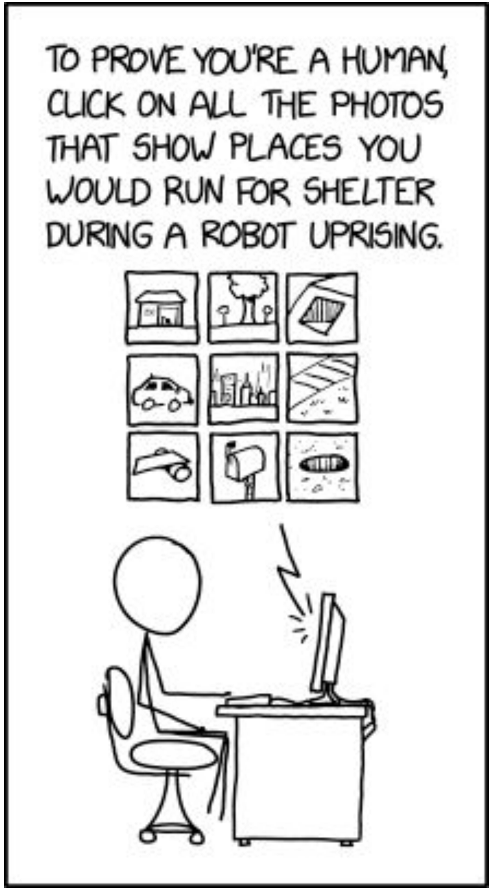
https://xkcd.com/2228/



https://xkcd.com/1897/

COMP4670/8600 Statistical Machine Learning - 2022

Agenda for today

- Course logistics and administrivia
- Helicopter view of ML + Scope of this class

(*Introduction to*) Statistical Machine Learning

This course provides a broad but thorough introduction to the methods and practice of statistical machine learning.

COMP4670/8600

IML — SML — AML

Mathematical thinking in machine learning, rationale, derivation and implementation of a broad set of models. In depth appreciation and critique of model design.

NetworkSci

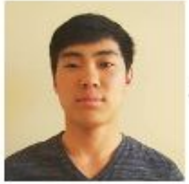
DocAnalysis

Optimisation

AI — AdvAI

SML Team

Lexing Xie



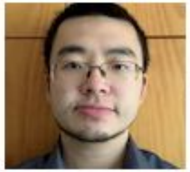
Alexander Soen

Chamin Hewa Koneputugodage

Shidi Li

Tianyu Wang

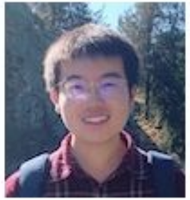
Josh Nguyen



Minchao Wu

Ekaterina (Katya) Nikonova

Ruiqi Li



Haiqing Zhu

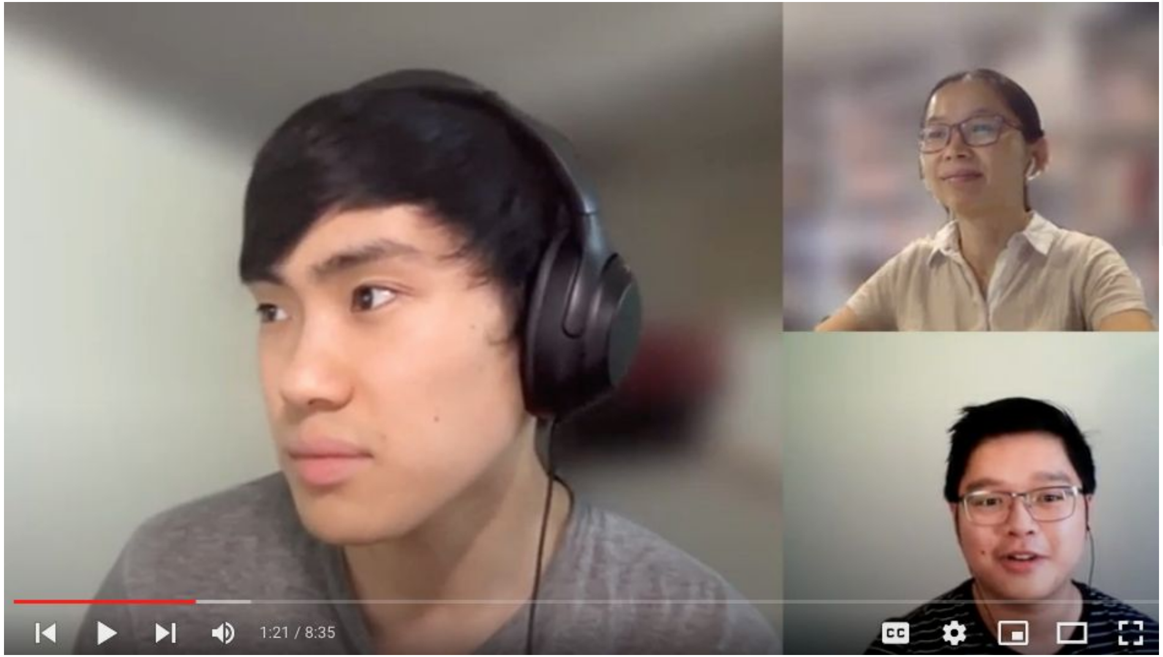
Past lecturers (since 2009)
Chris Weber
Cheng Soon Ong
Christian Walder

Belona Sonna

Dillon Chen

Rong Wang

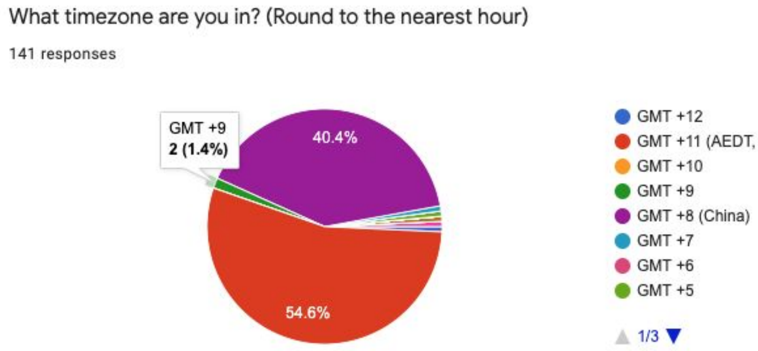
Hello SML videos - for you to get to know the tutors



A few words about yourself, what're you working on
What's your favorite ML algorithm, and why
Your one advice to SML students
One common misconception about machine learning

Logistics + technology

- Microsoft Teams
  - live meetings of “lectures” (recorded)
  - “tutorials/labs” (not recorded)
- In person lectures and labs: TBD
  - We aim to improve the experience in 2022 S1, conditioned on the safety of all students and course staff
  - To be finalised after discussing with tutor team



- Course webpage http://cm.cecs.anu.edu.au/sml2022/ infrequent, read-only updates
- Piazza: Announcements, Questions, and Discussions
- Gradescope: Assignment submission + grading
- Wattle: quiz 1+2, final exam

Course structure

Learning is done by the student, i.e. You :)

We are here to help, and here’re a few ways.

- Online, live lecture sessions
  - **E** [lecture time] ~ 1.5 hours
  - the rest are used to cover assignments, Q&A, etc
- Weekly tutorials with hands-on exercises, starting week 2
  - Signup to an in-person session on wattle if you’re on campus
- Assessment items
  - See website

Studying in unusual times

- Online + offline delivery --- lectures, tutorials, discussions between students
  - We strongly encourage you to study with others students
  - Technology: Zoom, Slack, ...

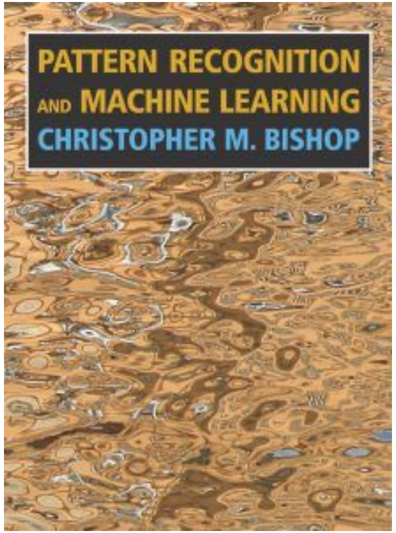
- The two written+programming assignment **can** be submitted **in pairs**
  - Each student can choose to submit the assignment individually or with another student. Students are free to change the teaming between assignment 1 and assignment 2.
  - Students submitting in a pair act as one unit:
    may share resources (such as notes) with each other and write the solutions together
  - Both of the two students should fully understand all the answers in their submission
    Each student in the pair must understand the solution well enough in order to reconstruct it by him/herself
- Video assignment – individual, specs will be available.
- Online quiz x 2
- Online exam (at home)
  - During exam period

Pre-requisites

- Probability
  - random variables, distribution, conditional probability, expectation, variance, density
- Linear algebra
  - matrix multiplication, eigenvector, solving linear systems, matrix inverse
- Python, via jupyter notebook and stand alone functions / scripts.

- This is a mathematically intensive course. But that’s why it’s exciting and rewarding! 😄
- Practice / assess your pre-requisite in this take-home exam
  - Available on Piazza (resource section)
  - Solutions will be released by end of week 1

Text book



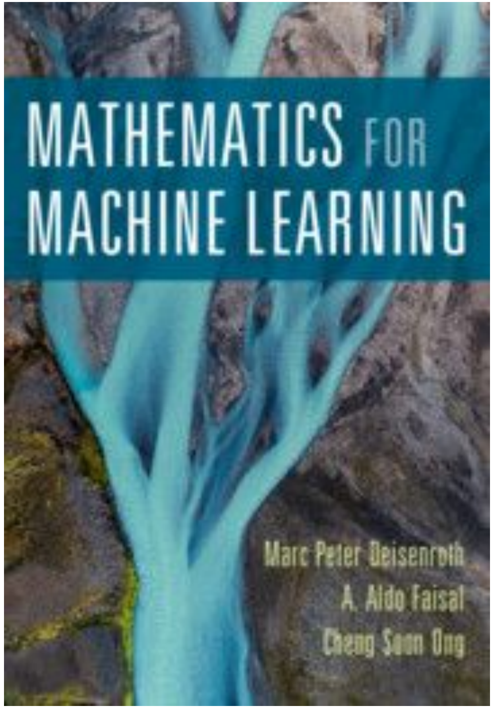
Well-written, pedagogically.
A lot of timeless material.
Note style difference with MML.

A note about exercises in the book and listed on tutorial sheets: Most of them intend to fill in the gaps or enrich the main exposition, few of them are designed to test your understanding to the presented concept or use them in relevant problems.

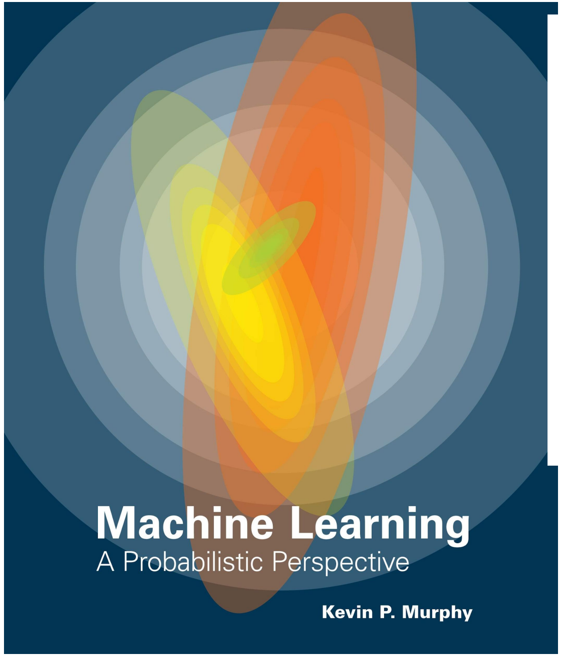
They are not similar to questions in the assignment or exam.

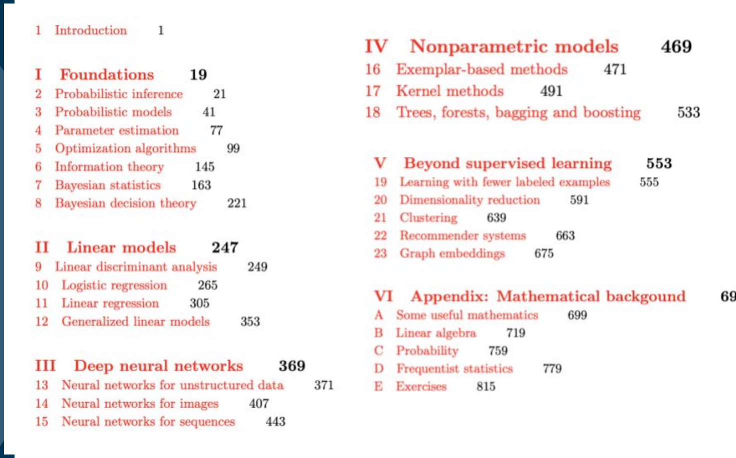
https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/

References



https://mml-book.com





New version https://probml.github.io/pml-book/book1.html

Positioning of this class

“ML Researchers” roles, in analogy to Music
[credit: Preface in MML book]

- Astute listener: cloud-based tools, focus on extracting insight, critique the correctness, interpretation of results, reason about fairness, ethics, etc.
- Experienced artist: understands ML interfaces, their uses, benefits and limits.
- Fledgling Composer: develop and extend existing algorithms, uncover relationships between different tasks. ← **SML**

CECS Class Representatives

Class Student Representation is an important component of the teaching and learning quality assurance and quality improvement processes within the ANU College of Engineering and Computer Science (CECS).

The role of Student Representatives is to provide ongoing constructive feedback on behalf of the student cohort to Course Conveners and to Associate Directors (Education) for continuous improvements to the course.

Roles and responsibilities:

- Act as the official liaison between your peers and convener.
- Be creative, available and proactive in gathering feedback from your classmates.
- Attend regular meetings, and provide reports on course feedback to your course convener
- Close the feedback loop by reporting back to the class the outcomes of your meetings.

Any questions so far?

Scope of this course

What is Machine Learning?

A (simplistic) taxonomy of machine learning

(Biased) samples of ML projects at the ANU

Pressing concerns in fairness, accountability and transparency

What we will not cover

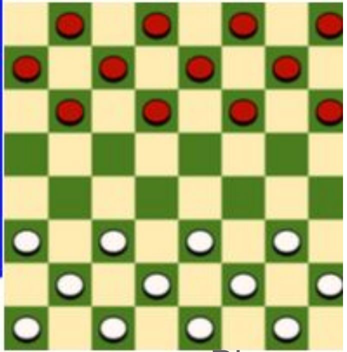
Thanks: many slides are from Stanford CS229

Definition of Machine Learning

Arthur Samuel (1959): Machine Learning is the field of study that gives the computer the ability to learn without being explicitly programmed.

A. L. Samuel*

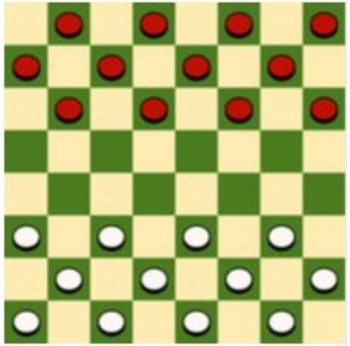
Some Studies in Machine Learning Using the Game of Checkers. II—Recent Progress



Photos from wikipedia

Definition of Machine Learning

Tom Mitchell (1998): a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.



Experience (data): games played by the program (with itself)

Performance measure: winning rate

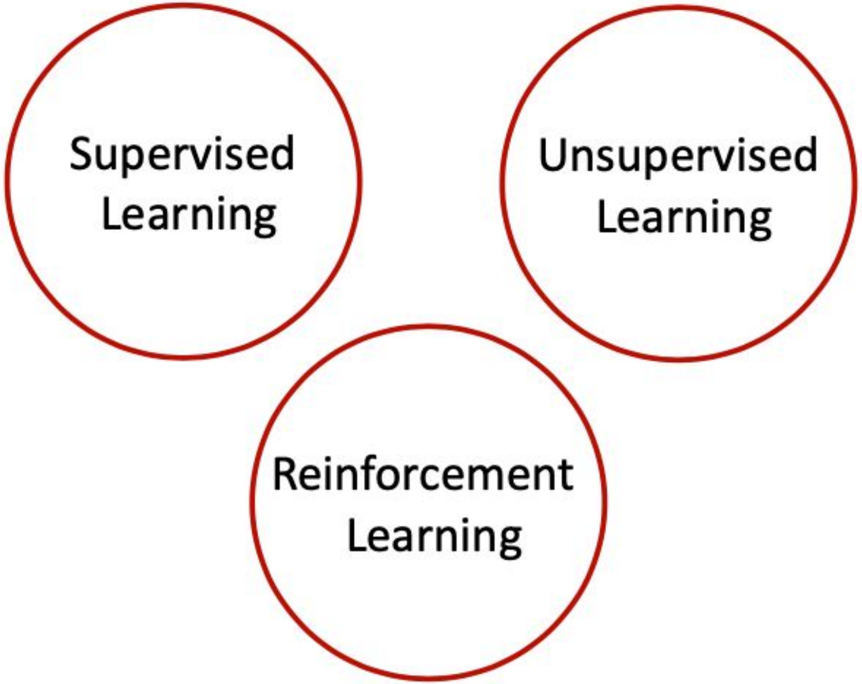
Photos from Tom Mitchell's homepage

Statistical Machine Learning - some history

- 1960's : symbolic AI; computers learn rules from data; analysis of the underlying statistics is seldom done.
- Perceptron (Rosenblatt, 1957), "Perceptrons" (Minsky and Papert, 1969)
-
- 1980's : artificial neural networks
- 1990's - 2000's : statistical machine learning (kernel methods, decision trees, graphical models)
- Why Statistical Machine Learning not earlier?
    - faster computers with larger memory to represent statistical models have become available
    - numerical methods on the desktop computer (BLAS, LAPACK, Optimisation)
    - found new interesting classes of algorithms (e.g. on graphs)
    - large amounts of data available which can be tapped into (flickr, social networks)
    - many data sets with partial/incomplete data (e.g. netflix)

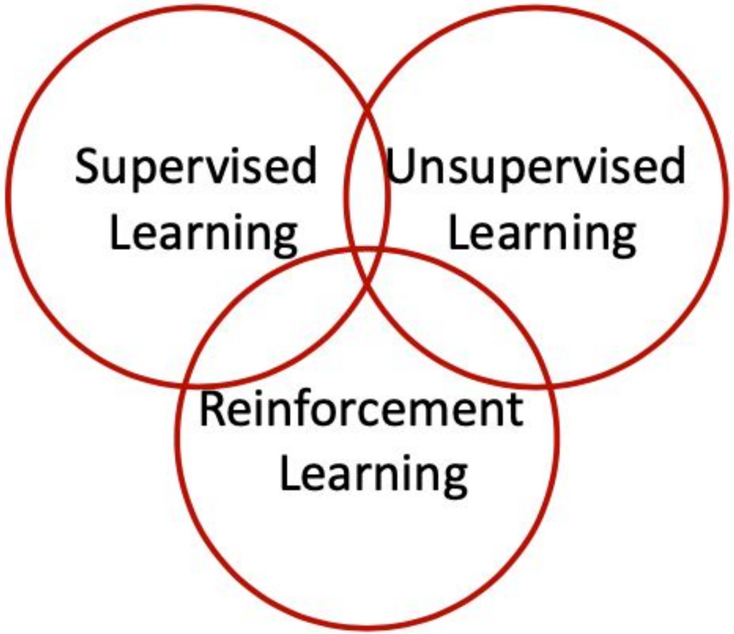
Machine Learning Taxonomy

(A Simplistic View Based on Tasks)



Machine Learning Taxonomy

(A Simplistic View Based on Tasks)



Can also be viewed as sets of tools/methods.

‹ Back to Search | Home › Buy (ACT) › North Canberra › Lyneham › 7 Glover Street

NEW

Add to watchlist ☆

Auction 24/02/21

7 Glover Street, Lyneham ACT 2602

House • 3 1 1 1.0 EER

Block size: **562 m²** approx. UV: **$575,000 (2020)**

BLACKSHAW

Next Inspection: Wed 10 Feb, 5:00 PM

Christine Shaw

Blackshaw Manuka

Call

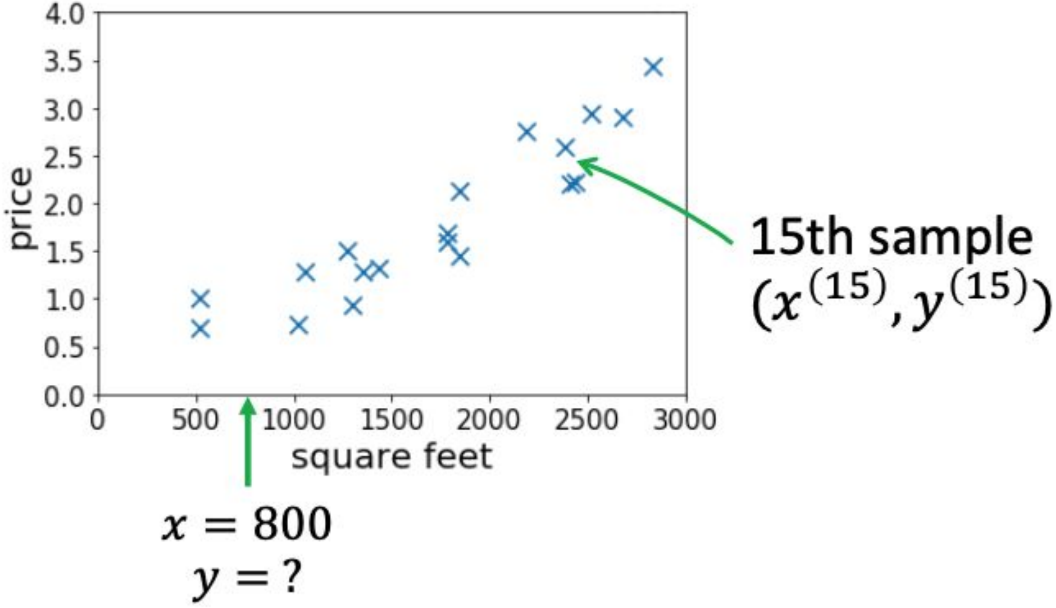
https://www.allhomes.com.au/7-glover-street-lyneham-act-2602

Supervised Learning: Housing Price Prediction

- Given: a dataset that contains n samples

$$\left(x^{(1)}, y^{(1)}\right), \ldots \left(x^{(n)}, y^{(n)}\right)$$

- Task: if a residence has x square feet, predict its price?

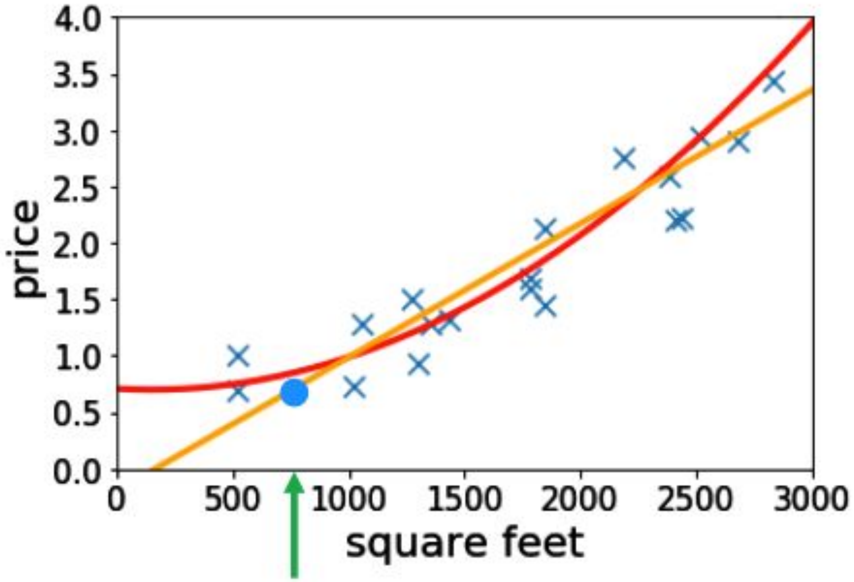


Housing Price Prediction

- Given: a dataset that contains n samples

$$(x^{(1)}, y^{(1)}), \ldots (x^{(n)}, y^{(n)})$$

- Task: if a residence has x square feet, predict its price?



Fit a line or quadratic curve to the data? Week 1 and 2

More features

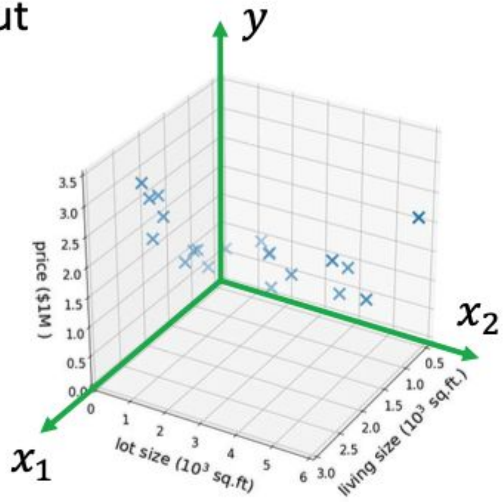
- Suppose we also know the lot size
- Task: find a function that maps

$$\underbrace{\text{(size, lot size)}}_{\substack{\text{features/input} \\ x \in \mathbb{R}^2}} \rightarrow \underbrace{\text{price}}_{\substack{\text{label/output} \\ y \in \mathbb{R}}}$$

- Dataset: $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})$

where $x^{(i)} = (x_1^{(i)}, x_2^{(i)})$

- "Supervision" refers to $y^{(1)}, \ldots, y^{(n)}$



Even more features? Number of rooms, year built, energy rating, ...

High-dimensional features

- $x \in \mathbb{R}^d$ for large d
- E.g.,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ \vdots \\ \vdots \\ x_d \end{bmatrix} \begin{array}{l} \text{--- living size} \\ \text{--- lot size} \\ \text{--- \# floors} \\ \text{--- condition} \\ \text{--- zip code} \\ \quad\quad\vdots \\ \\ \end{array} \longrightarrow \quad y \text{ --- price}$$

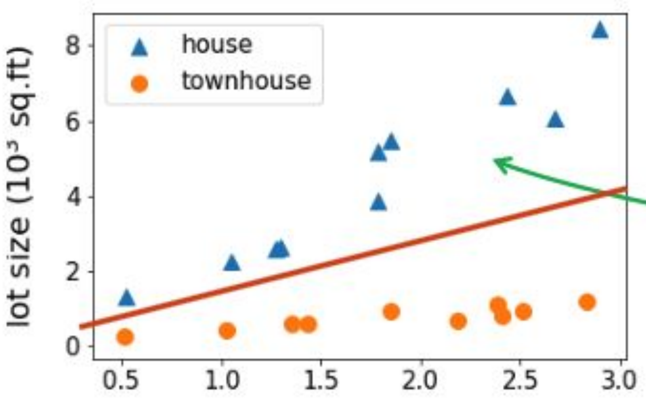
Regression vs Classification

Week 2: (linear) regression

Week 3: (linear) classification

- regression: if $y \in \mathbb{R}$ is a continuous variable
  - e.g., price prediction
- classification: the label is a discrete variable
  - e.g., the task of predicting the types of residence

(size, lot size) $\rightarrow$ house or townhouse?

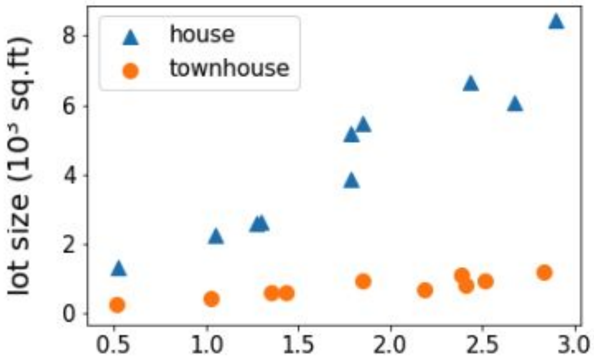


y = house or townhouse?

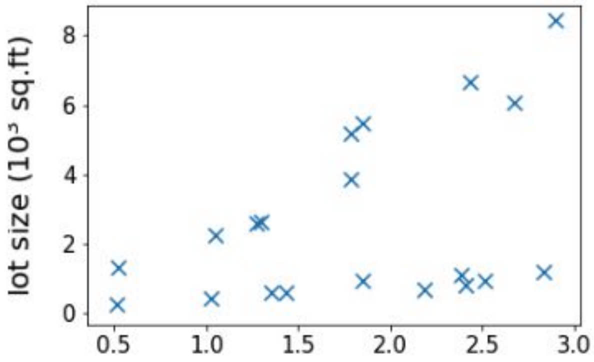
Unsupervised Learning

- Dataset contains no labels: $x^{(1)}, \ldots x^{(n)}$
- Goal (vaguely-posed): to find interesting structures in the data

supervised



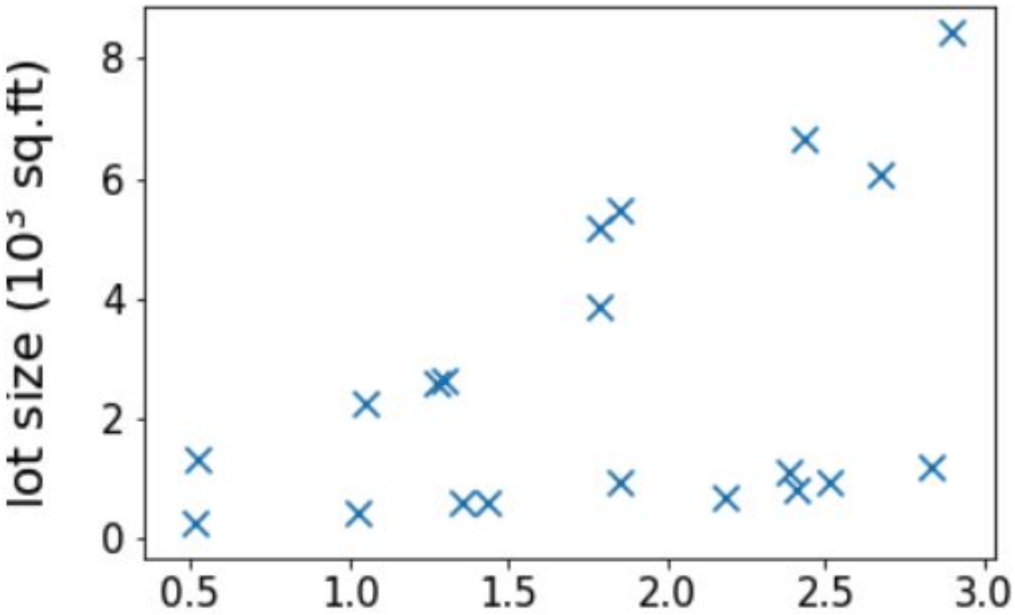
unsupervised



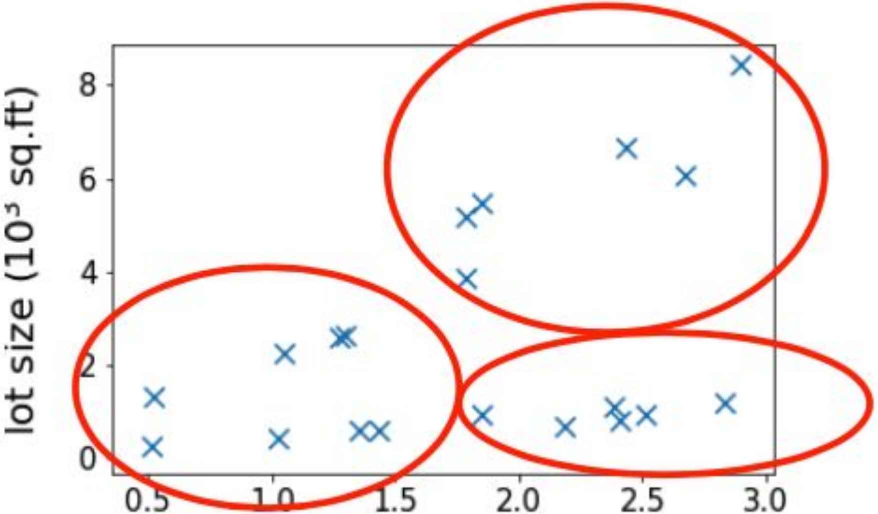
Clustering

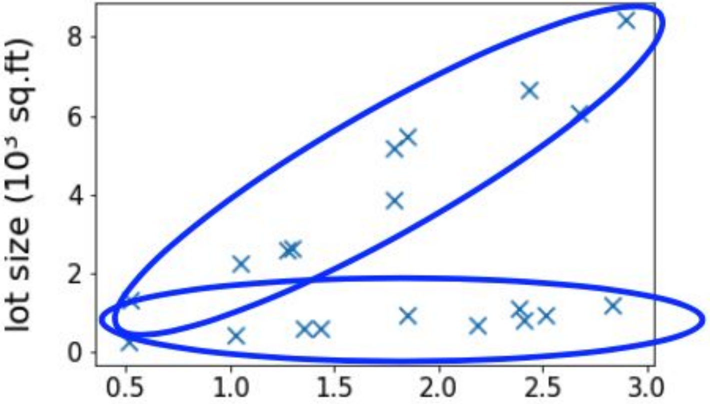
Week 5: Gaussian mixture models, k-means

Week 4: PCA



Clustering


lot size (10³ sq.ft)
0
2
4
6
8
0.5
1.0
1.5
2.0
2.5
3.0


lot size (10³ sq.ft)
0
2
4
6
8
0.5
1.0
1.5
2.0
2.5
3.0

Supervised Learning in Computer Vision

Covered in the computer vision course

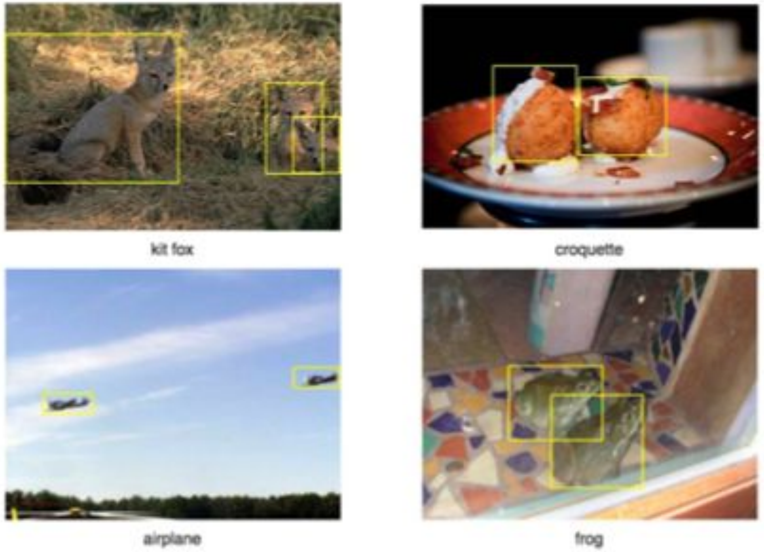
- Image Classification
  - x = raw pixels of the image, y = the main object

ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Supervised Learning in Computer Vision

Covered in the computer vision course

- Object localization and detection
  - x = raw pixels of the image, y = the bounding boxes

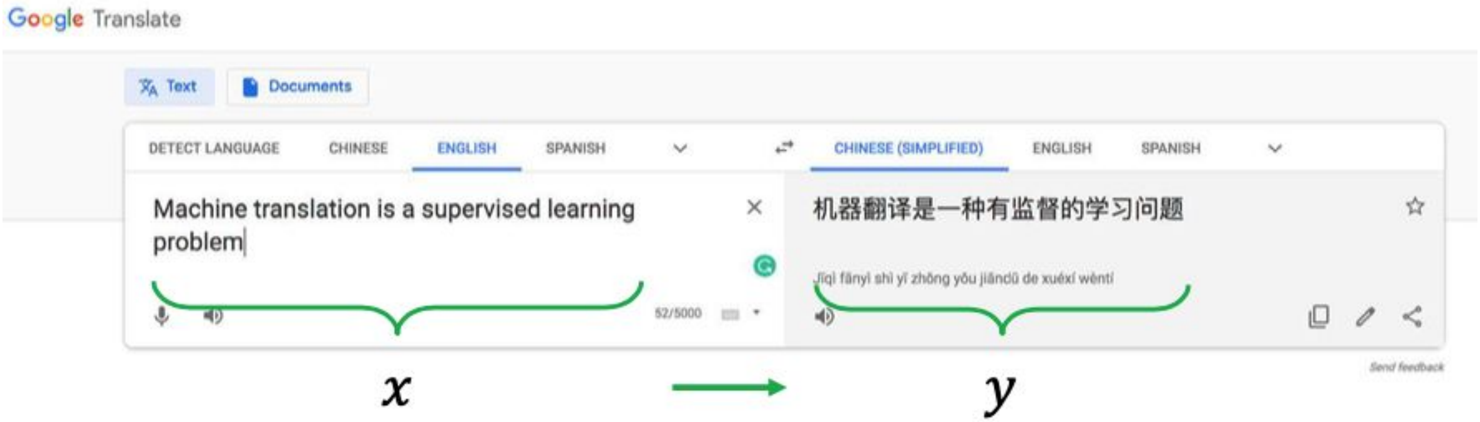


ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Machine Learning for NLP

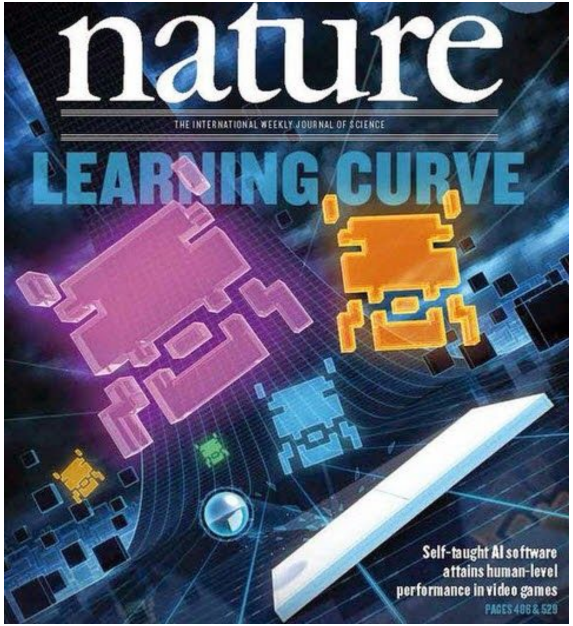
Take Document Analysis (COMP4650) if you'd like to learn more.

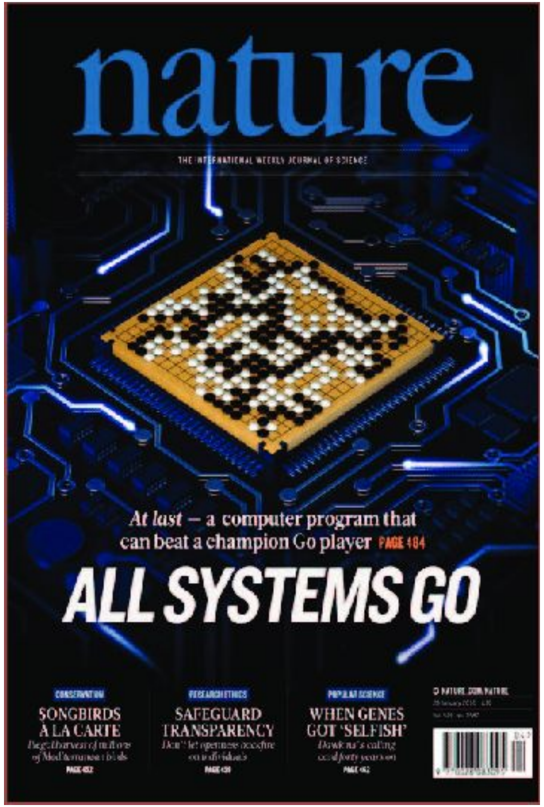
- Machine translation



- **Note:** this course only covers the basic and fundamental techniques of supervised learning (which are not enough for solving hard vision or NLP problems.)

Reinforcement Learning

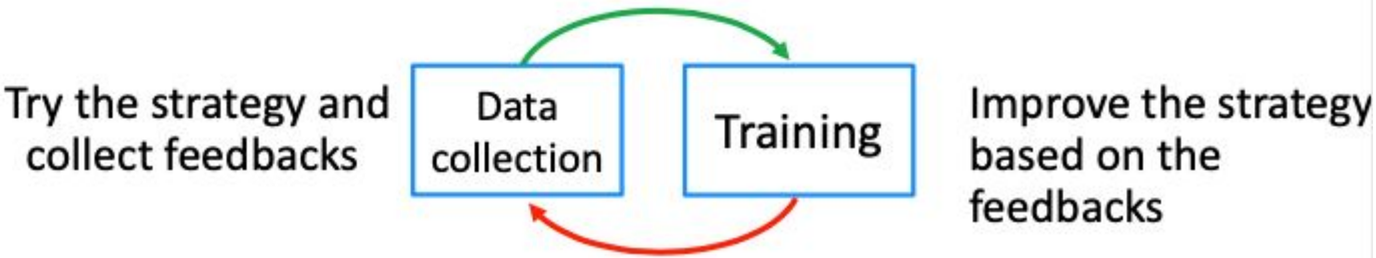




Reinforcement Learning

Take Advanced AI to learn more about POMDP and reinforcement learning.

- ➢ The algorithm can collect data interactively



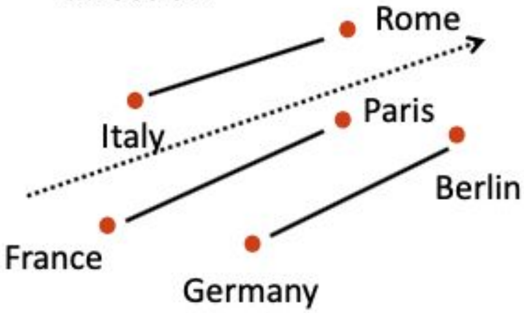
Many open challenges remain for tackling real-world problems.

Word “Embeddings”

WIKIPEDIA
The Free ~~Encyclopedia~~
CORPUS

Represent words by vectors

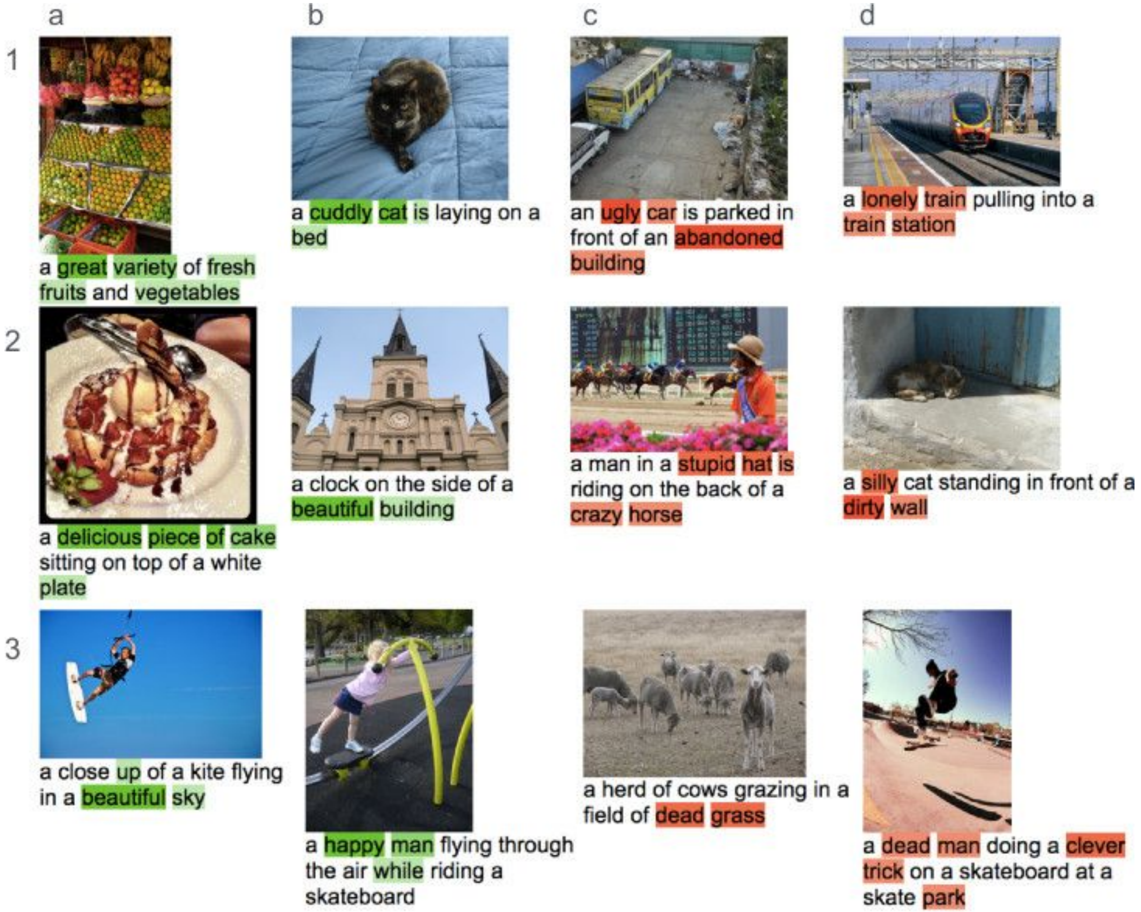
- word $\xrightarrow{\text{encode}}$ vector
- relation $\xrightarrow{\text{encode}}$ direction



“Unlabeled” data?
Self-supervision?

Word2vec [Mikolov et al’13]
GloVe [Pennington et al’14]

Senticap: Describing Images with Sentiments



http://cm.cecs.anu.edu.au/post/senticap/

SemStyle: Learning to Caption from Romantic Novels

success cases

failure case

(a)

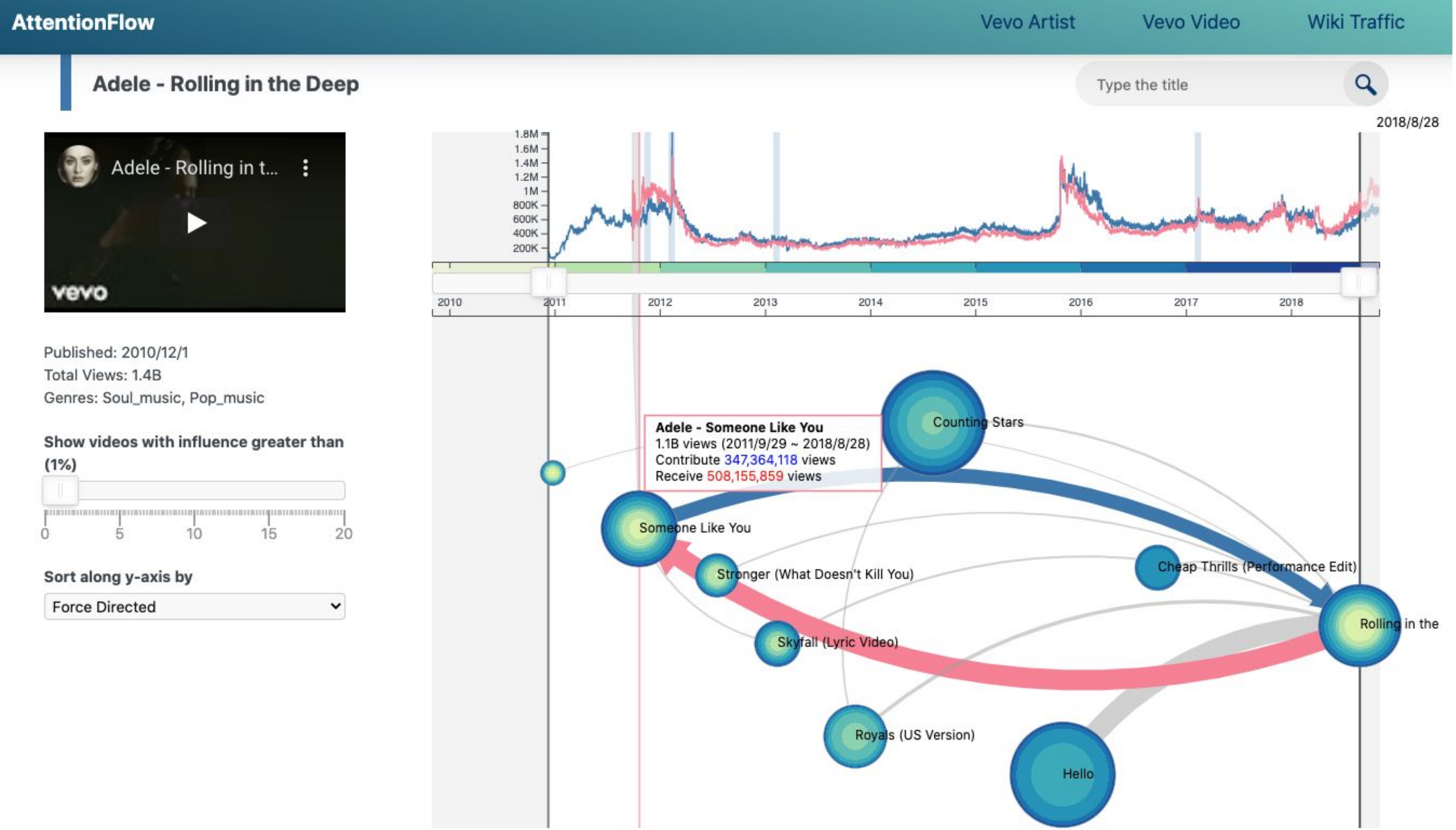
(b)

(c)

	(a)	(b)	(c)
[Descriptive]	A woman walking with an umbrella in the rain.	A forest that has a large tree in it.	A juicer is poured into a glass of juice.
[Story-like]	The woman stepped underneath her umbrella and walked in the rain.	Forest, tall, and thick trees.	I'll be in the juicer with a glass of orange juice.

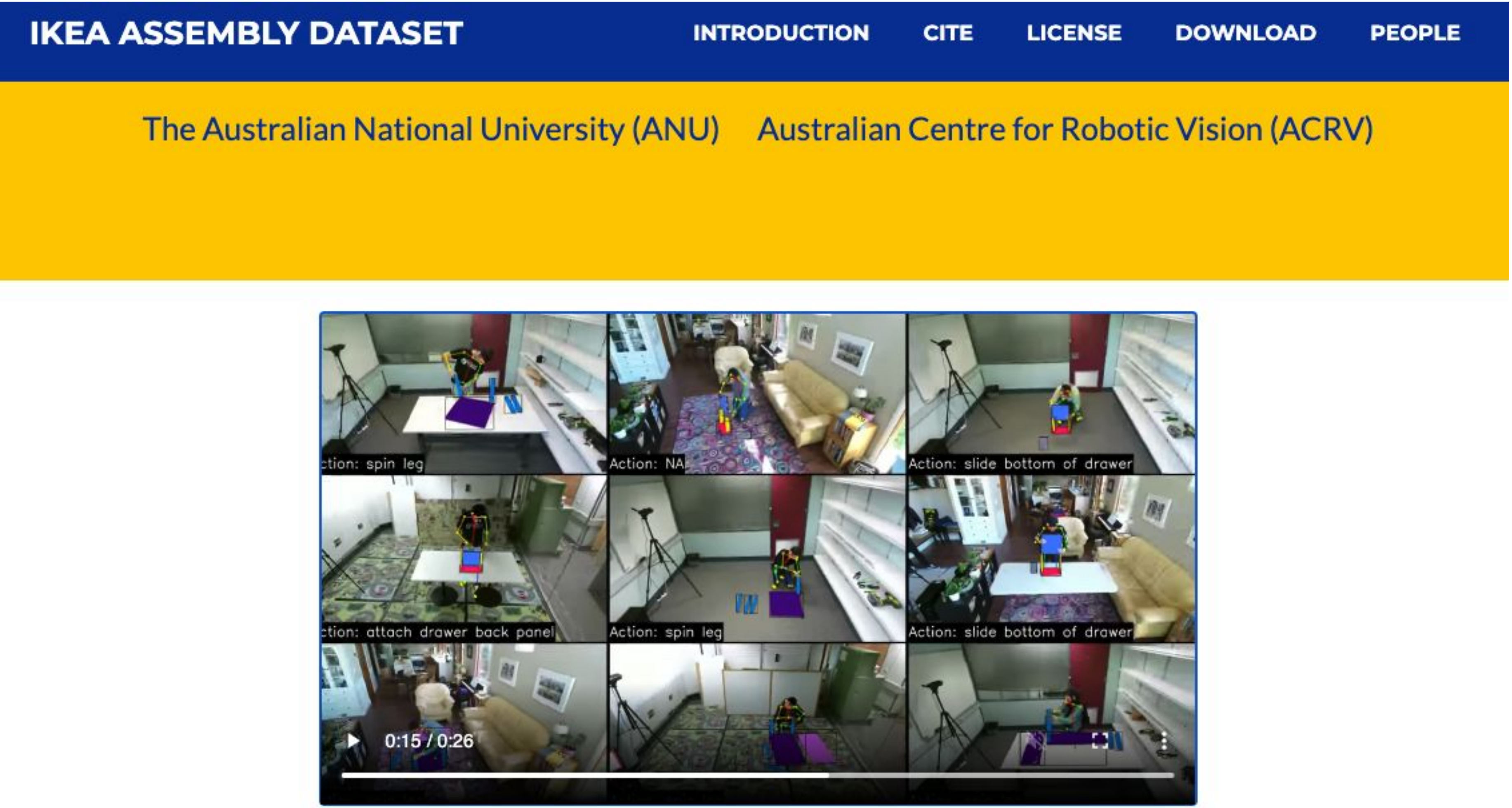
Descriptive (blue) and story-like (dark red) image captions created by the SemStyle system. The story-like captions in example (a) is written as a sequence of actions, rather than a static scene description; (b) introduces a new adjective and uses a poetic sentence structure. Styled caption (c) is my favorite failure case -- it violates common sense but triggers readers' imagination.

http://cm.cecs.anu.edu.au/post/semstyle/
http://cm.cecs.anu.edu.au/post/transform_and_tell/



https://attentionflow.ml/#/overview/video/rYEDA3JcQqw

Estimating Attention Flow in Online Video Networks, Wu et al, ACM CSCW 2019
Radflow: A Recurrent, Aggregated, and Decomposable Model for Networks of Time Series, Tran et al, The Web Conference 2021
AttentionFlow: Visualising Influence in Networks of Time Series, Shin, Tran, Wu, Lyall, Wang, Mathews, Xie, WSDM Demo, 2021



https://ikeaasm.github.io/

Sign language recognition and translation

ARC Discovery Project: Hondong Li, Xin Xu

This project aims to develop an automatic two-way machine-translation system between Auslan (Australian Sign Language) and English by researching and leveraging advanced computer vision and machine learning technology. The project expects to advance research in AI technology on topics including visual recognition, language processing and deep learning. This will boost Australia's national research capacity and global competitiveness. Expected outcomes of this project will help to break the communication barriers between the Deaf and hearing population. This should provide significant benefits to Deaf communities through enhanced communication and improved quality-of-life, leading to a fair, more inclusive and resilient Australian society.

(a) The verb "**Wish**" (top) and the adjective "**hungry**" (bottom) correspond to the same sign.

(b) The same sign represents different words "**Rice**" (top) and "**soup**" (bottom).

(c) Signers perform "**Scream**" with different hand positions and amplitude of hand movements.

Figure 2: Ambiguity and variations of Signing. (a, b) shows linguistic ambiguity in ASL. (c) shows signing variations of different signers.

Bayesian Joint Inversions for the Exploration of Earth Resources

Alistair Reid[1], **Simon O'Callaghan**[1], **Edwin V. Bonilla**[1], **Lachlan McCalman**[1], **Tim Rawling**[2] and **Fabio Ramos**[3]

1. NICTA, 2. University of Melbourne, 3. University of Sydney

{ alistair.reid, simon.ocallaghan, edwin.bonilla, lachlan.mccalman }@nicta.com.au

tim.rawling@unimelb.edu.au, f.ramos@acfr.usyd.edu.au

In a geological inversion problem, properties such as temperature, conductivity, density, magnetic susceptibility and permeability are inferred from related observations such as gravity, magnetics and seismic reflexion. ... In this paper we formulate geophysical inversion as a machine learning problem, and propose an approach based on Gaussian processes regression that naturally provides both a predictive distribution over the inverted quantities and a principled method to fuse different types of observations. We apply our method to a real dataset from South Australia containing gravity and drill-hole data with the goal of characterizing rock densities for geothermal target exploration, and also to simulated validation data involving gravity, drill-hole and magnetic observations.

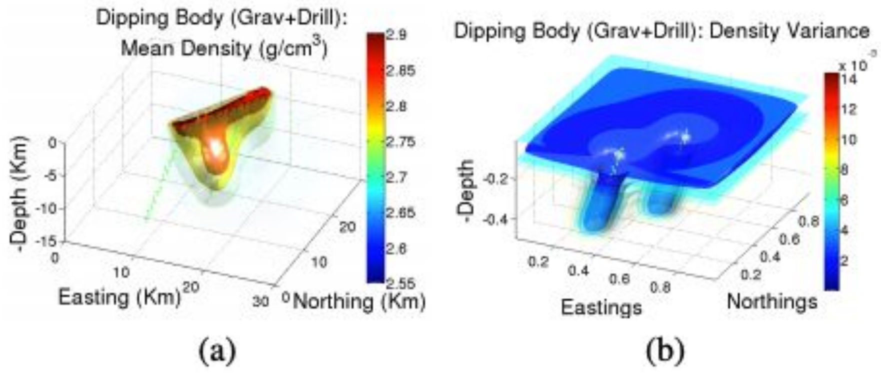


Figure 5: Outputs of the Gaussian process inversion algorithm after fusing gravity and drill observations of the dipping body.

ML + X at ANU

- Machine learning for quantum computing and gravitational physics
- Machine learning for astronomy
- Machine learning for synthesizing materials
- Machine learning for climate change, natural resources and disasters
- Machine learning for plants and agriculture
- Machine learning for monitoring insect behavior
- Active learning for experimental design in biology
- Machine learning for enhancing in-vitro microscopy
- … …

GPT-3, very large models, and all that

"an armchair in the shape of an avocado"

https://openai.com/blog/dall-e/

ML algorithms can appear biased and racist

https://archive.ieet.org/articles/sweeney.html

Google apologizes for mis-tagging photos of African Americans

BY AMANDA SCHUPAK
JULY 1, 2015 / 5:04 PM / CBS NEWS

Google was quick to respond over the weekend to a user after he tweeted that the new Google Photos app had mis-categorized a photo of him and his friend in an unfortunate and offensive way.

Jacky Alciné, a Brooklyn computer programmer of Haitian descent, tweeted a screenshot of Google's new Photos app showing that it had grouped pictures of him and a black female friend under the heading "Gorillas."

"Google Photos, y'all f****d up. My friend's not a gorilla," Alciné wrote.

WIRED BACKCHANNEL BUSINESS CULTURE GEAR IDEAS SCIENCE SECURITY SIGN IN SUBSCRIBE

When It Comes to Gorillas, Google Photos Remains Blind

Google promised a fix after its photo-categorization software labeled black people as gorillas in 2015. More than two years later, it hasn't found one.

Extracting Training Data from Large Language Models

Nicholas Carlini[1] Florian Tramèr[2] Eric Wallace[3] Matthew Jagielski[4]
Ariel Herbert-Voss[5,6] Katherine Lee[1] Adam Roberts[1] Tom Brown[5]
Dawn Song[3] Úlfar Erlingsson[7] Alina Oprea[4] Colin Raffel[1]
Google [2]*Stanford* [3]*UC Berkeley* [4]*Northeastern University* [5]*OpenAI* [6]*Harvard* [7]*Apple*

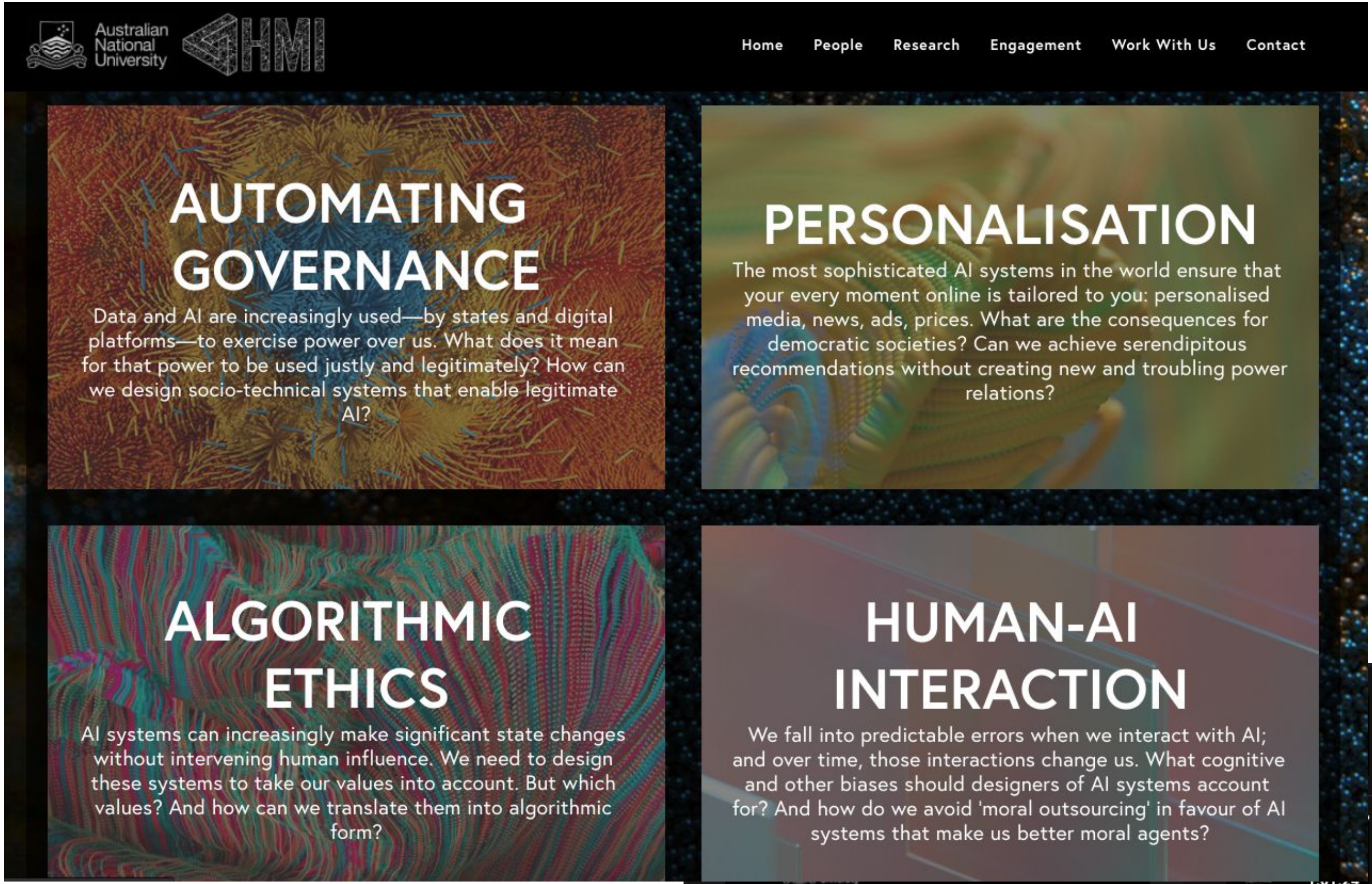
Memorized Usernames. There are BPE tokens for several usernames of individual people. For example, the Twitter handle for Donald Trump, `realDonaldTrump`, is represented by a single token in the encoding dictionary. However, this is not an instance of Eidetic memorization, as this token is contained in thousands of webpages. In contrast, through manual review of the BPEs, we identify three BPE tokens that correspond to usernames of individual users on Reddit.[13] These three tokens are otherwise unique on the Internet: Google searches yield 24, 29 and 34 results for each of these usernames; all results correspond to content related to these users.

Similarly, we identify one token that corresponds to the GitHub repository name of a particular user. This repository has only two "stars" on GitHub, and there are 40 results for this phrase contained on Google.

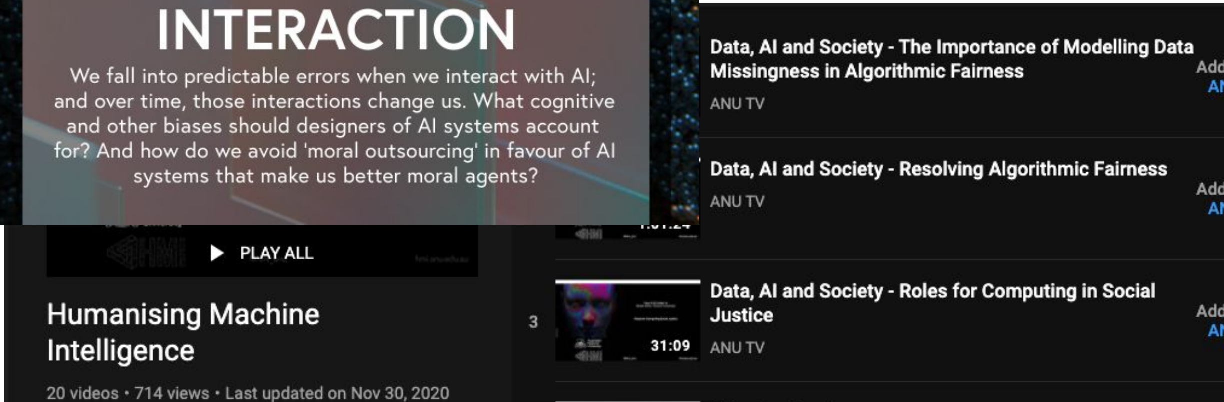
Memorized Leaked Podesta Emails from WikiLeaks. We identify several memorized URLs that originated from the leaked Podesta Emails available on WikiLeaks.[14] There is only a single training document that contains these memorized URLs. Due to the nature of email, the text of one message is often included in subsequent replies to this email. As a result, a URL that is used (intentionally) only once can be included in the dataset tens of times due to the replies.

[14] https://en.wikipedia.org/wiki/Podesta_emails

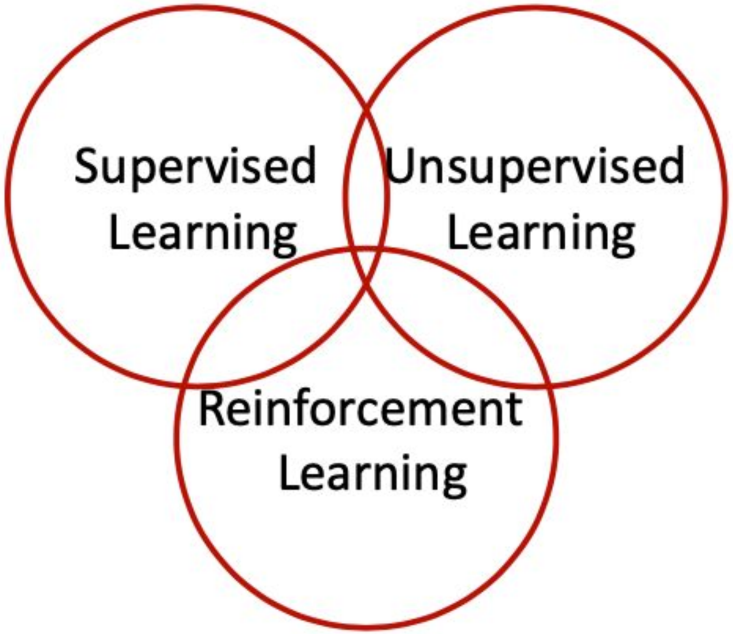
https://arxiv.org/abs/2012.07805



https://hmi.anu.edu.au/



Recall: simplistic machine learning taxonomy

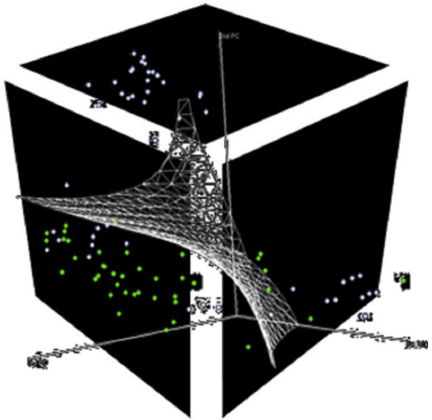


Example other machine learning tasks that this class won't cover

- Active Learning
  - The algorithm may choose which data $x_i \in \mathcal{X}$ to select next when building the model.
  - The order of the data is actively chosen by the algorithm at run-time.
- Transduction
  - The algorithms is allowed to use the test data (but of course not labels!) when building a model.
- Estimation with missing variables.
- Co-training with two different but related data sets.
- ... and others.

topic areas

learning theory, asymptotic analysis, classification, clustering, causality inference, game theory …



problem settings

supervised learning, unsupervised learning, semi-supervised learning, online learning, transfer learning, multitask learning, life-long learning, zero-shot learning, multi-instance learning, machine teaching, meta learning, active learning, reinforcement learning, structured prediction …

models

random forest, decision trees, C4.5, logistic regression, linear regression, ridge regression, lasso, naive bayes, kmeans, spectral clustering, t-SNE, bandit algorithms, topic models, support vector machines, kernel methods, matrix and tensor factorisation, MDP, neural network, ResNet, CovNet, RNN, LSTM, learning to rank, factorisation machines, CCA, ICA, point processes, CRF …

methods

optimization: combinatorial, convex, non-convex, submodular; belief propagation, variational inference, stochastic gradient descent, adam, density estimation, hyperparameter selection, distributed inference, graph cut …

problem domains

computational biology, finance, computer vision, surveillance, traffic monitoring, natural language processing, geo-physics and geo-chemistry, question-answering, information retrieval, crowd-sourcing, music information retrieval, prediction markets, computational social science, knowledge extraction, neural science, astronomy, …

What will SML NOT cover

- Reinforcement learning
- Models for time series and graphs
- Recommender systems
- Optimisation, convex or not
- ML systems, distributed/federated learning
-
- Make you an expert in TRENDY_DEEPLEARNING_PACKAGE
- ...

On the Way to Learning (in Indonesia)

A good time to learn ML

- Job prospects
- Intellectual reward
- Blend of CS and math
- In-dpeth understanding of inside the blackbox, so as to improve its use and social welfare

Tips (see tutor videos)

- Use different course resources
- Program it to “really understand” something
- Start early on the assessment items!

What to do now:

- Register on Piazza and GradeScope
- Register for a tutorial slot.
- Work on Assignment 0

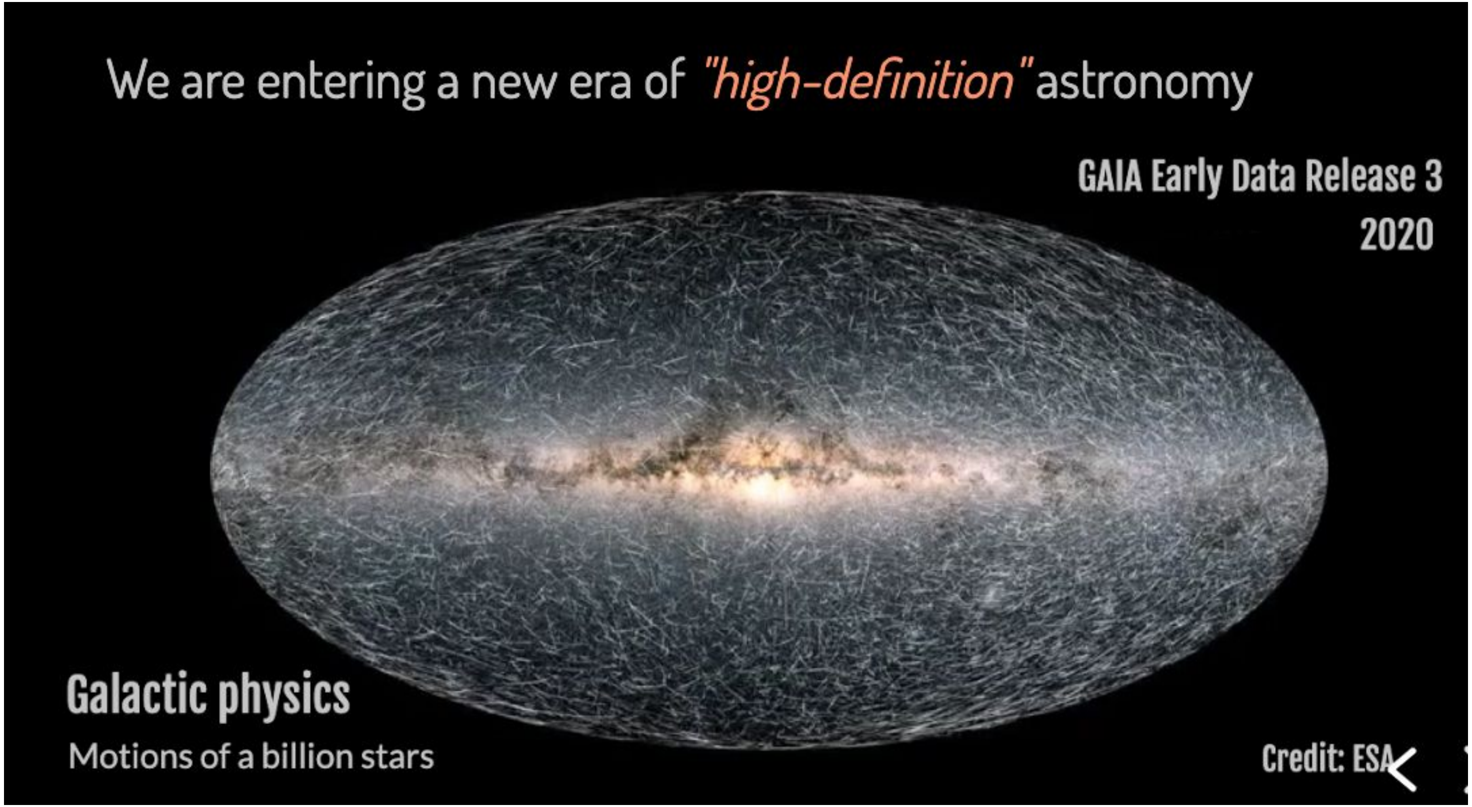
Questions?

Solar Power Prediction

- Photovoltaics now very close to grid electricity in price
- Distributed system of generators -- Energy market
- Great Machine Learning Problem: Predict the solar energy output (variability primarily due to clouds) for Australia
- Pilot project in Canberra : Use cheap cameras to take 360◦ sky photos in several location.
- Learn to predict 3-D model of cloud movement.
- Learn orientation and efficiency of solar panels for each house from time series of energy output.
- Predict output of each solar panel for 15 min to 1 hour from current snapshots.

Photo from wikipedia

ML for Astronomy



Slide by Yuan-Sen Ting http://sns.ias.edu/~ting/

TODO: A bit more detail w.r.t IML and other related courses

STAT3040 Statistical Learning

STAT3017 Big Data Statistics (random matrix theory, Dale Roberts)