

Like 2 Dislike 0



honestly right now to me it's a buzzword, but hopefully this class will change that :)

Like 1 Dislike 0



Definitions of Machine Learning

“We are drowning in information, but we are starved for knowledge”

John Naisbitt, Megatrends

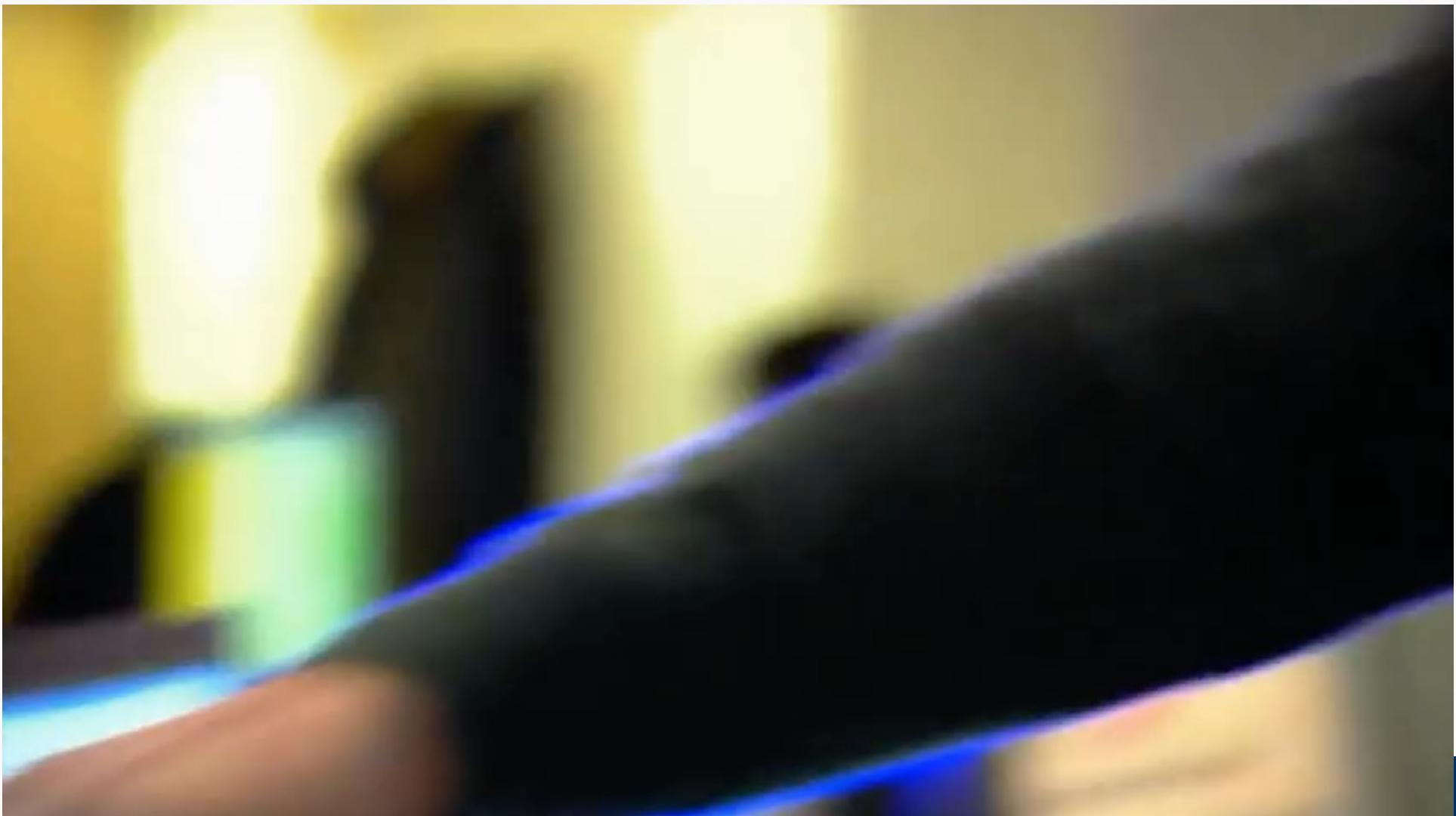


Our definition of Machine Learning

automatic extraction of *valid, novel, useful* and *comprehensible knowledge* (rules, regularities, patterns, constraints, models, ...) from arbitrary sets of **data**



Let's see a short video



Definitions of Machine Learning

Learning what?

- **Task** to accomplish a goal, e.g.,
 - Assign a label to inputs (finding letter A)
 - Group inputs into known classes (email → {spam, no-spam})
 - Understand regularities in the data

Learning from what?

- **Data**
- Where do the data come from? Is it reliable? Representative?

How do we learn?

- define a **model** that explains how to get from input to output
- derive a **learning algorithm** to find the best model parameters

How do we know learning is happening?

- The algorithm improves at its task with exposure to more data
- We need to be able to **evaluate** performance objectively



- **Topics:**
 - classification,
 - clustering,
 - optimization,
 - semi-supervised learning,
 - neural networks,
 - anomaly detection
 - and more...
- All from a theoretical and practical perspective
- Refreshers on maths and programming basics
- Theory in the lectures
- Hands-on experience in workshops and projects
- **Guest lecture 1:** academic writing skills
- **Guest lecture 2:** Industry talk with a focus on large language model (LLM) and Generative AI



Intended learning outcomes

On completion of this subject students are expected to be able to:

- ✓ ILO 1 - Apply elementary mathematical concepts used in machine learning
- ✓ ILO 2 - Derive machine learning models from first principles
- ✓ ILO 3 - Design, implement, and evaluate machine learning systems for real-world problems
- ✓ ILO 4 - Identify the correct machine learning model for a given real-world problem



Expected Background

Programming concepts

- We will be using **Python** and **Jupyter Notebooks**
- Basic familiarity with libraries (NumPy, scikit-learn, SciPy, pandas)
- You need to be able to write code to process your data, apply different algorithms, and evaluate the output
- Optional practice/demo Jupyter notebooks (from week 1) [available on Canvas](#)

Mathematical concepts

- formal maths notation
- basic probability, statistics, calculus, geometry, linear algebra
- (why?)



What Level of Maths are we Talking?

$$\ln \frac{P(y = \text{true}|x)}{1 - P(y = \text{true}|x)} = w \cdot f$$

$$\frac{P(y = \text{true}|x)}{1 - P(y = \text{true}|x)} = e^{w \cdot f}$$

$$P(y = \text{true}|x) = e^{w \cdot f} / (1 + e^{w \cdot f})$$

$$P(y = \text{true}|x) + e^{w \cdot f} P(y = \text{true}|x) = e^{w \cdot f}$$

$$P(y = \text{true}|x) = h(x) = \frac{e^{w \cdot f}}{1 + e^{w \cdot f}} = \frac{1}{1 + e^{-w \cdot f}}$$

$$P(y = \text{false}|x) = \frac{1}{1 + e^{w \cdot f}} = \frac{e^{-w \cdot f}}{1 + e^{-w \cdot f}}$$



COMP90049
HouseKeeping

COMP90049 – Teaching Staff



Hasti Samadi
Coordinator



Kris Ehinger
Lecturer (Week 1-4)



Ting Dang
Lecturer (Week 5-12)



Jia
Head Tutor

- First point of contact
- Consultations
- Assignment logistics
- Discussion board



COMP90049 – Teaching Staff

| | | |
|-------------|-------------------|--|
| Coordinator | Hasti Samadi | hasti.samadi@unimelb.edu.au |
| Lecturers | Kris Ehinger | kris.ehinger@unimelb.edu.au |
| | Ting Dang | ting.dang@unimelb.edu.au |
| Head Tutor | Jiayang Ao | jiayanga@student.unimelb.edu.au |
| Tutors | Jiayang Ao | jiayanga@student.unimelb.edu.au |
| | Kazi Adnan | kazi.adnan@unimelb.edu.au |
| | Mojgan | mkouhounesta@student.unimelb.edu.au |
| | Kouhounestani | |
| | Behzad Moradi | moradib@unimelb.edu.au |
| | Sayantan Dasgupta | sayantan@unimelb.edu.au |
| | Reyhaneh Goli | reyhaneh.goli@unimelb.edu.au |

Find more about our Staff on Canvas



Lecture Materials

- The following outline is broadly indicative of the content that we expect to cover; it is subject to change throughout the semester.
- Most weeks, we will release an ungraded quiz and some example Python code
- Optional further readings are listed [here](#) (and at the end of each slide deck)

| Week | Tuesday Lecture | Thursday Lecture | Assessment | Other Materials |
|---------------------|--|---|---------------|---------------------------|
| 1 (27 Feb & 29 Feb) | 1: Subject Introduction [ slides  <td>2: Machine Learning Concepts & Linear Regression [slides]</td> <td></td> <td>[quiz] [code]</td> | 2: Machine Learning Concepts & Linear Regression [slides] | | [quiz] [code] |
| 2 (5 Mar & 7 Mar) | 3: K-Nearest Neighbours [slides] | 4: Probability (recap) [slides] | | [quiz1] [code] [quiz2] |
| 3 (12 Mar & 14 Mar) | 5: Decision Trees [slides] | 6: Evaluation I [slides] | | [quiz1] [code] [quiz2] |
| 4 (19 Mar & 21 Mar) | 7: Exact Optimisation & Naive Bayes I [slides] | 8: Naive Bayes II [slides] | A1 (Rel: Mon) | [quiz1] [code] [quiz2] |
| 5 (26 Mar & 28 Mar) | 9: Feature Selection [slides] | 10: Iterative Optimisation [slides] | | [quiz1] [code] [quiz2] |
| BREAK | | | | |
| 6 (9 Apr & 11 Apr) | 11: Logistic Regression [slides] | 12: Evaluation II [slides] | A1 (Due: Fri) | [quiz1] [code] [quiz2] |
| 7 (16 Apr & 18 Apr) | 13: Perceptron [slides] | 14: Neural Networks [slides] | A2 (Rel: Mon) | [quiz1] [code] [quiz2] |
| 8 (23 Apr) | 15: Backpropagation | 16: Academic Writing (Pre-recorded lecture) | | [quiz11] [code] |

Let's
check
the
subject
Canvas
page



This subject is offered **on campus only**

- The lectures are **on campus**
- Workshops are **on campus**
- All lectures are **live-streamed** through Lecture Capture
- All lectures will be recorded. All recordings and other materials will be made available online through Canvas
- Attending the workshops is **highly recommended**.
- From week 2, we provide **additional** coding practice sessions for students seeking extra support to enhance their coding skills.
- We can have weekly Math consultation sessions based on the interest.



Coding Practice Session

- Start from week 2
- For people who are not confident in their coding skills
- Fri 3-5 PM and Wednesday 5-7 PM
- Please enrol [here](#)
- Bring your laptops with you



Coding Workshop Content

- Practical exercises
- Revising concepts from the lectures via coding examples
- You will have access to the provided content via Canvas at the beginning of each week



Math Consultation Session

- Start from week 2
- For people who have questions about the math parts
- Time slot(s) will be announced shortly
- Please enrol [here](#)
- Will be online and via Zoom



Consultation Content

- Unlike Coding Practice sessions, there is no content for math
- It will be Q&A sessions, where you show up with your questions



Lectures

Lecture 1

Tue 13:00-14:00

Sidney Myer Asia Centre - B02
(Carrillo Gantner Theatre)

Lecture 2

Thu 14:15-15:15

Peter Hall Building – G01
(JH Michell Theatre)

Lecture content

- Theory
- Derivation of ML algorithms from scratch
- Motivation and context
- Some coding demos in Python



Workshops

- Start from week 2
- 1 hour per week
- 11 slots, please sign up and stick to one
- All workshops will be on campus

Workshop Content

- Practical exercises
- Working through numerical examples
- Revising theoretical concepts from the lectures
- You will have access to the provided content via Canvas at the beginning of each week
- The solution will be released at the end of each week (Friday Morning)



COMP90049 – Workshops Schedule

| Tutorial Code | Tutorial Day | Tutorial Time | Tutorial Venue | Tutor |
|---------------|--------------|---------------|-----------------|----------------------|
| T01/01 | Monday | 11:00AM | PAR-379-B1-B117 | Jiayang ao |
| T01/02 | Monday | 01:00PM | PAR-379-B1-B116 | Jiayang ao |
| T01/03 | Wednesday | 11:00AM | PAR-193-L1-122 | Mojgan Kouhounestani |
| T01/04 | Monday | 10:00AM | PAR-193-L1-122 | Behzad Moradi |
| T01/05 | Tuesday | 05:15PM | PAR-379-B1-B132 | Sayantan Dasgupta |
| T01/06 | Wednesday | 05:15PM | PAR-379-B1-B132 | Kazi Abir Adnan |
| T01/07 | Tuesday | 09:00AM | PAR-379-B1-B132 | Behzad Moradi |
| T01/08 | Thursday | 04:15PM | PAR-278-L1-108 | Mojgan Kouhounestani |
| T01/09 | Wednesday | 06:15PM | PAR-379-B1-B132 | Kazi Abir Adnan |
| T01/10 | Tuesday | 06:15PM | PAR-379-B1-B132 | Sayantan Dasgupta |
| T01/11 | Thursday | 11:00AM | PAR-379-B1-B115 | Reyhaneh Goli |



For general questions

- **Default:** Post on the Ed discussion board
- **Backup option 1:** Email the head tutor (Jia) or your tutor
- **Backup option 2:** Email the Subject Coordinator

Ed

- Actively engage by **asking and answering** questions. Peer teaching is the most effective way of learning!
- (Of course, no assignment solutions should be given away. Doing – or asking for – this is academic misconduct.)

Personal/private concerns: Email the head tutor or lecturer, e.g.,

- With specific assignment questions
- With private or personal concerns
- Constructive feedback is always very welcome!
- Please include COMP90049 in the email subject



We need 2 or 3 student representatives

- Communication channel between class and teaching team
- Collect and pass on (anonymous) feedback or complaints
- Attend a student-staff meeting during the semester (TBD)
- Represent the **diversity** of the class

Interested? Send me an email with a short paragraph on why you want this role.



Weekly Consultations

- Every Tuesday 10-11 AM consultations by the Hasti online and on-site.
- On-site: Room3202, Lv3, Melbourne Connect
- Online:
<https://unimelb.zoom.us/j/85462515749?pwd=bUY1K3pmZXF0M25tbytZbjJEbm1TZh09&from=addon>

Lecturer's Consultation

- If you require consultation with the lecturer for that week, kindly send them an email to arrange a consultation hour.

Assignment consultations

- 1-2 sessions per assignment for clarification.
- Usually, half a week to a week before submission.

Discussion Board

- Ask all your questions on the discussion Board (Ed)
- All questions will be answered within 48 hrs (usually less)



Assessment

One small coding project (20%)

- Release week 4, due week 6
- Jupyter notebooks; Read in data, apply ML algorithm(s), evaluate.

Open-ended research project (30%)

- Release week 7, due week 11&12
- You will be given a data set to formulate a research question and write a short research paper on your findings.

Final exam (50%)

- during exam period
- 2 hours; pen and paper
- **Hurdle requirement:** you have to pass the exam ($\geq 50\%$).

If you have an Academic Adjustment Plan (AAP) please send it to Jia (and Me) as soon as possible to be added to your records.



Academic Honesty

- Videos & Quiz
- Linked from Canvas 'Home' page (or in Modules)
- CIS-specific scenarios

▼ CIS Academic Honesty Training Complete All Items

Videos

-  **Getting help from non-student friends**
Mark as done
-  **Copying the answer from a fellow student**
Mark as done
-  **Getting help from fellow students**
Mark as done
-  **Copying the answer from online sources**
-  **Do not outsource assignments**
Mark as done
-  **Lock screen when leaving your monitor**
Mark as done
-  **Protect your code (Do not share on Github)**
Mark as done
-  **Working and discussing with friends in the right way**
Mark as done

Quiz

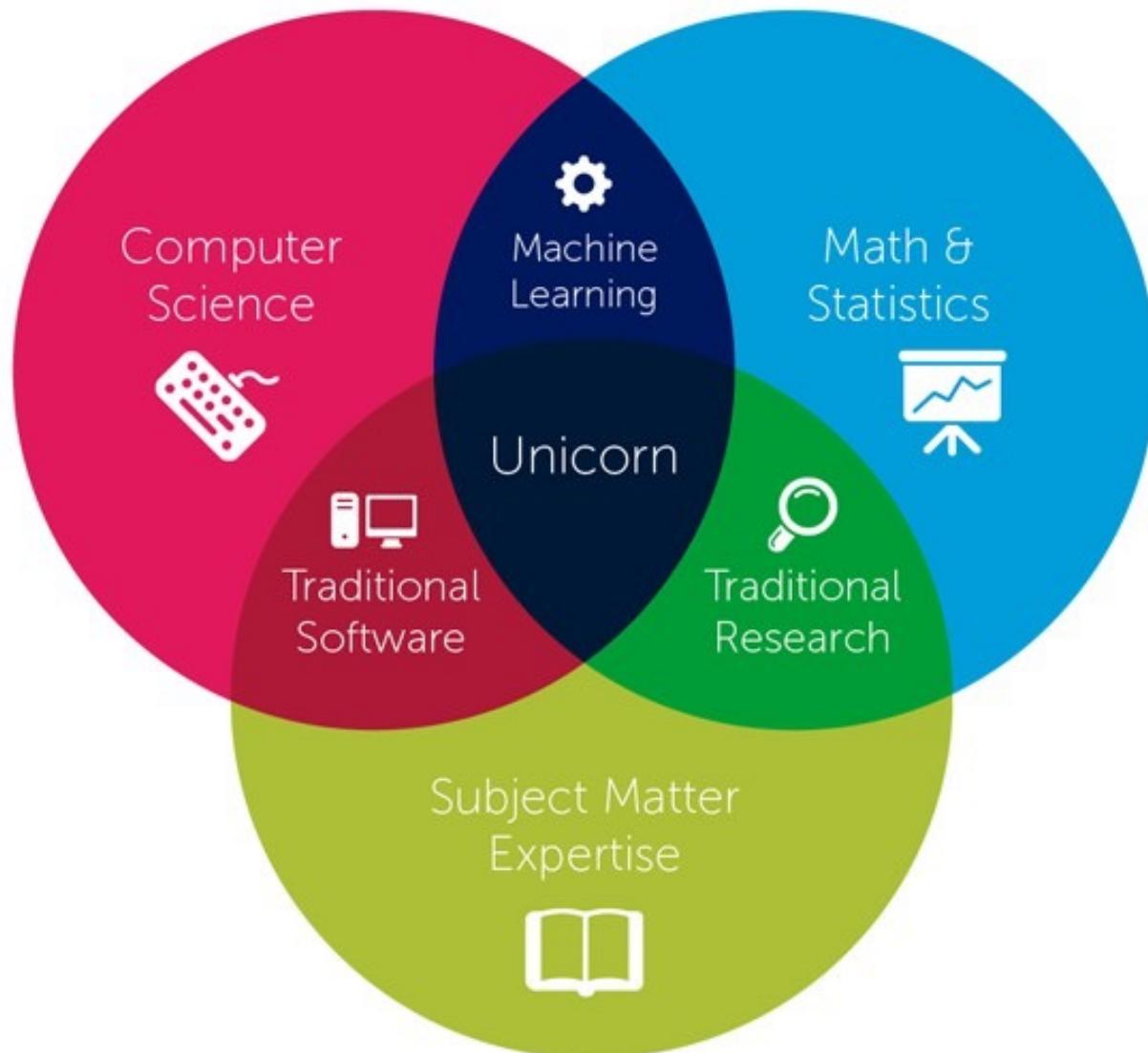
Further information

-  **Academic Integrity Principles at Unimelb**
-  **Further Resources**



What and Why of Machine Learning?

What is Machine Learning?



Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this image,
provided that this copyright notice remains intact.



What is Machine Learning?



<https://xkcd.com/1838/>

You're sitting in the right class! ☺

Top 5 Occupations In Predicted Percentage Growth

Source:

<https://insightglobal.com/blog/in-demand-careers-2024/>

1. Wind turbine service technician
Predicted growth rate: 45%
Predicted new jobs: 5,000
Median annual salary: \$57,320
2. Nurse practitioner
Predicted growth rate: 45%
Predicted new jobs: 118,500
Median annual salary: \$121,610
3. Data scientist
Predicted growth rate: 35%
Predicted new jobs: 59,400
Median annual salary: \$103,500
4. Statistician (Data science, data analyst, etc.)
Predicted growth rate: 32%
Predicted new jobs: 10,500
Median annual salary: \$98,920
5. Information Security Analyst
Predicted growth rate: 32%
Predicted new jobs: 53,200
Median annual salary: \$112,000



Three ingredients for machine learning

... and related questions



Three ingredients for machine learning

... and related questions

1. Data

- Discrete vs continuous vs ...
- Big data vs small data
- Labeled data vs unlabeled data
- Public vs sensitive data



Three ingredients for machine learning

... and related questions

2. Models

- function mapping from inputs to outputs
- motivated by a *data-generating hypothesis*
- probabilistic machine learning models
- geometric machine learning models
- parameters of the function are unknown



Three ingredients for machine learning

... and related questions

3. Learning

- Improving (on a task) after data is taken into account
- Finding the best model parameters (for a given task)
- Supervised vs. unsupervised learning



More Idea Sharing

Think of **3 words** that describe your expectations of this subject.

Please go to: pollev.com/hs2024



Think of 3 words that describe your expectations of this subject.

practical prediction rewarding standing
want take models high-level helps make profession
passable brain-burning baby stuff hands-on world
start intense awesome easier curve help painful
confusing data learning fun future applicable wam
practice analysis learn hd help skill
rich happy interesting h1 pass work exiting
math barely useful easy cool build final skills
engaging advanced aim machine career exam high
robots expert algorithm difficult startup enjoyable using
maths life demanding challenging group mastering
research knowledge hard graduate master
supportive smooth processing training

The words are presented in a cloud-like arrangement, with three specific words circled in red: "interesting", "useful", and "pass".



ML Example Problem

ML Example Problem

- Scenario 1

You are an archaeologist in charge of classifying a mountain of fossilized bones, and want to quickly identify any “finds of the century” before sending the bones off to a museum

- Solution:



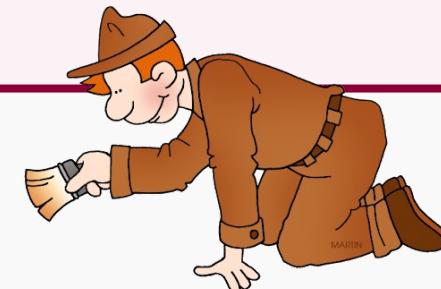
- Scenario 1

You are an archaeologist in charge of classifying a mountain of fossilized bones, and want to quickly identify any “finds of the century” before sending the bones off to a museum

- Solution:

Identify bones which are of different size/dimensions/characteristics to others in the sample and/or pre-identified bones

CLUSTERING/OUTLIER DETECTION



ML Example Problem

- Scenario 2:

You are an archaeologist in charge of classifying a mountain of fossilized bones, and want to come up with a consistent way of determining the species and type of each bone which doesn't require specialist skills

- Solution:



- Scenario 2:

You are an archaeologist in charge of classifying a mountain of fossilized bones, and want to come up with a consistent way of determining the species and type of each bone which doesn't require specialist skills

- Solution:

Identify some easily measurable properties of bones (size, shape, number of “lumps”, ...) and compare any new bones to a pre-classified database of bones

SUPERVISED CLASSIFICATION ;



ML Example Problem

- Scenario 3:

You are in charge of developing the next “release” of Coca-Cola, and you want to be able to estimate how well-received a given recipe will be

- Solution:



- Scenario 3:

You are in charge of developing the next “release” of Coca Cola, and want to be able to estimate how well received a given recipe will be

- Solution:

Carry out taste tests over various “recipes” with varying proportions of sugar, caramel, caffeine, phosphoric acid, coca leaf extract, ... (and any number of “secret” new ingredients), and estimate the function which predicts customer satisfaction from these numbers

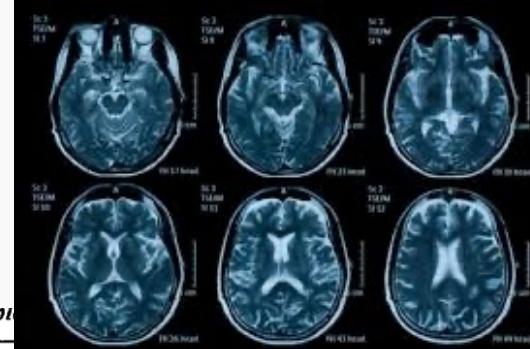
REGRESSION



More Applications

- natural language processing
- image classification
- stock market prediction
- movie recommendation
- web search
- medical diagnoses
- spam/malware detection
- ...

NETFLIX



| No. | Names | Sex | Age | Religion | Work | Role | Mother Tongue | Other Languages |
|-----|---------------|-----|-----|-----------|-----------|---------|---------------|-----------------|
| 1. | Anwar | M | 32 | Muslim | Teacher | Teacher | Thai | English |
| 2. | Serena | F | 25 | Christian | Housewife | Singer | Thai | English |
| 3. | Min etc... | | | | ist | | | |

Google
Ask®
YAHOO!
bing
Aol.



ChatGPT

amazon



Today

- COMP90049 Overview
- What is machine learning?
- Why is it important? Some use cases.

Next lecture: Concepts in machine learning

