Introduction to Machine Learning

Fundamentals of Machine Learning (G6061)

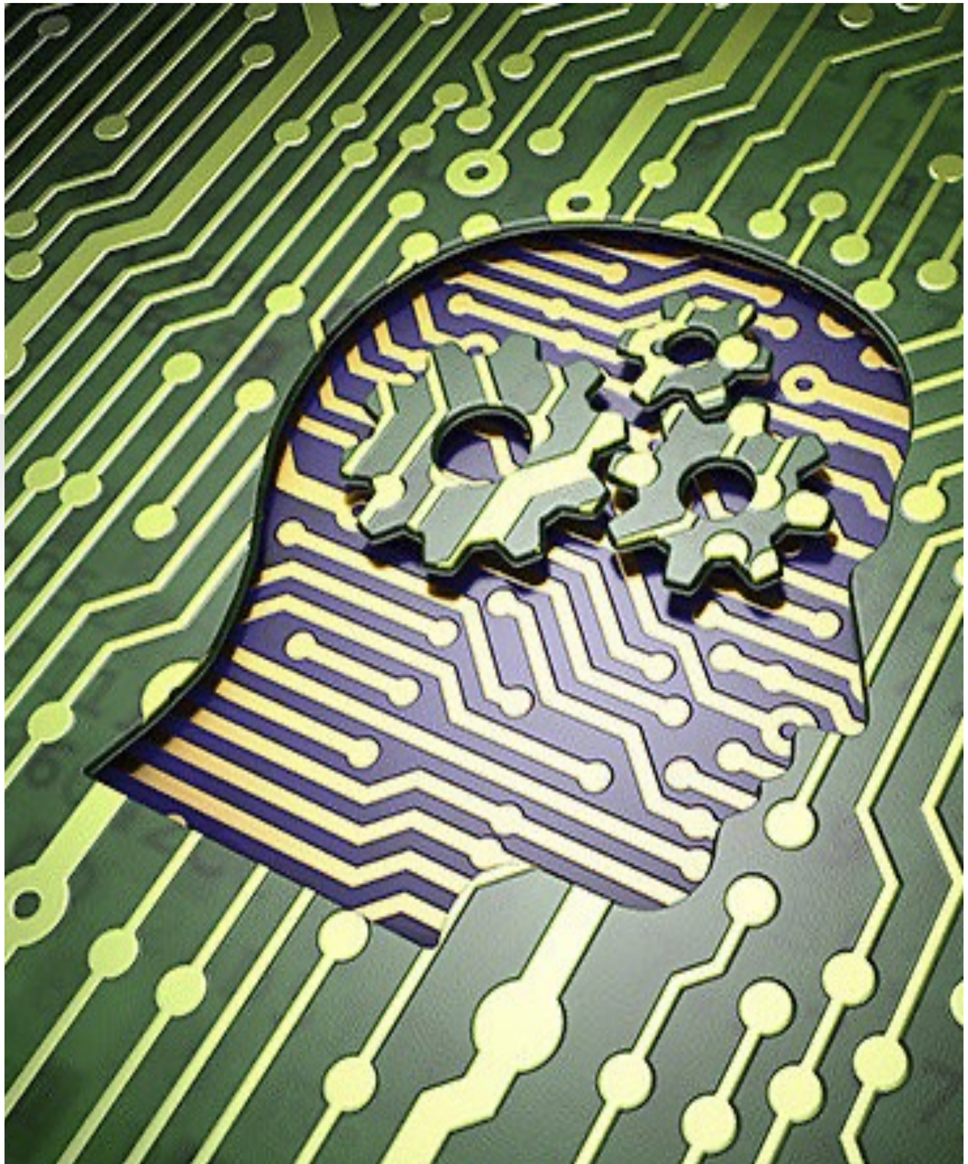
Dr. Benjamin Evans

University of Sussex

What is Machine Learning?

A subset of AI involving computer programmes, often referred to as models, which:

- are trained to achieve some task,
- **without** requiring programming explicit instructions for how to do it
  - rather they enable computers to perform pattern recognition.
- these patterns are learned from some example data.
- A key factor in ML is getting the computer to learn a generalisable mechanism for achieving the task.



Why is Machine Learning Interesting?

- Potential to delegate trivial, repetitive or dangerous jobs to machines.
- Faster processing of information
- Vast volumes of data are being generated and stored digitally, more than any human could ever hope to look at/learn from.
  - ► Potential for deeper insights to be learned from more data

55 million articles

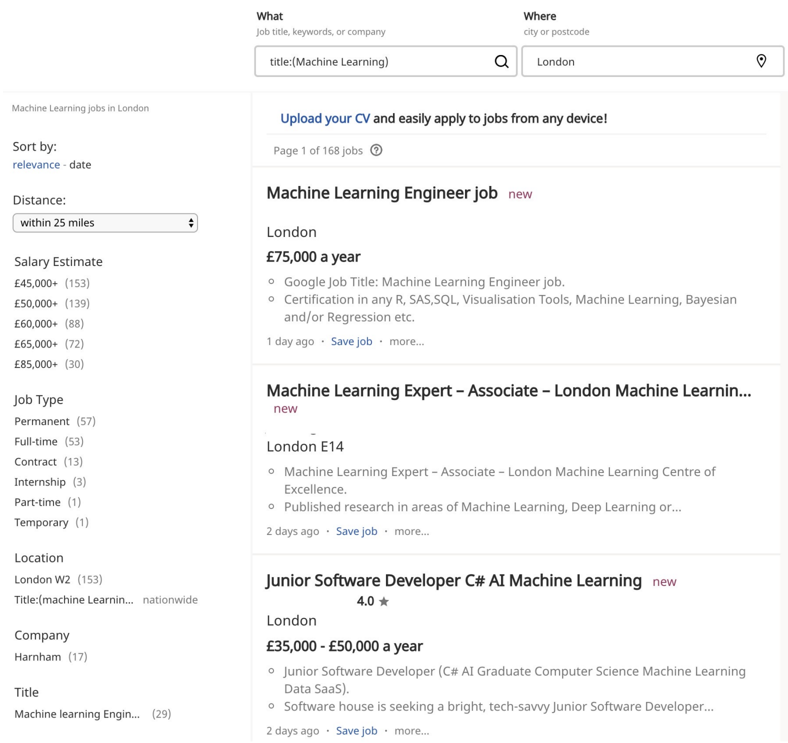
2.7 billion users

500 million tweets a day

4.2 billion webpages

Why Should Machine Learning be Interesting to You?

- Curiosity – machine learning is tasked with designing frameworks to <u>learn</u> and reason about data without needing to articulate explicit decision rules.
- It is becoming ubiquitous: you interact with a number of ML systems in day-to-day life.
- Combination of programming, applied mathematics and machine learning should fit well within your skillset at graduation.
- Job prospects are very good in this field at the moment!



University of Sussex

How do you interact with ML in your life?

Let’s find out what you know about ML and how it might be involved in your day-to-day life: PollEv.com/bdevans

This will help us to think about different possible tasks, and their requirements.

What applications of ML have you used?

Nobody has responded yet.

Hang tight! Responses are coming in.

Outline for today

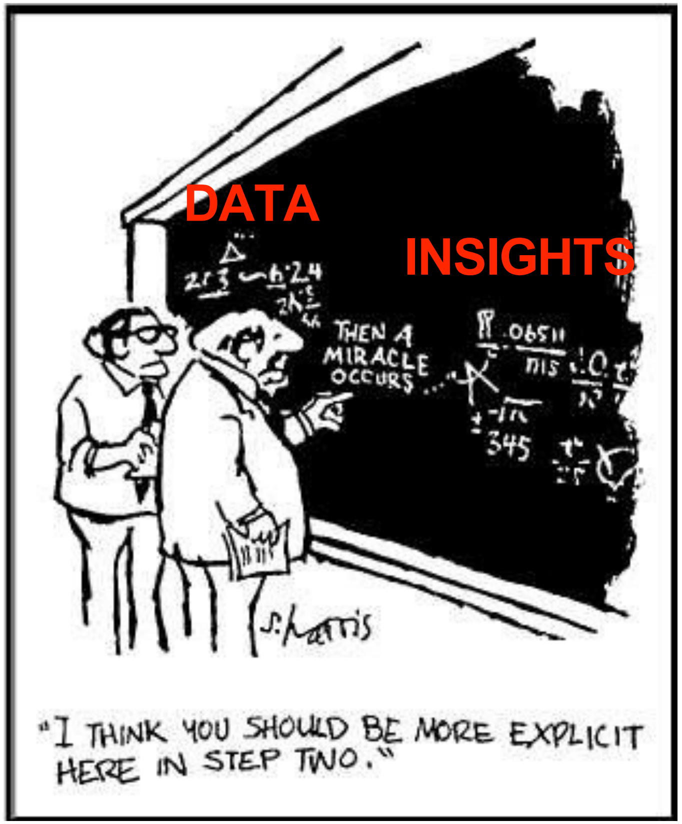
- About the module: logistics, evaluation, and syllabus overview
- The *What* and *Why* of machine learning
- Ingredients of machine learning: **tasks**, **features**, and **models**

Learning outcomes

- Understand the variety of machine learning tasks such as classification, regression, clustering, dimensionality reduction, and collaborative filtering.
- Learn what models are.
- Recognise the importance of features as an input to machine learning.

About This Module

- This is an introduction course to the field of machine learning.
- The module aims to provide a unified view of the field and to present key ideas and techniques.
- We will take a fairly applied, example-driven approach, but will also try to cover important theoretical concepts.



University of Sussex

Module Learning Outcomes

By the end of this module a successful student should be able to:

- Demonstrate basic knowledge of several **supervised** and **unsupervised** machine learning **models** including: *multi-layer perceptron*, *support vector machine*, *random forest*, *K-means*, and *PCA*.
- Map machine learning models to tasks based on reasoned arguments.
- Explain and exploit practical concepts such as **cross-validation** and **learning curves**.
- Use machine learning toolboxes to solve **classification**/**regression** problems with real-world data, including **pre-processing** of the data and incorporating prior knowledge.

Logistics

- Canvas webpage:
  - ► All teaching materials (lecture notes, lab exercises & solutions, assessment details, additional reading, etc.) will be made available at **https://canvas.sussex.ac.uk/courses/28155**.
  - ► **Module Discussion Forum** on Canvas: If you have a question, it is likely someone else is wondering the same thing, so please use this forum to ask any questions and share resources for everyone's benefit.
- **Lectures** covering key concepts and techniques will be in person:
  - ► **Tuesdays 9-10am**.
  - ► **Thursdays 9-10am**.
  - ► Lectures may be supplemented by pre-recorded segments on Canvas.
- **Office hours**:
  - ► Dr. Johanna Senk: Wednesdays: 11-12pm, Thursdays: 11-12pm.
  - ► Dr. Ben Evans: Tuesdays: 12-1pm, Thursdays: 3-4pm.
  - ► Please send an email first to make sure we'll be available.

US
University of Sussex

Labs

- Support the lecture content.
  - ▶ **Start in Week 2**.
  - ▶ They will cover the **previous week's material**.
  - ▶ Take longer than the 1-hour slot.
  - ▶ Prepare in advance.
  - ▶ Attempt the extensions if you can.
- Labs will take place in person with a manual attendance register
- Check Sussex Direct for which of the six slots you have been assigned to (Tuesdays, Wednesdays x2 or Fridays x3).
- Solutions will be made available for the labs in the following week and walkthroughs will be posted on Canvas.

Evaluation

1. Multiple choice exam (on Canvas), 10% weighting, 1 hour, **Friday Week 5 from 17.00**.
2. Multiple choice exam (on Canvas), 10% weighting, 1 hour, **Friday Week 9 from 17.00**.
3. Data Analysis Competition + **Report**
   - This is worth 80% of the total grade.
   - A 1200-word report detailing the implementation, critical evaluation, and **creative** solution of a machine learning application.
   - Due for submission: **4pm, Thursday 23rd May**.

NO EXAM!!! 🥳

University of Sussex

Source Materials

Some of the good options:

- C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2007.
- S. J. D. Prince. Computer Vision: Models, Learning, and Inference. 2012. http://www.computervisionmodels.com/.
- S. J. D. Prince. Understanding Deep Learning. 2023. https://udlbook.github.io/udlbook/.
- K. Murphy. Probabilistic Machine Learning: An introduction. 2021. https://probml.github.io/pml-book/
- MP Deisenroth, AA Faisal, CS Ong, Mathematics for Machine Learning 2020. Online version available at: https://mml-book.github.io/book/mml-book.pdf

Useful video materials:

- https://www.coursera.org/course/ml and http://videolectures.net → search for "machine learning summer school".

Mathematics and Machine Learning

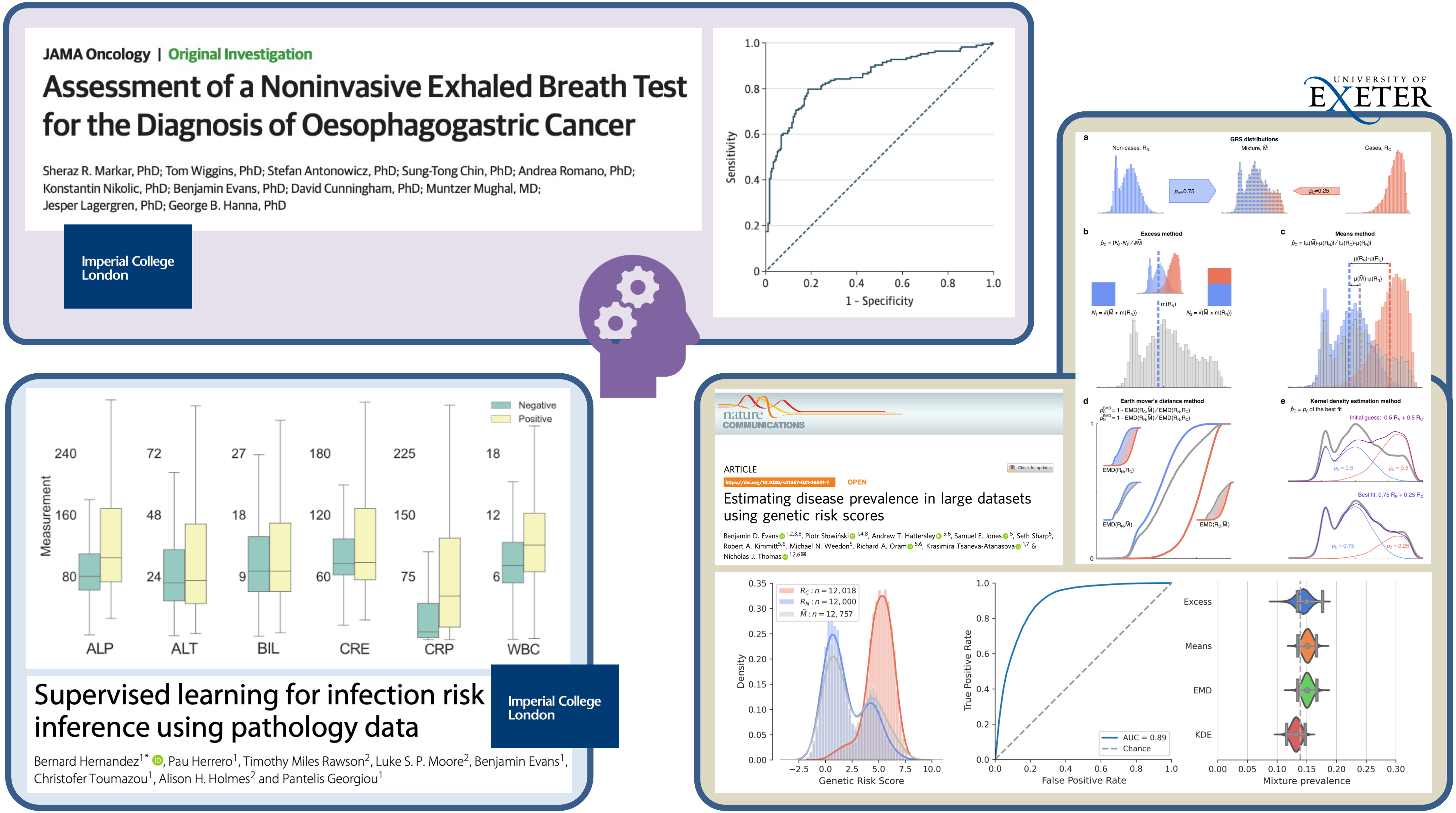
- Machine learning is underpinned by a variety of mathematical theories and formalisms.
- Difficult to present without at least some of the relevant mathematics:
  - probability theory
  - linear algebra
  - multivariable calculus
  - information theory
  - logic and set theory
- It **is** important to try to understand the intuition behind the mathematics. This influences the choice of what method to use in which situation.

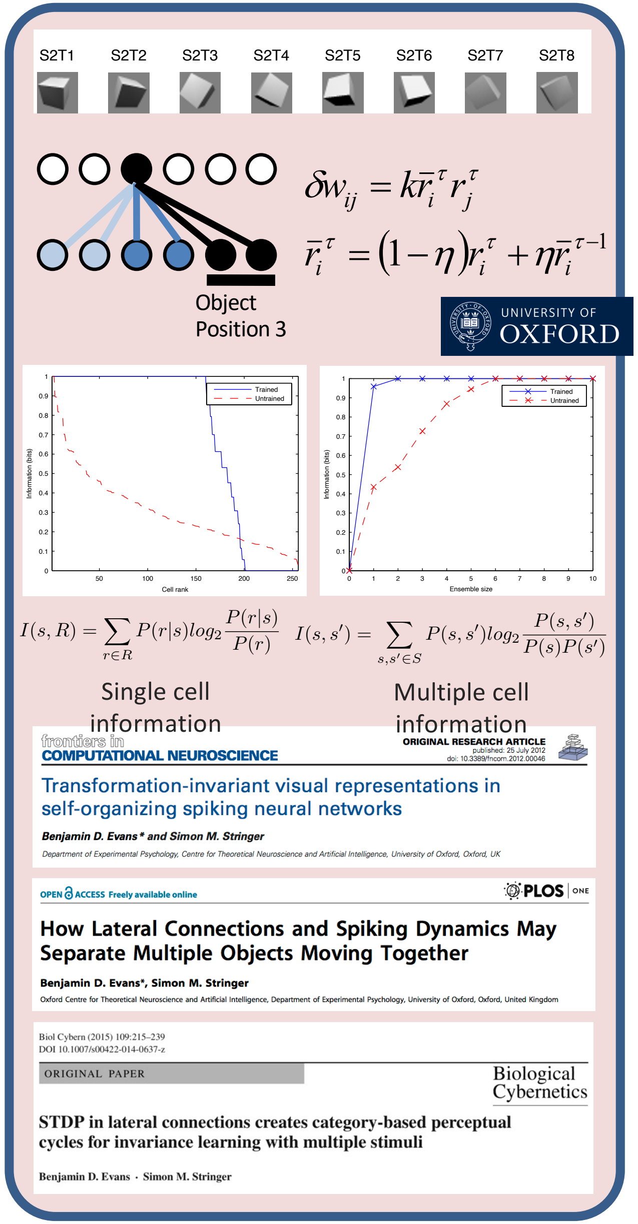
Syllabus Overview

1. Introduction to Machine Learning / Classifiers (BE)
2. Probability theory form ML (JS)
3. Linear Models for Regression and Classification (JS)
4. Regularisation and Model Selection (JS)
5. Perceptron and Backpropagation, Clustering (BE)
6. Catch-up week (BE)
7. Pre-processing / Dimensionality Reduction (JS)
8. Multilayer Perceptron (JS)

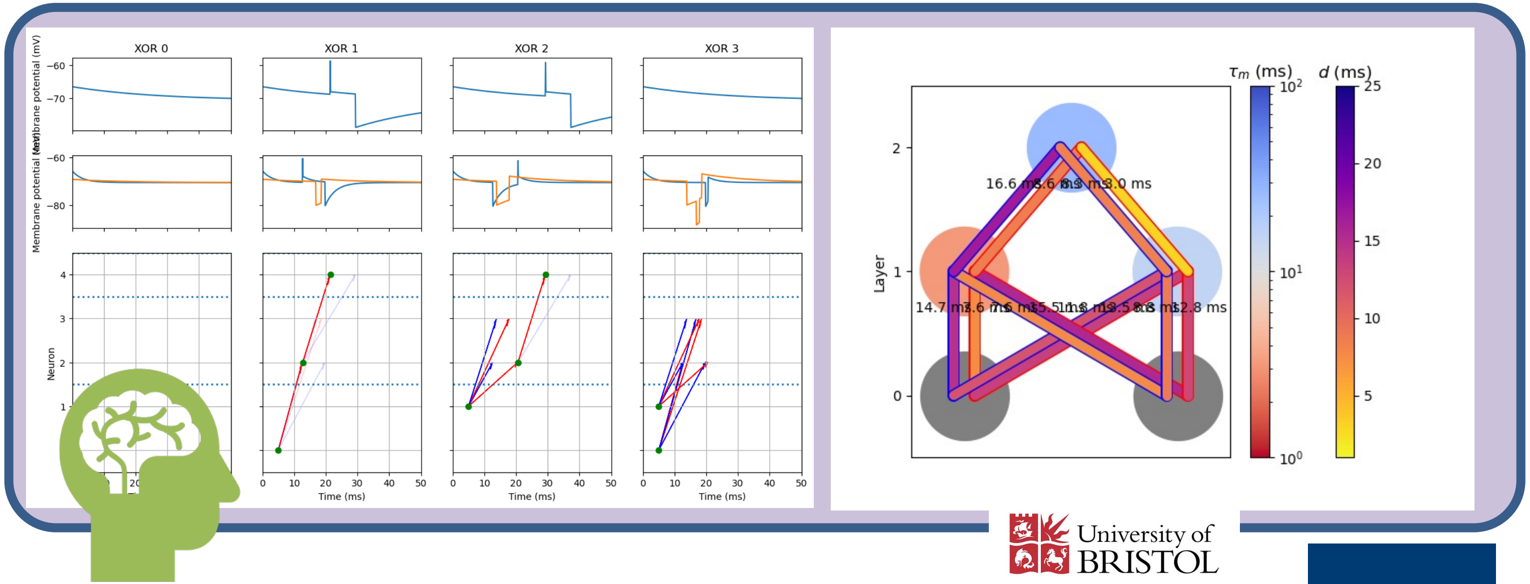
   ~~~ Easter Break ~~~

9. CNNs and Deep Learning (JS)
10. Support Vector Machines (JS)
11. Random Forests and Advanced Models (BE)
   ~~~

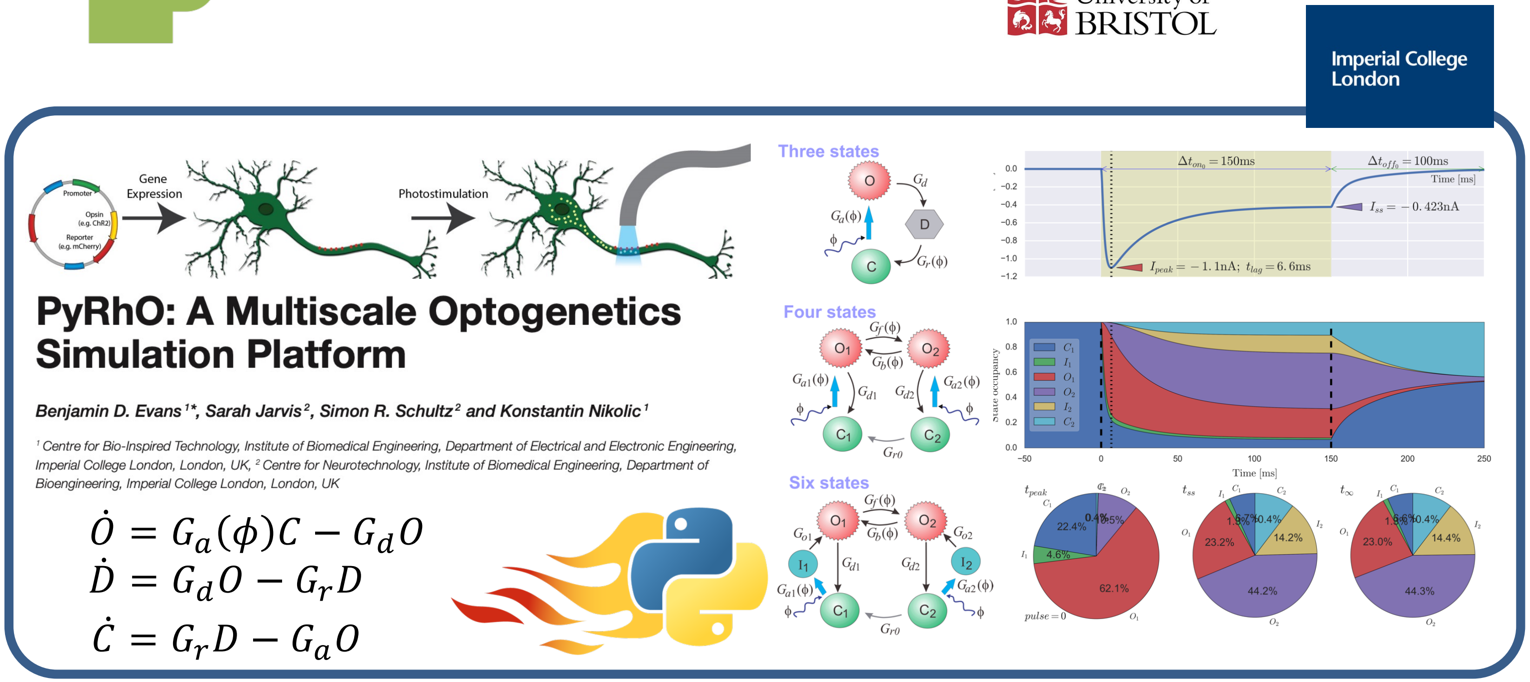

JAMA Oncology | Original Investigation
Assessment of a Noninvasive Exhaled Breath Test for the Diagnosis of Oesophagogastric Cancer
Sheraz R. Markar, PhD; Tom Wiggins, PhD; Stefan Antonowicz, PhD; Sung-Tong Chin, PhD; Andrea Romano, PhD; Konstantin Nikolic, PhD; Benjamin Evans, PhD; David Cunningham, PhD; Muntzer Mughal, MD; Jesper Lagergren, PhD; George B. Hanna, PhD
Imperial College London
Sensitivity
1 - Specificity
UNIVERSITY OF EXETER
Negative
Positive
Measurement
ALP
ALT
BIL
CRE
CRP
WBC
Supervised learning for infection risk inference using pathology data
Imperial College London
Bernard Hernandez, Pau Herrero, Timothy Miles Rawson, Luke S. P. Moore, Benjamin Evans, Christofer Toumazou, Alison H. Holmes and Pantelis Georgiou
nature COMMUNICATIONS
ARTICLE
OPEN
Estimating disease prevalence in large datasets using genetic risk scores
Benjamin D. Evans, Piotr Słowiński, Andrew T. Hattersley, Samuel E. Jones, Seth Sharp, Robert A. Kimmitt, Michael N. Weedon, Richard A. Oram, Krasimira Tsaneva-Atanasova & Nicholas J. Thomas
GRS distributions
Excess method
Means method
Earth mover's distance method
Kernel density estimation method
Density
Genetic Risk Score
True Positive Rate
False Positive Rate
AUC = 0.89
Chance
Excess
Means
EMD
KDE
Mixture prevalence



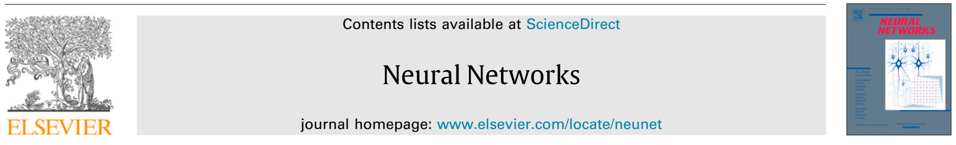
S2T1 S2T2 S2T3 S2T4 S2T5 S2T6 S2T7 S2T8
$\delta w_{ij} = k\bar{r}_i^{\tau} r_j^{\tau}$
$\bar{r}_i^{\tau} = (1-\eta) r_i^{\tau} + \eta \bar{r}_i^{\tau-1}$
Object
Position 3
UNIVERSITY OF OXFORD
Trained
Untrained
Cell rank
Ensemble size
$I(s,R) = \sum_{r \in R} P(r|s) log_2 \frac{P(r|s)}{P(r)}$
$I(s,s') = \sum_{s,s' \in S} P(s,s') log_2 \frac{P(s,s')}{P(s)P(s')}$
Single cell information
Multiple cell information
frontiers in COMPUTATIONAL NEUROSCIENCE
ORIGINAL RESEARCH ARTICLE
Transformation-invariant visual representations in self-organizing spiking neural networks
Benjamin D. Evans* and Simon M. Stringer
Department of Experimental Psychology, Centre for Theoretical Neuroscience and Artificial Intelligence, University of Oxford, Oxford, UK
OPEN ACCESS Freely available online
PLOS ONE
How Lateral Connections and Spiking Dynamics May Separate Multiple Objects Moving Together
Benjamin D. Evans*, Simon M. Stringer
Oxford Centre for Theoretical Neuroscience and Artificial Intelligence, Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom
Biol Cybern (2015) 109:215–239
DOI 10.1007/s00422-014-0637-z
ORIGINAL PAPER
Biological Cybernetics
STDP in lateral connections creates category-based perceptual cycles for invariance learning with multiple stimuli
Benjamin D. Evans · Simon M. Stringer



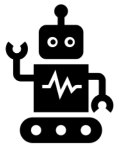
XOR 0
XOR 1
XOR 2
XOR 3
Membrane potential (mV)
Neuron
Time (ms)
Layer
τ_m (ms)
d (ms)
University of BRISTOL

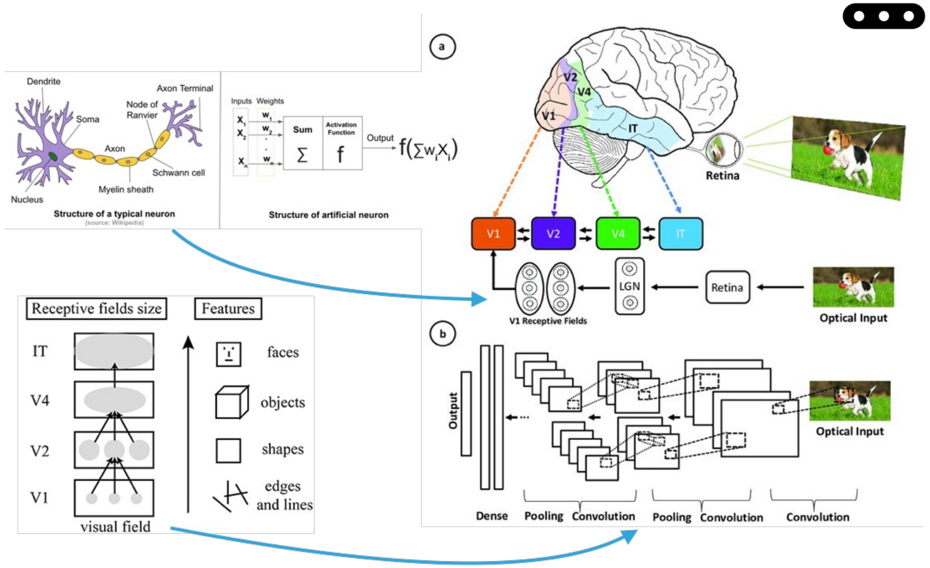


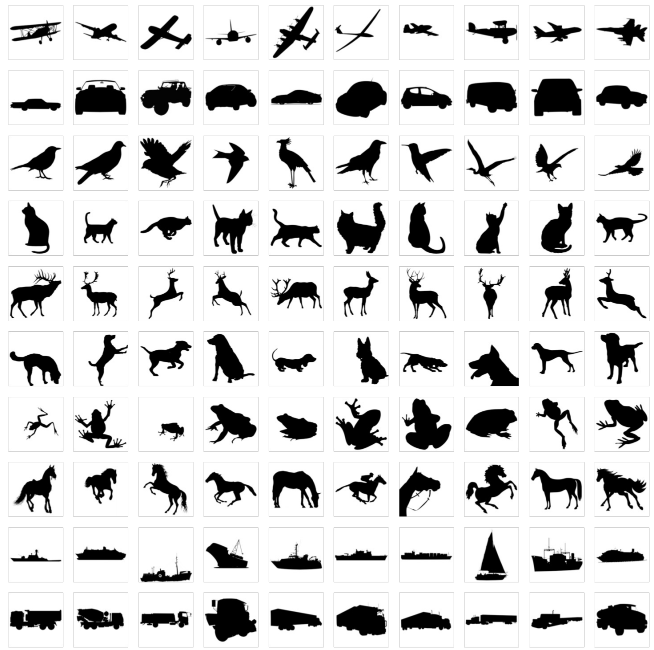
Imperial College London
Gene Expression
Photostimulation
PyRhO: A Multiscale Optogenetics Simulation Platform
Benjamin D. Evans [1]*, Sarah Jarvis [2], Simon R. Schultz [2] and Konstantin Nikolic [1]
[1] Centre for Bio-Inspired Technology, Institute of Biomedical Engineering, Department of Electrical and Electronic Engineering, Imperial College London, London, UK, [2] Centre for Neurotechnology, Institute of Biomedical Engineering, Department of Bioengineering, Imperial College London, London, UK
$\dot{O} = G_a(\phi)C - G_d O$
$\dot{D} = G_d O - G_r D$
$\dot{C} = G_r D - G_a O$
Three states
Four states
Six states
$\Delta t_{on_0} = 150$ms
$\Delta t_{off_0} = 100$ms
$I_{ss} = -0.423$nA
$I_{peak} = -1.1$nA; $t_{lag} = 6.6$ms
Time [ms]
State occupancy

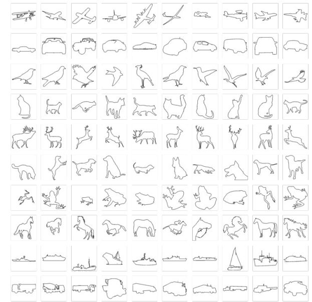

Neural Networks 148 (2022) 96–110
Contents lists available at ScienceDirect
Neural Networks
journal homepage: www.elsevier.com/locate/neunet
2021 Special Issue on AI and Brain Science
Biological convolutions improve DNN robustness to noise and generalisation
Benjamin D. Evans *, Gaurav Malhotra, Jeffrey S. Bowers
School of Psychological Science, University of Bristol, 12a Priory Road, Bristol BS8 1TU, UK

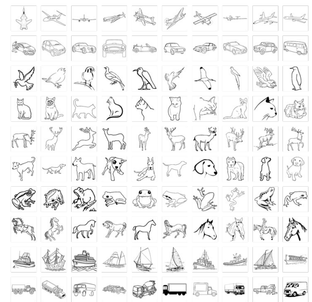
Architecture

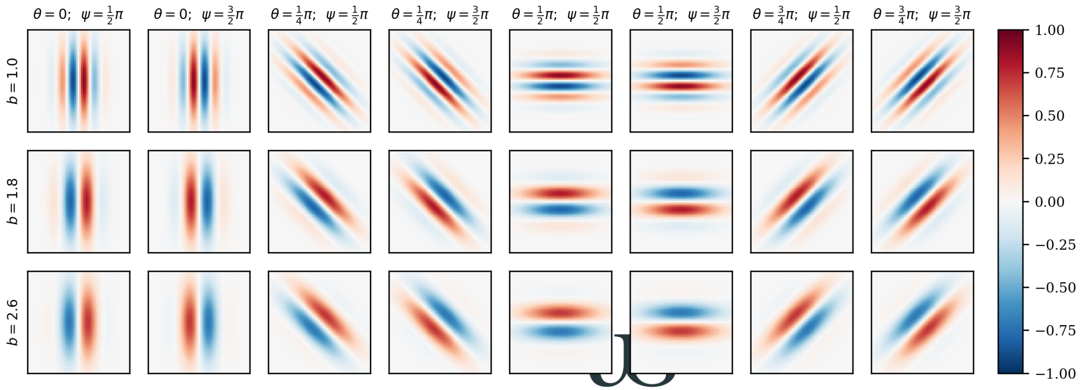










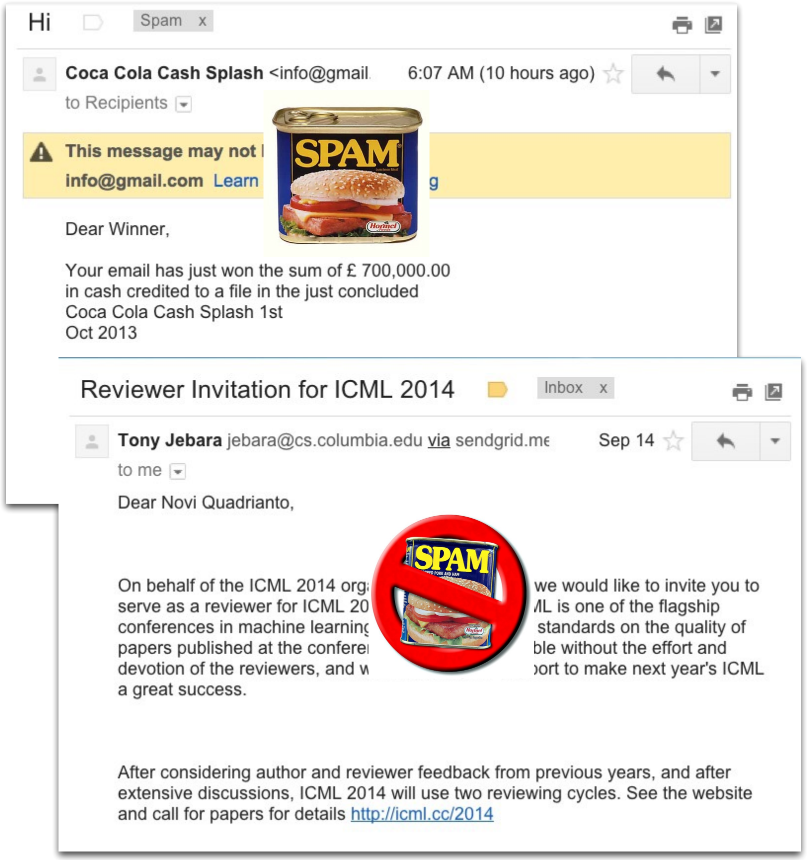


University of Sussex

Why Machine Learning?

Data can be used to perform all kinds of tasks:

1. Identifying whether an email contains irrelevant information (spam) or not (not spam).

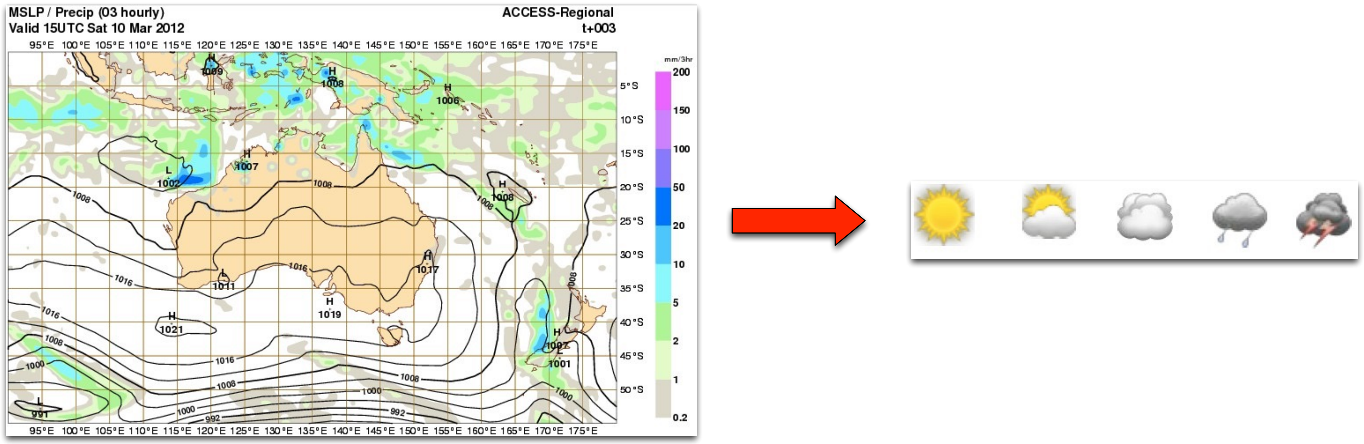


University of Sussex

Why Machine Learning?

Data can be used to perform all kinds of tasks:

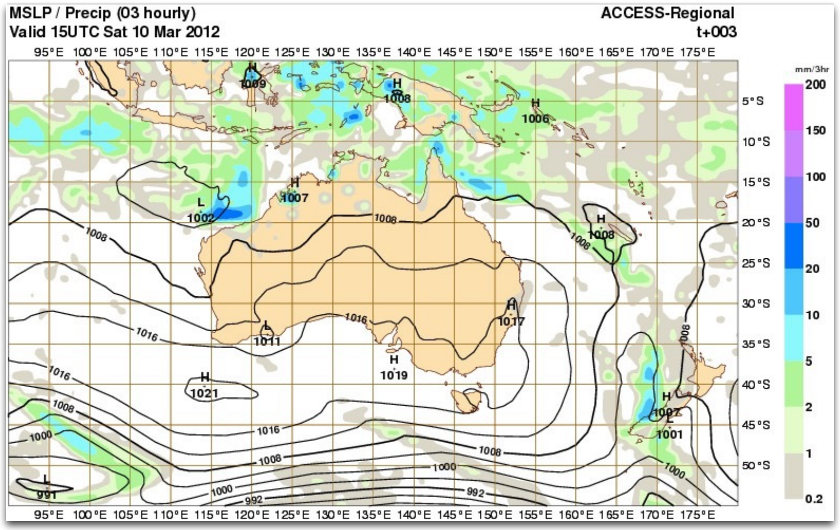
2 Predicting from satellite images and environmental sensors whether tomorrow will be sunny or partly cloudy or cloudy or rainy or thunderstorm.

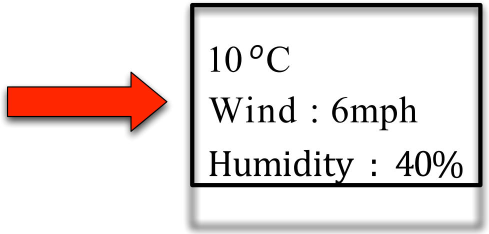


Why Machine Learning?

Data can be used to perform all kinds of tasks:

3 Predicting from satellite images and environmental sensors for tomorrow's temperature, wind, and humidity values.



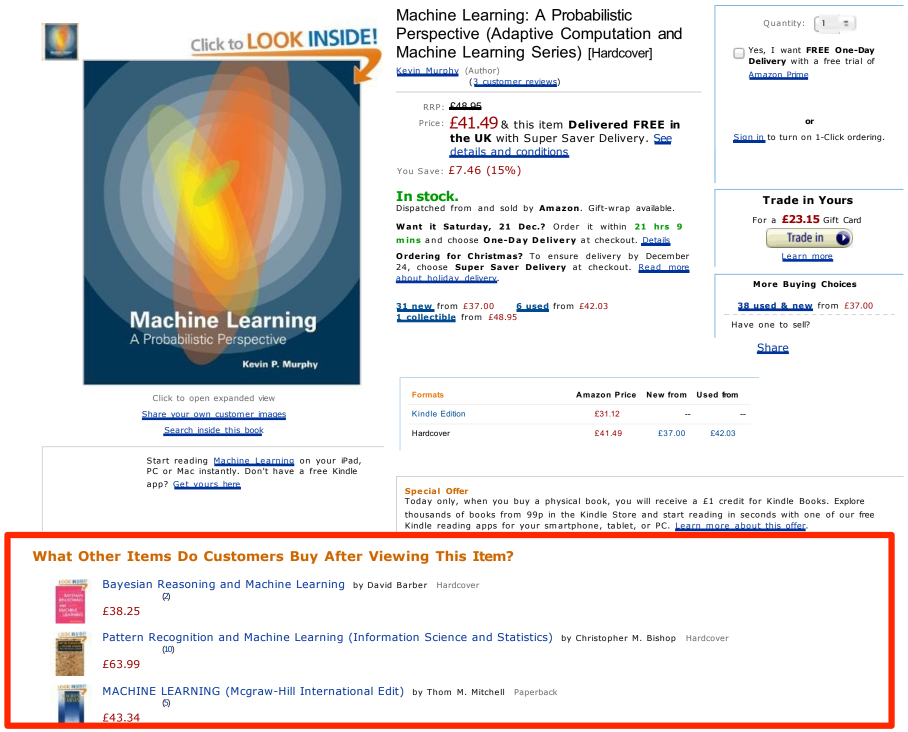


University of Sussex

Why Machine Learning?

Data can be used to perform all kinds of tasks:

- Recommending products and services on the basis of customer purchase history and the decisions made by similar customers.

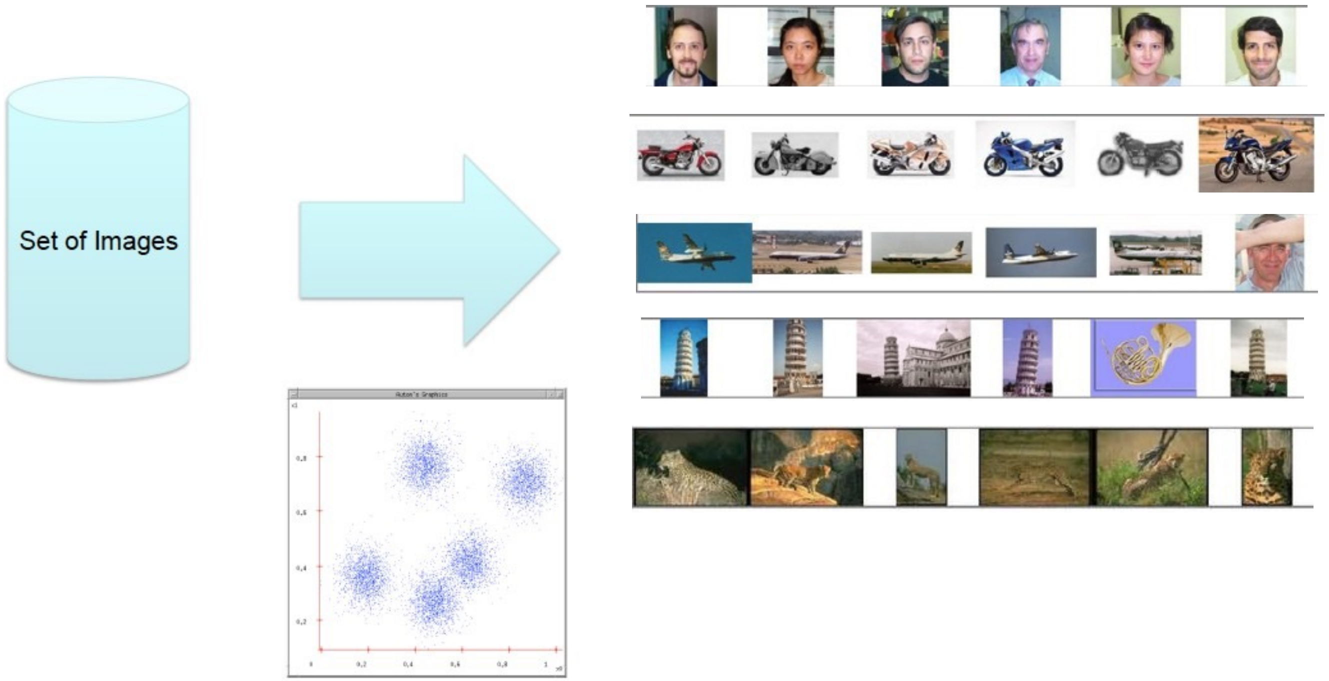


University of Sussex

Why Machine Learning?

Data can be used to perform all kinds of tasks:

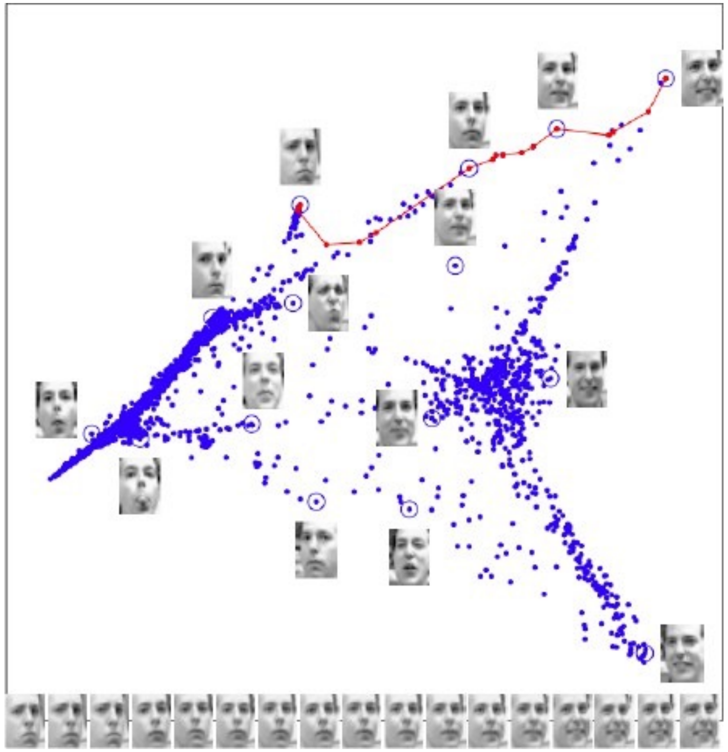
5 Grouping together images according to semantic similarities.



Why Machine Learning?

Data can be used to perform all kinds of tasks:

6 Visualising high dimensional data in a 2-dimensional space.



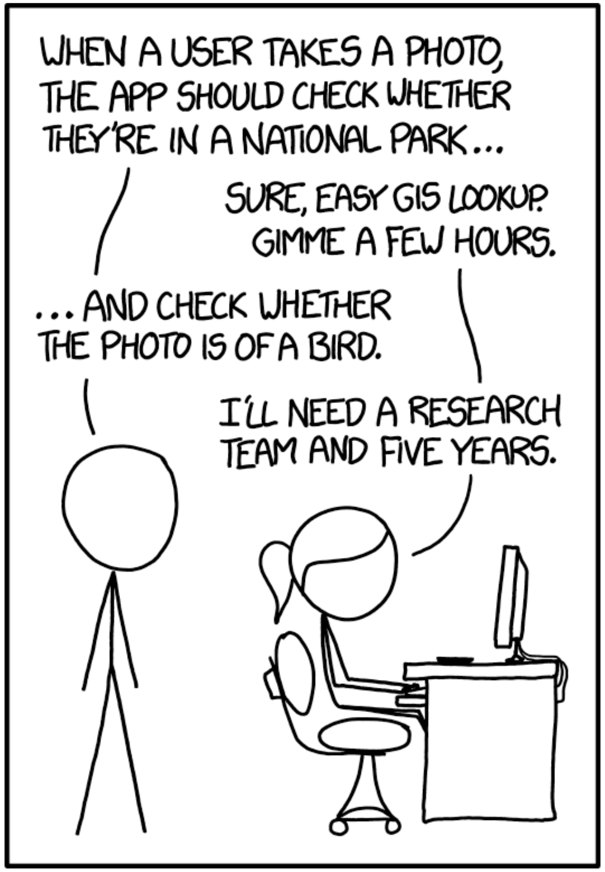
Saul & Roweis, 2003.

University of Sussex

What problems can ML solve that were previously very difficult or impossible?

Nobody has responded yet.

Hang tight! Responses are coming in.



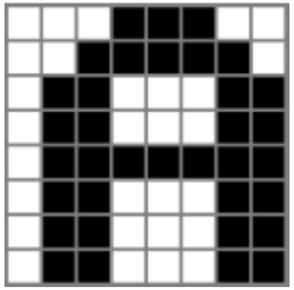
IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

Why Machine Learning?

Some problems are difficult or even impossible to formalise as a computer science problem, but humans can provide examples or feedbacks.

- Given an image of a handwritten character, which character is this?

Character Recognition



It is difficult to program a solution to this.

```
if (I[0,5]<128) & (I[0,6] > 192) & (I[0,7] < 128):
  return 'A'
elif (I[7,7]<50) & (I[6,3]) != 0:
  return 'Q'
else:
  print "I don't know this letter."
```

Why Machine Learning?

Some problems are difficult or even impossible to formalise as a computer science problem, but humans can provide examples or feedbacks.

- Given an image of a handwritten character, which character is this?

Character Recognition

It is difficult to program a solution to this.

```
if (I[0,5]<128) & (I[0,6] > 192) & (I[0,7] < 128):
  return 'A'
elif (I[7,7]<50) & (I[6,3]) != 0:
  return 'Q'
else:
  print "I don't know this letter."
```


University of Sussex

Why Machine Learning?

Some problems are difficult or even impossible to formalise as a computer science problem, but humans can provide examples or feedbacks.

- Given an image of an animal, which animal is this?

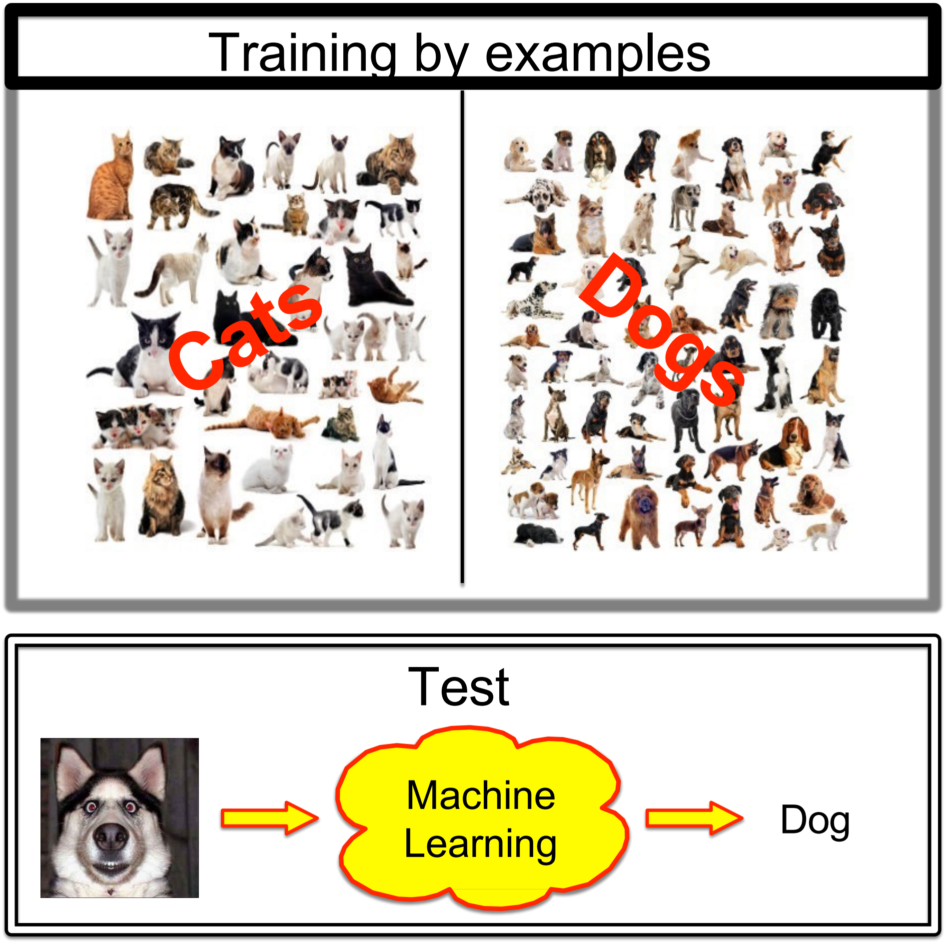
Object Category Recognition

It is *impossible* to program a solution to this.

```
if ???
```

Why Machine Learning?

The good news: we have examples.

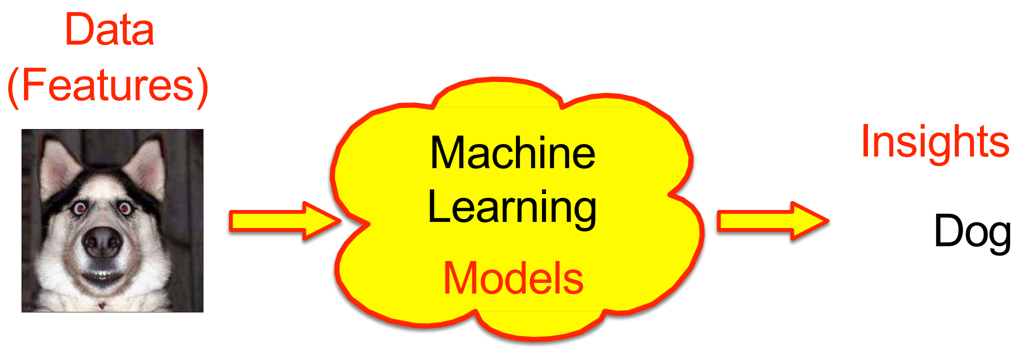


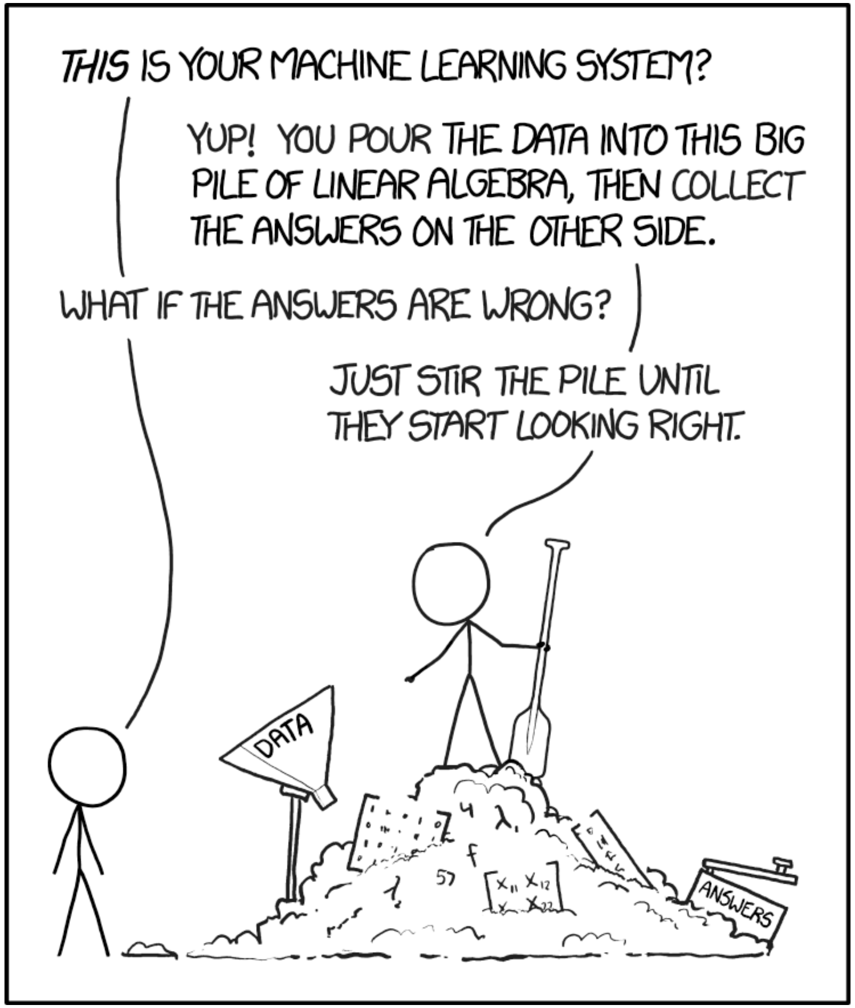
University of Sussex

Tasks, Models, and Features

Three main ingredients of machine learning:

- **Tasks**:
  - ▶ problems that require a mapping from data to desired outputs (insights).
- **Features**:
  - ▶ characteristics of the data used to describe domain objects.
- **Models**:
  - ▶ encode the required task mapping.





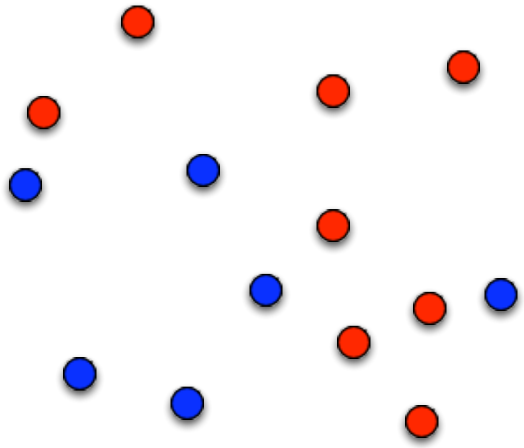
THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

(Some) Machine Learning Tasks if You have Health Data

- Classification
  - Binary ($C = 2$) classification involves separating data into two distinct groups (+ and −) e.g. distinguishing people at risk from heart disease (+), from those not at risk (−).
  - Generalise to $C > 2$ different classes
- Regression
  - Involves mapping from data items to real values
    e.g. quantifying the risk of heart disease on the basis of personal health records
- Clustering
  - Separating data into different *clusters/concepts* on the basis of their characteristics.
    e.g. grouping people according to their genetic characteristics.
- Collaborative filtering
  - Identifying rules or associations from data
    e.g. "recommending" products or films on the basis of patient records and diseases of similar patients.
- Dimensionality reduction
  - Visualising the data

Teaching your models : Supervised Learning

- There are different teaching paradigms in machine learning.
- In supervised machine learning, each element of the training data has associated labels.
  - ► These labels are the appropriate "output" for this data example
    e.g. binary classification: patient data already labelled with "at risk of heart disease" or "not at risk of heart disease".
  - ► Model should generalise from training data
    i.e. be good at predicting output for unseen data.

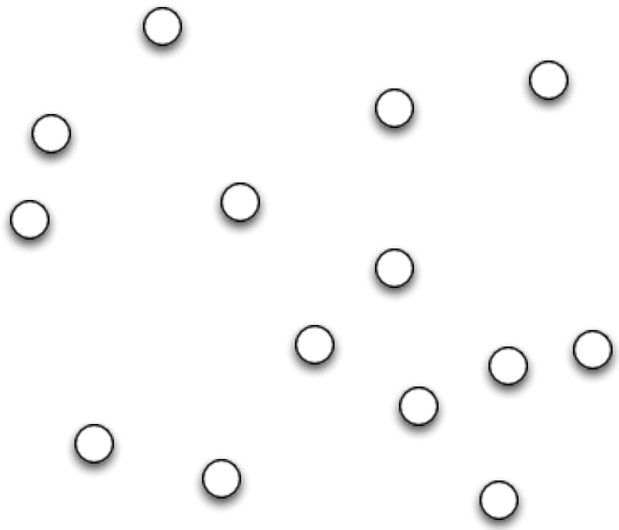


Supervised Learning

University of Sussex

Teaching your models : Unsupervised Learning

- Unsupervised machine learning does not require labels.
- Learns structure and patterns based on the similarities and differences in the data features of the training examples.
  e.g. clustering: grouping people on basis of genetic similarity in order to discover interesting sub-groups.

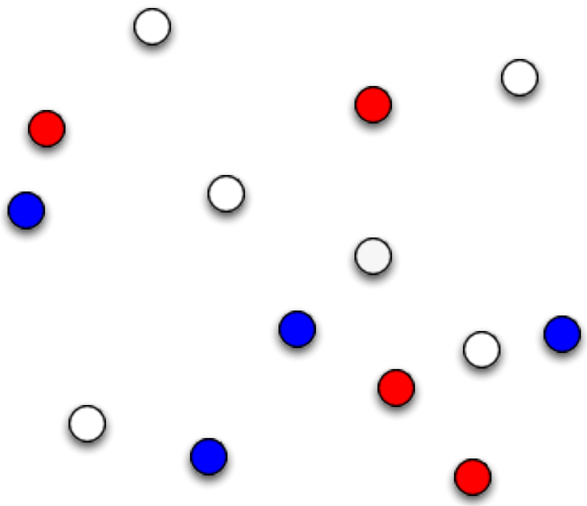


Unsupervised Learning

University of Sussex

Teaching your models : Degrees of supervision

- The choice is not always full/no supervision.
- Semi-supervised machine learning methods use labels on a subset of the data.



- This is particularly relevant when the labels are expensive

University of Sussex

Data and Features

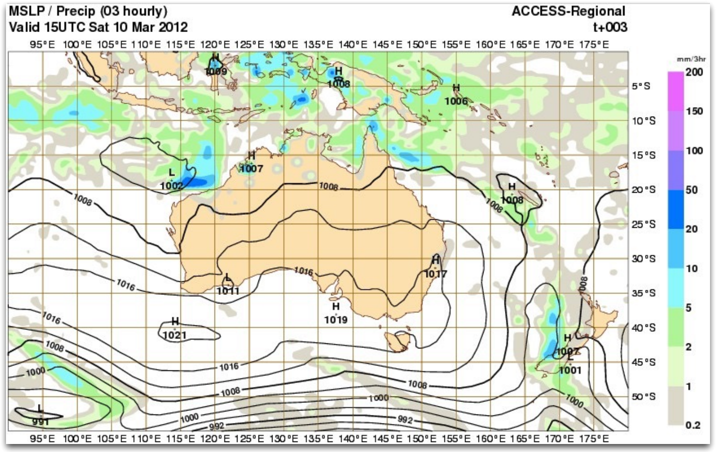
- Features are the inputs to machine learning:
  - ▶ Used to describe data objects
  - ▶ Represent particular characteristics of an instance
  - ▶ May be numerical (e.g. age, weight, income), boolean (e.g. is employed, is male), ordinal (e.g. SES, satisfaction), or nominal (e.g. nationality).
- A model is only as good as the information it sees:
  - ▶ Choice of features is therefore extremely important
  - ▶ Techniques for removing redundant features or transforming features in various ways are often crucial

Machine Learning Models

- Models form the central concept in machine learning:
  - ► A framework for making predictions from features
  - ► With parameters learned from data
  - ► Models encode the mapping needed to solve a task
- We will talk about several different models during this module.

Quiz Time!

- Go to https://pollev.com/bdevans



10 °C
Wind : 6mph
Humidity : 40%

Which machine learning task is this?

A Classification

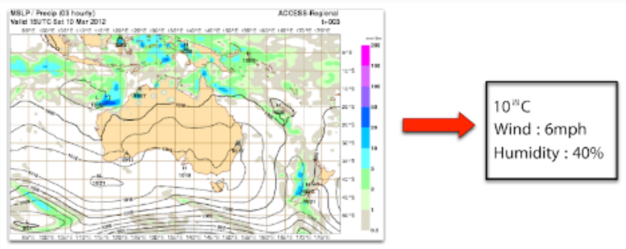
B Regression

C Clustering

D Collaborative Filtering

E Dimensionality Reduction

What type of ML task is this?



Clustering

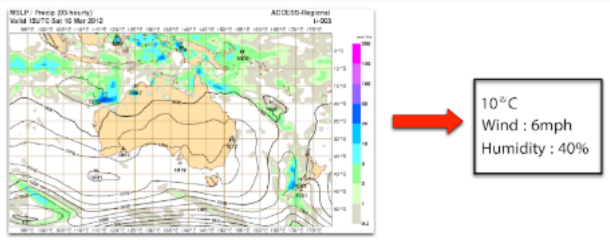
Regression

Classification

Collaborative Filtering

Dimensionality Reduction

What type of ML task is this?



Clustering
0%

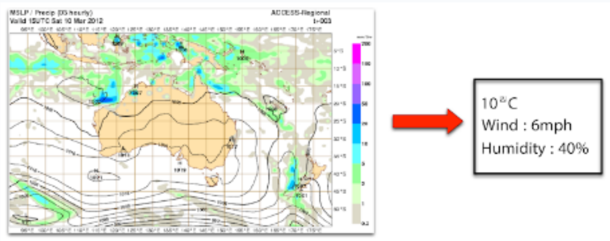
Regression
0%

Classification
0%

Collaborative Filtering
0%

Dimensionality Reduction
0%

What type of ML task is this?



Clustering

0%

Regression

0%

Classification

0%

Collaborative Filtering

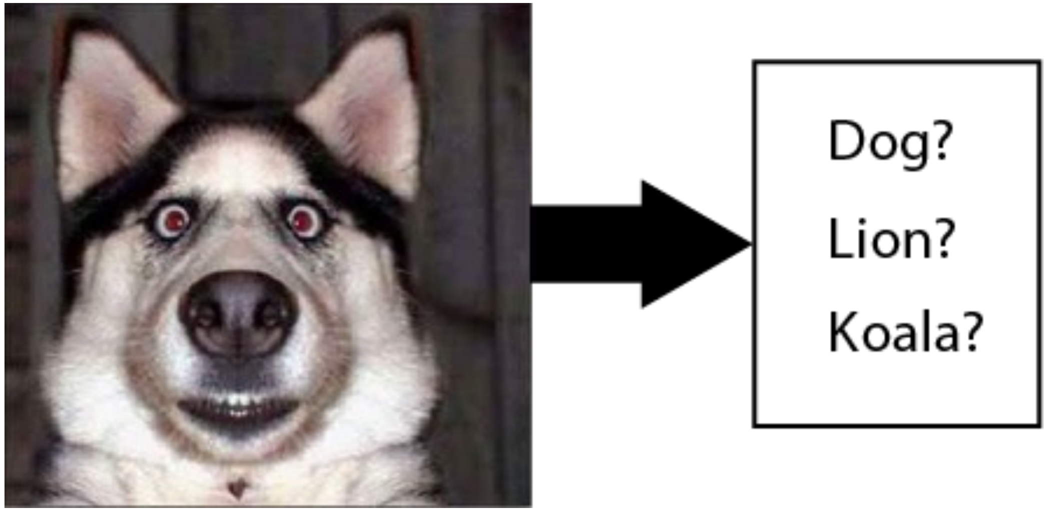
0%

Dimensionality Reduction

0%

Quiz Time!

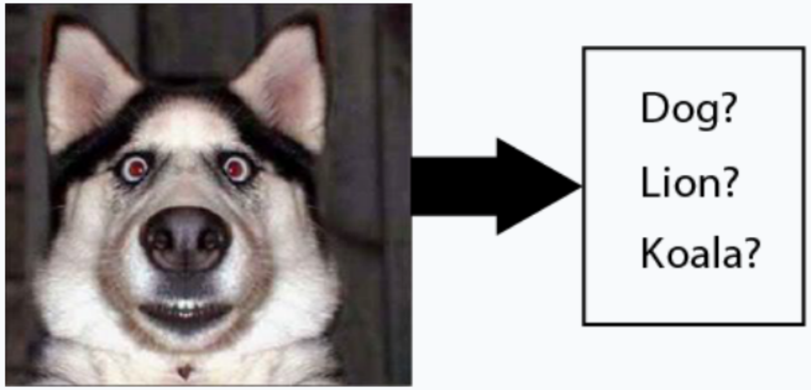
- Go to https://pollev.com/bdevans



Which machine learning task is this?

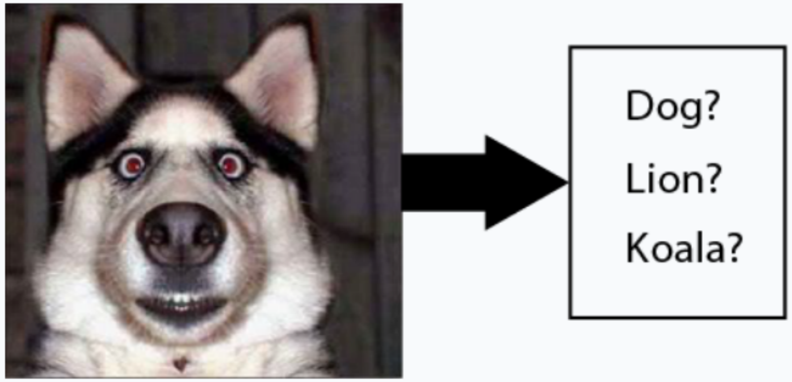
A Classification
B Regression
C Clustering
D Collaborative Filtering
E Dimensionality Reduction

What type of ML task is this?



- Clustering
- Regression
- Classification
- Collaborative Filtering
- Dimensionality Reduction

What type of ML task is this?



Clustering 0%

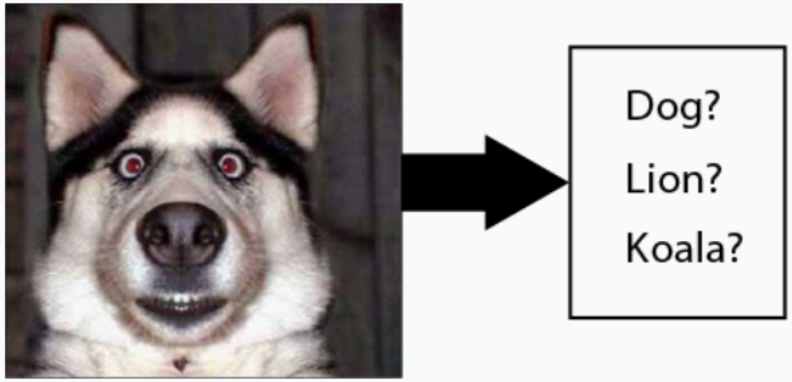
Regression 0%

Classification 0%

Collaborative Filtering 0%

Dimensionality Reduction 0%

What type of ML task is this?



Clustering — 0%

Regression — 0%

Classification — 0%

Collaborative Filtering — 0%

Dimensionality Reduction — 0%