University of Exeter

Introduction to Machine Learning

Coding for Reproducible Research

March 2025

Collaborative doc: https://tinyurl.com/fd3cd22b

Course Leader:

- Simon Kirby

Course Helpers:

- Finley Gibson
- Sam Fletcher

Sign in here →

Code of Conduct

- Our ethos is to provide a welcoming and supportive environment for all people, regardless of background or identity. By registering to attend this workshop, participants are agreeing to abide by the Researcher Development Code of Conduct.

- Our goal is to support you to develop your programming skill sets to enable you to do cutting edge research. We want to create a positive and professional learning environment and therefore encourage the following kinds of behaviours:

    - Show courtesy and respect towards all who attend a workshop or engage in community events.
    - Be respectful of different viewpoints and experiences.
    - Gracefully accept constructive criticism.
    - Be patient if there are technical glitches. While we know something about how to use computers, we are not immune to internet or hardware issues.
    - Respect our policy on not recording workshops to protect the nature of the sessions and ensure we are GDPR compliant.

Programme Funding

The CfRR training programme is supported by:

- Research Software Analytics Group
- Institute for Data Science and Artificial Intelligence (IDSAI)
- University of Exeter Reproducibility Leadership Team
- EPSRC Research Software Engineering Fellowship
- Community of academics who volunteer their time to support delivery

To make the case for continued investment, please help us demonstrate the impact of these sessions by attending all courses you register for and providing feedback at the end of the course.

University of Exeter

Intro to Machine Learning

Part 1 - What is machine learning?

Course contents

Session 1

- **Slides: what is machine learning?**
- Tutorial: linear regression
- Slides: model selection and evaluation

Session 2

- Tutorial: model selection and evaluation
- Slides: the machine learning pipeline
- Tutorial: machine learning pipeline task

Session 3

- Continue with machine learning pipeline task
- Tutorial: unsupervised learning

What is machine learning?

University of Exeter

Discuss with your neighbour!

What is machine learning?

University of Exeter

What is machine learning?

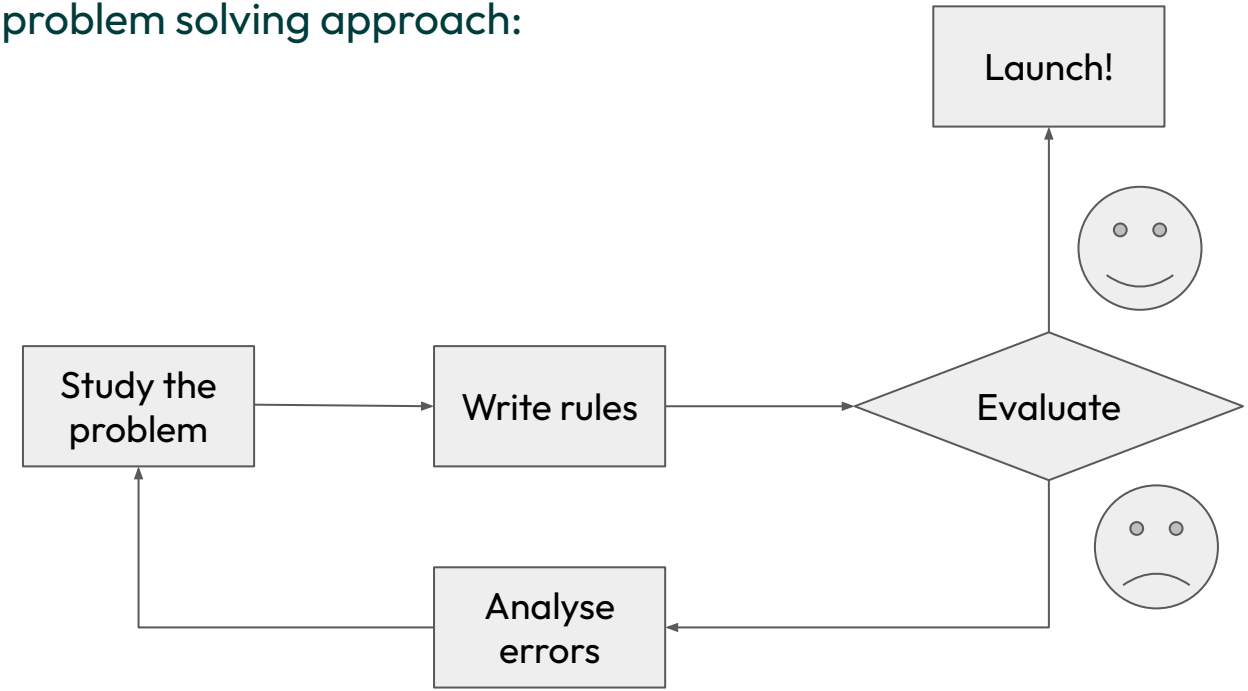
“Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed.” - Arthur Samuel, 1959

“A computer program is said to learn from experience E with respect to some task T and some performance measure P, if the performance on T, as measured by P, improves with experience E.” - Tom Mitchell, 1997

The science (and art) of programming computers so they learn from data.

What is machine learning?

Traditional problem solving approach:

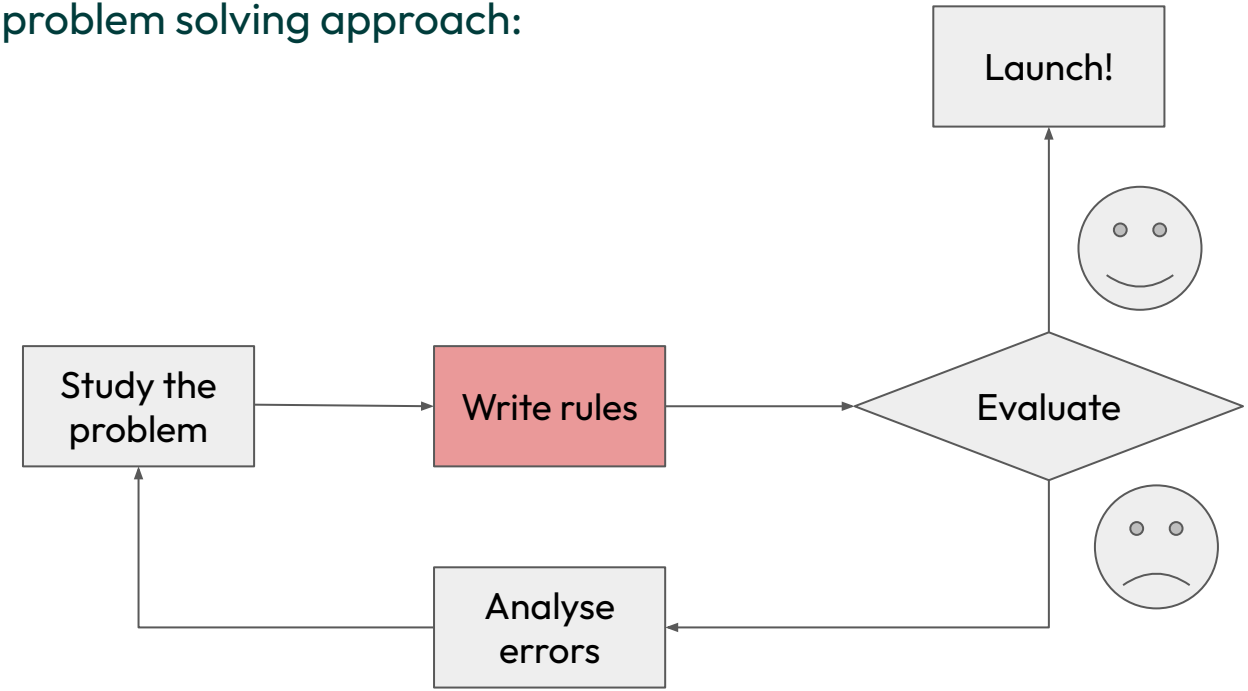


Diagrams adapted from Hands On Machine Learning with scikit-learn, Keras and Tensorflow by Aurelien Geron

What is machine learning?

University of Exeter

Traditional problem solving approach:

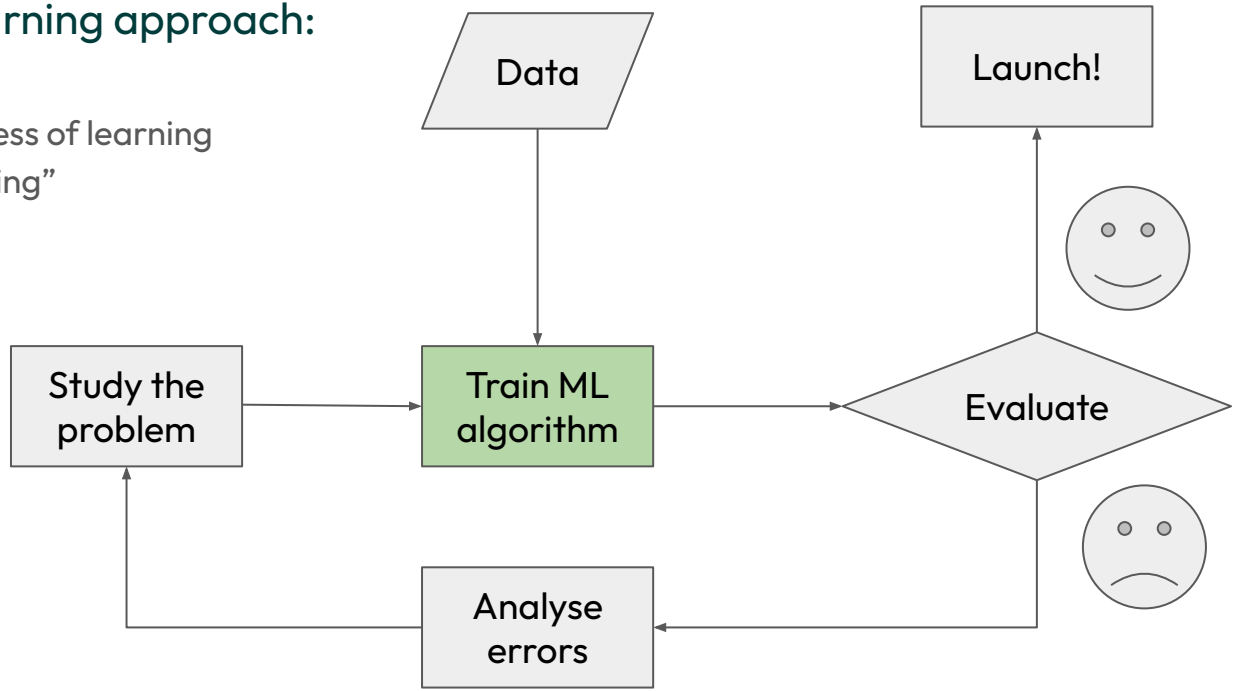


Diagrams adapted from Hands On Machine Learning with scikit-learn, Keras and Tensorflow by Aurelien Geron

What is machine learning?

Machine learning approach:

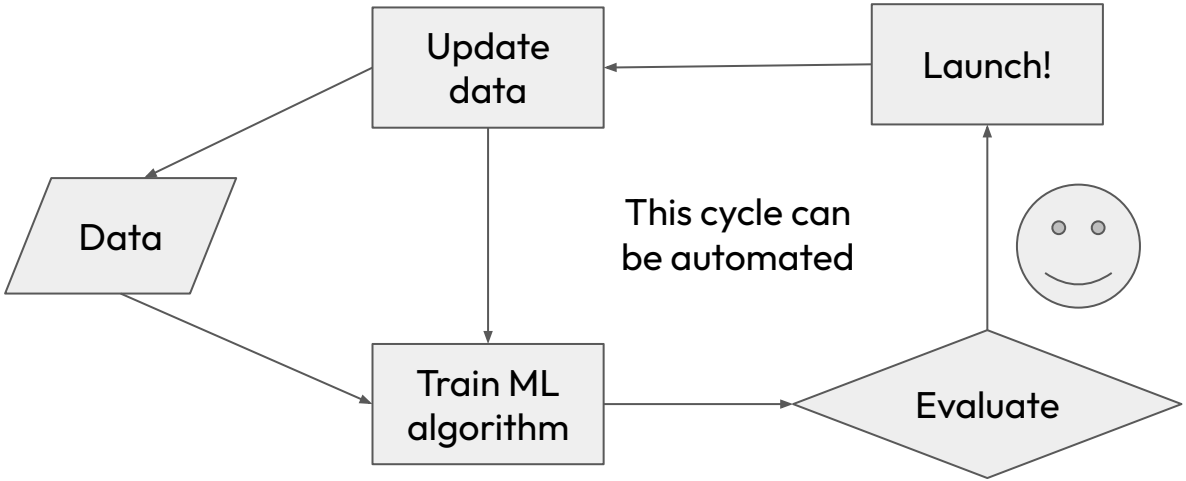
We call this process of learning from data “training”



Diagrams adapted from Hands On Machine Learning with scikit-learn, Keras and Tensorflow by Aurelien Geron

What is machine learning?

Automatically adapting to change:

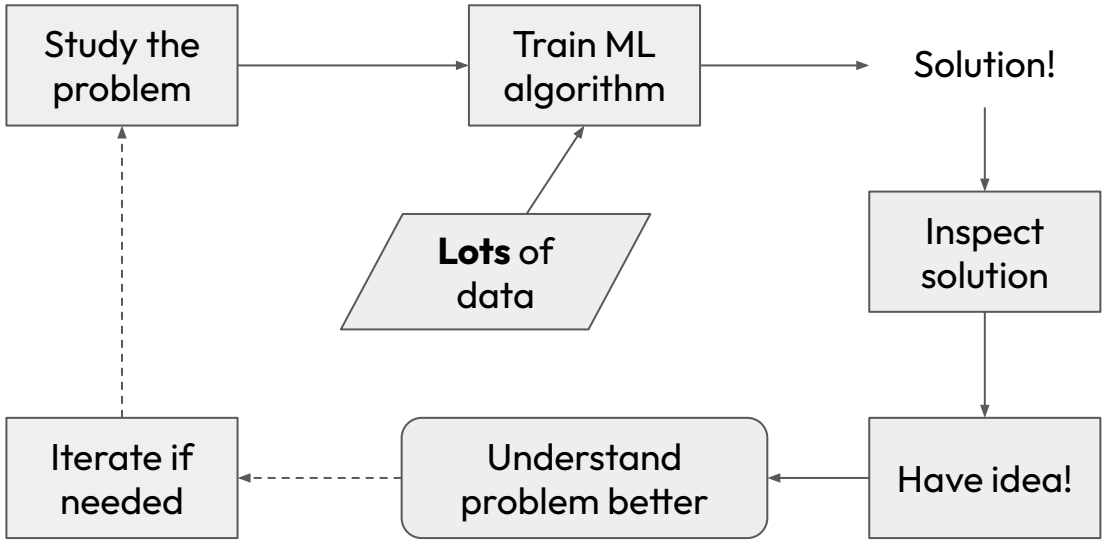


Diagrams adapted from Hands On Machine Learning with scikit-learn, Keras and Tensorflow by Aurelien Geron

What is machine learning?

University of Exeter

ML can help humans learn, especially in research:



Diagrams adapted from Hands On Machine Learning with scikit-learn, Keras and Tensorflow by Aurelien Geron

What is machine learning good for?

Problems that require a lot of fine tuning, or long lists of rules. ML-based approaches can simplify the process.

Complex problems where traditional approaches do not yield a solution. Maybe ML can find a solution.

Fluctuating environments: where approaches need to adapt to new data.

Getting insights about complex problems and large amounts of data

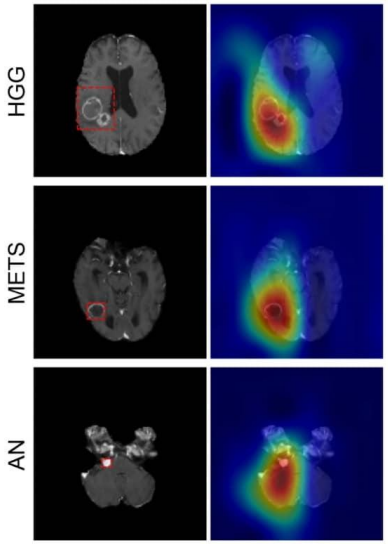
Example applications

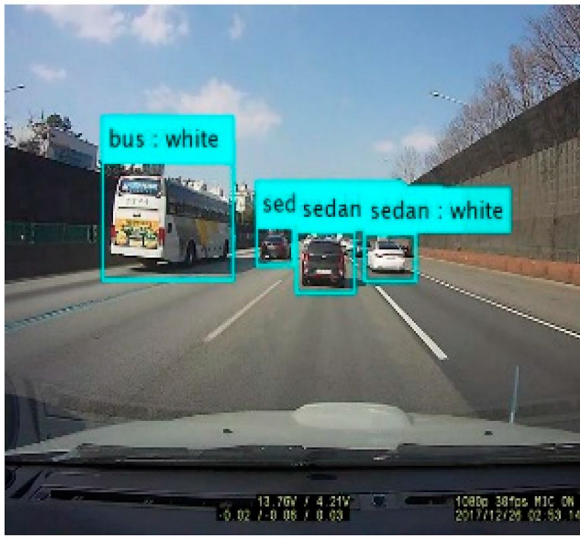
University of Exeter

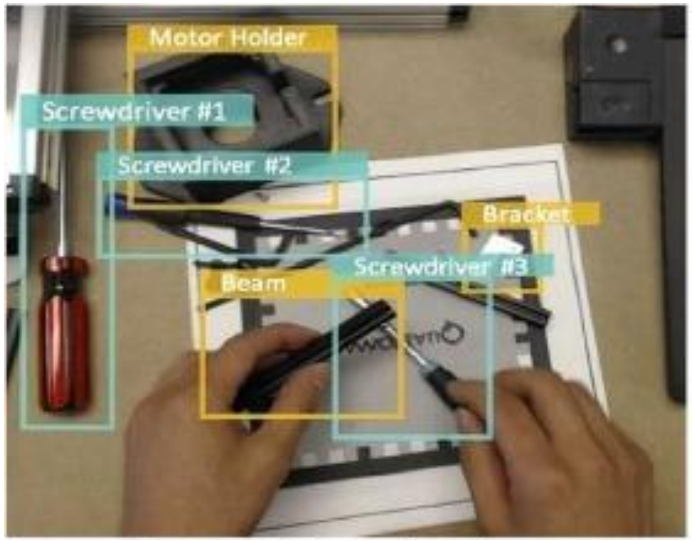
Discuss with your neighbour!

Example applications

Identifying brain tumours, classifying auto traffic, AR/VR object detection

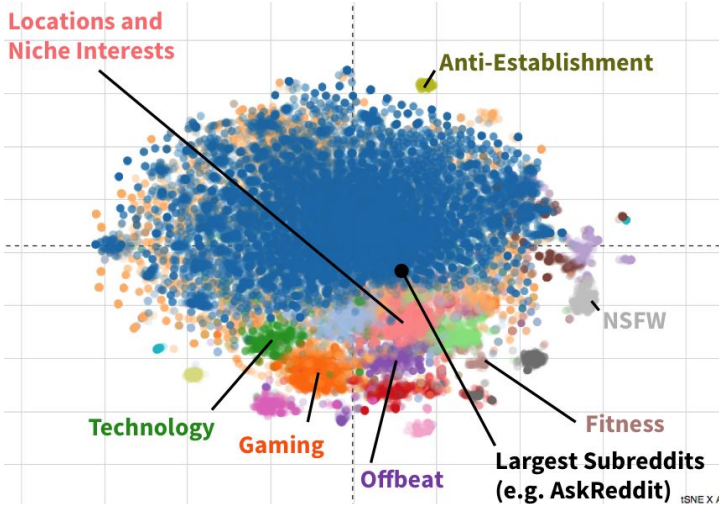






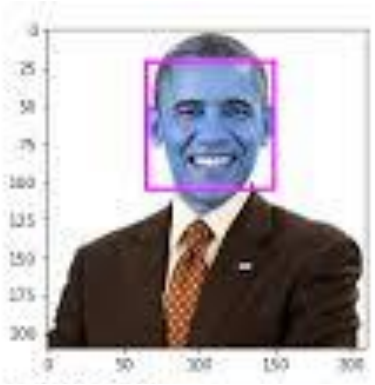
Example applications

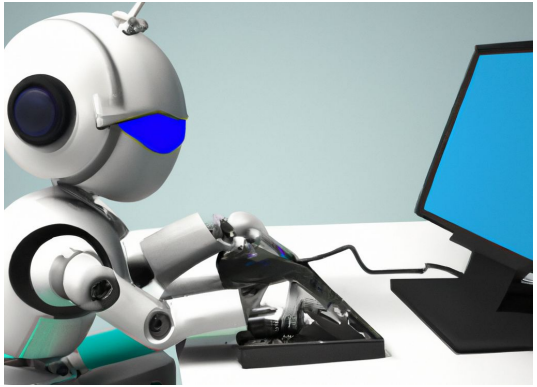
Identifying fraudulent payments, forecasting sales growth, identifying groups of social media users



Example applications

Face recognition, building a bot for a game, building a chatbot.





ChatGPT

And many, many more!

Types of machine learning task

All of these examples are solved using different machine learning algorithms.

We can classify both the type of machine learning task, and the genre of algorithm used to approach the problem.

Problem type: classification vs regression

Approach type: supervised vs unsupervised vs reinforcement learning

Types of machine learning task - classification

University of Exeter

Classification: given a new sample, classify it into a predefined group

Output: a predicted label name from a predefined group

Examples:

- Is this new email spam, or not spam?
- Is the next pixel in a scan cancerous or not cancerous?
- What object is in front of me on the road?
- Did person X survive the Titanic disaster?
- Will a customer on my website purchase the item they are looking at?

Types of machine learning task - regression

University of Exeter

Regression: given a new sample, predict a continuous numerical value from the samples input features.

Output: a continuous numerical value

Examples:

- What is the predicted fuel efficiency of a particular model of car?
- What is the predicted sale value of a house, given its rooms, size & location?
- How old is the person in a recording, given their facial/vocal features and behaviour?
- How much money will a film earn, given its release date, genre, cast, etc?

Types of machine learning systems/algorithms

We can categorise ML approaches based on the amount of supervision they require when learning from data:

- **Supervised learning** - requires the data and the solutions (labels)
- **Unsupervised learning** - requires the data, but no solutions (labels) provided
- **Reinforcement learning** - observe system, update policy based on reward

Supervised learning

- Supervised learning algorithms train using input data that has been labelled.
- A model is trained and its performance evaluated.
- This model can make predictions on a new sample (that is unlabelled.
- The model has learnt to generalise from the set of examples.

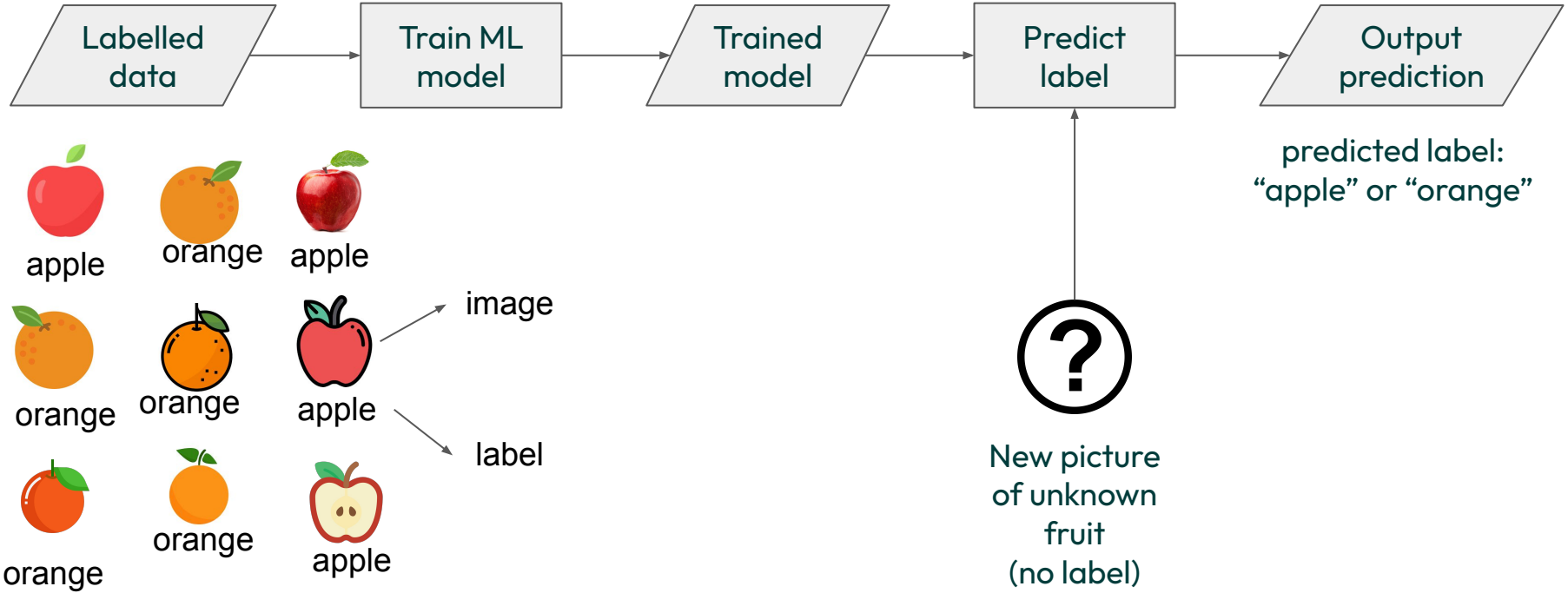
Example algorithms:

- Linear/logistic regression, SVMs, random forests, etc
- Neural networks

Supervised learning - binary image classification

University of Exeter

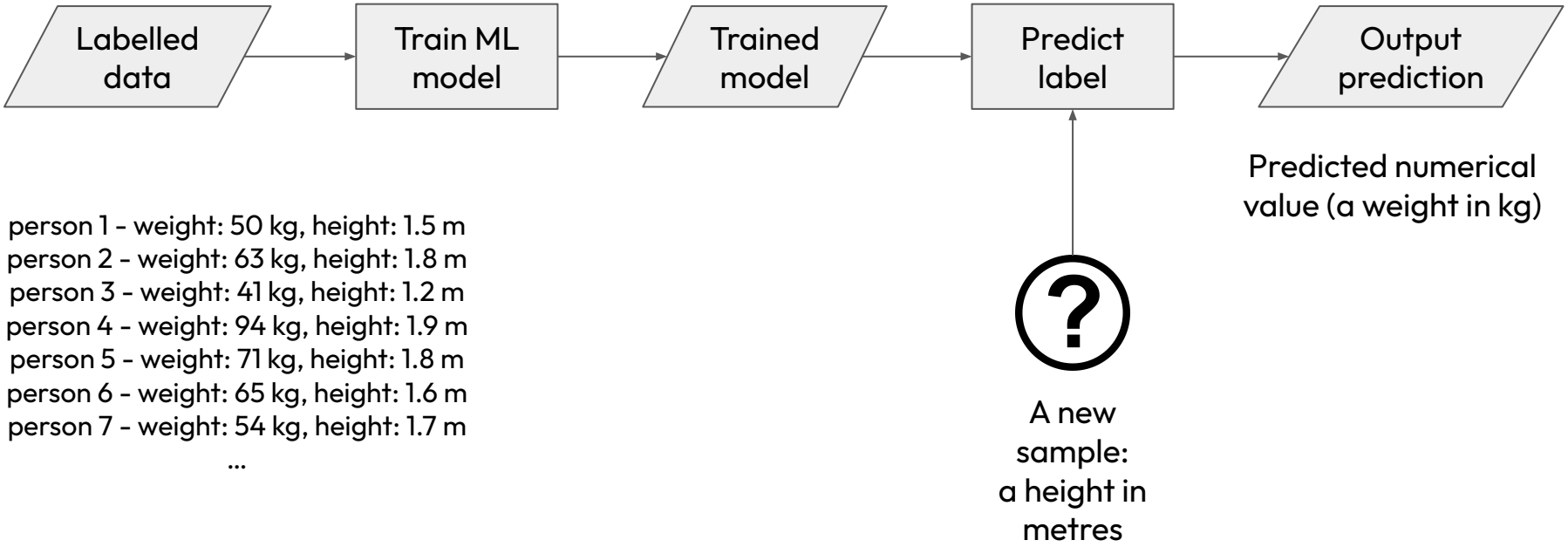
Predict whether a picture of an unknown fruit is an apple or an orange.



Supervised learning - linear regression

University of Exeter

Predict a weight based on height



Unsupervised learning

- Unsupervised learning algorithms train using data that is not labelled.
- They are often used to identify groups in data without human intervention. However, this is not a magic bullet! Results can be variable.
- This task is commonly called clustering.

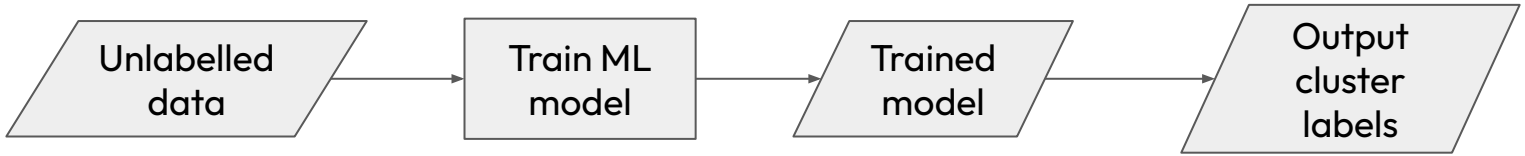
Example algorithms:

- K-means
- Principal component analysis (PCA), used for dimensionality reduction

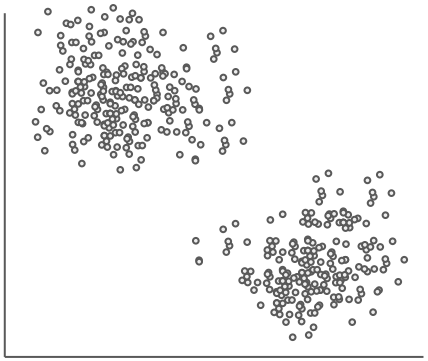
Unsupervised learning - K-means clustering

University of Exeter

Predict k cluster labels (identify groups) in a given dataset

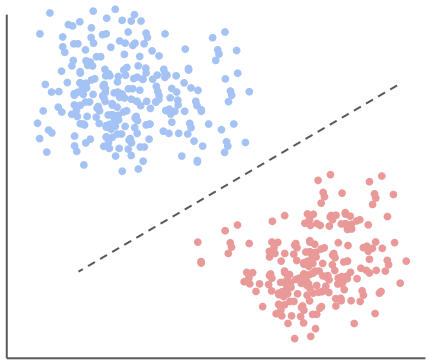


Unlabelled data: before K-means



After K-means: data coloured by cluster label

With K = 2



Reinforcement learning (RL)

University of Exeter

- RL algorithms (or agents) are trained to make decisions to achieve a goal with optimal results. Actions that move the agent towards their goal are reinforced, and those that move the agent away are ignored/penalised.
- This is a reward/punishment approach to improve the agents policy, that can often find solutions humans cannot anticipate.
- RL is often used in complex tasks, famously in AlphaGo.

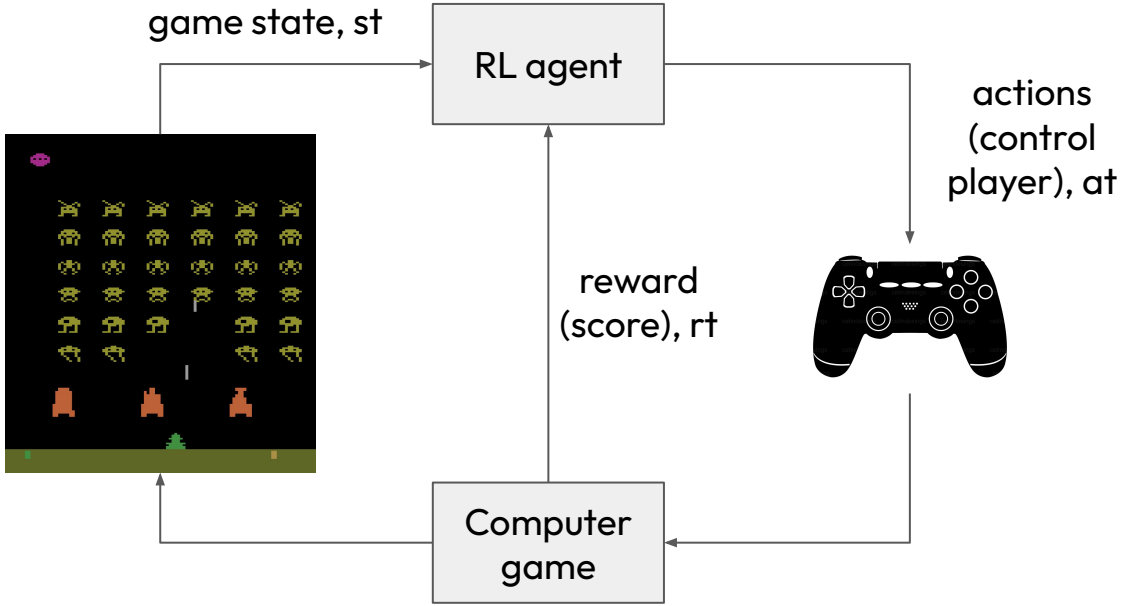
Example algorithms:

- Q-learning, policy gradient methods

Reinforcement learning

University of Exeter

Goal: play a 2D game with no initial knowledge of the rules



- Agent observes system state.
- Agent takes action using its policy (moves the player).
- Agent observes change in reward i.e. score.
- Agents policy is gradually improved by rewarding successful actions, and penalising unsuccessful ones.

Challenges with machine learning

University of Exeter

Data challenges

- Quantity: not enough training data to generalise across new samples
- Quality: garbage in, garbage out
- Bias: non-representative or biased data can lead to poor predictions

Algorithm challenges

- Overfitting: training a model that predicts well on training data, but not on new samples
- Underfitting: using a model too simple to make good predictions
- Picking an algorithm: no free lunch theorem

Thank you!

University of Exeter

Post-workshop Anonymous Feedback Form 2024-25
https://tinyurl.com/2d8fys7e