Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Fundamentals of Machine Learning

Introduction

WEDI'I YRRU GAN FFYNIANT BRO | POWERED BY LEVELLING UP

Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Sessions

Week 1:	Week 2:	Week 3:	Week 4:
Introduction	Supervised Learning	Neural Networks	Reinforcement Learning

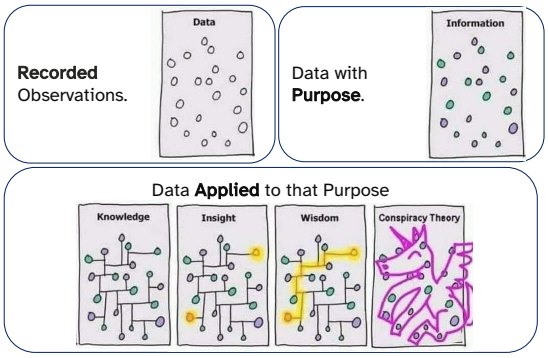
Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Data

Funded by UK Government Wedi ei ariannu gan Llywodraeth y DU

institute of CODING in wales technocamps

technocamps



Funded by UK Government Wedi ei ariannu gan Llywodraeth y DU

institute of CODING in wales technocamps

technocamps

Machine Learning as a Tool

- It can be good to think of machine learning as a tool, using parts of other fields to solve real-world issues based on data.
- We may want to find patterns:
  - Detect anomalies or outliers
  - Cluster groups of similar things
  - Identify relationships between things
  - Apply a label to an observation
- These identified patterns are only useful if they allow us to then do something, we want actionable insights.

Funded by UK Government Wedi ei ariannu gan Llywodraeth y DU

institute of CODING in wales technocamps

technocamps

Machine Learning as a Tool

- A common example is to predict whether email is spam or not.
- Let's say that I've seen numerous emails over the last few years which tell me I can "make money quick" in some form or another. Those emails have turned out to be spam.
- My friends and colleagues send me genuine, non-spam emails, and they don't contain any mention of making money quick.

Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING technocamps in wales | technocamps

Machine Learning as a Tool

- We could create a **rule** that says:

  IF subject contains “make money quick”

  then the email is spam, otherwise it is not.
- Given a new email, we **apply our rule** and **predict** whether it should be labelled as spam or not.
- We **classify** it’s **label** from the set of possible labels

  {spam, non-spam}
- Based on something we observed (the subject content).

Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING technocamps in wales | technocamps

Machine Learning as a Tool

- However, it is quite weak rule, easy to bypass and pontentially could get it wrong (misclassify).
- If I can spot that trend as a human, then what is the point of throwing machine learning at it?
- Instead we often want to find patterns/insights which humans could not find easily:
  - Complex patterns
  - Massive amounts of data

Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING technocamps in wales | technocamps

Data

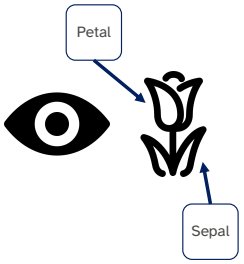
- Dataum
  - “A single piece of information”
  - “A item of data”
  - “A single recorded observation”
- A dataum is often an observation or measurement of something, recording information about that thing
- A collection of multiple observations can form from data

Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Data

- Suppose we observe a flower
- What could we measure/record about that flower?

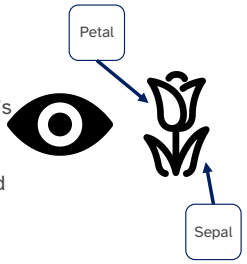
- Colour/ Number of petals
- Length/ Width of petals
- Length/ Width of sepals



Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Data

- Why might we want to?
- Well perhaps we have a hypothesis that we might be able to tell what the flower's **species** is by these measurements.
- Similar to our example example, we might want to predict the species based on our data, so that if we see a new flower, we can **classify it**



Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

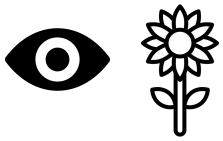
Data

- Our one observation becomes a single data point:
  - Colour: Black
  - Number of petals: 3
  - Length of petal: 1.5cm
  - Width of petal: 2cm
  - Length of speal: 0.7cm
  - Width of sepal: 0.3cm
  - Species: Black rose

Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Data

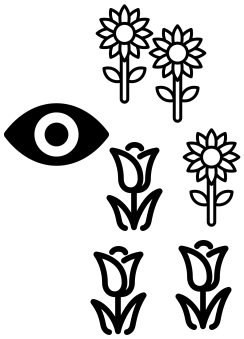
- Our one observation becomes a single data point:
  - Colour: Yellow
  - Number of petals: 12
  - Length of petal: 13cm
  - Width of petal: 2cm
  - Length of speal: 3.5cm
  - Width of sepal: 2cm
  - Species: Sunflower



Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Data

- If we then look at a lot of flowers, lots of observations perhaps we can create a large dataset on which we can apply our machine learning methods
- Maybe there are some trends in that data we collected which can help us solve our task



Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Data

- Data comes in a variety of shapes, sizes and types.
- We may collect numerical data, text data, images…
- We may collect pieces of information about a single observation. E.g. Name, age, address …
- We may collect tens, thousands, or even millions of data points

Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Data Types

Categorical

Nominal
- No order

Examples:
- Eye colour
- Postcode
- Star sign

Ordinal
- Ordered

Examples:
- Race position
- Ratings
- Degree level

Numerical

Discrete
- Countable

Examples:
- Class size
- Cars owned
- Fake teeth

Continuous
- Measurable

Examples:
- Salary
- Height
- Time

Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Dataset

Features

Sepal Length	Sepal Width	Petal Length	Petal Width	Species
5.1	5.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
7.0	3.2	4.7	1.4	versicolor
6.4	3.2	4.7	1.4	versicolor
6.3	3.3	6.0	2.5	virginica
5.8	2.7	5.1	1.9	virginica

- **Samples**
- **Instances**
- **Data Points**

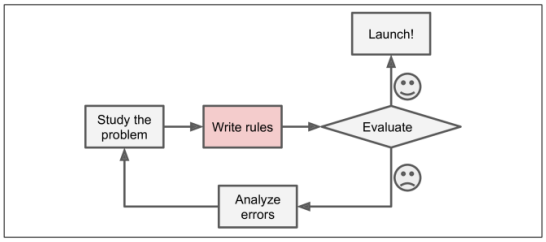
Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Machine Learning: Motivation

Funded by UK Government Wedi ei ariannu gan Llywodraeth y DU

institute of CODING in wales technocamps

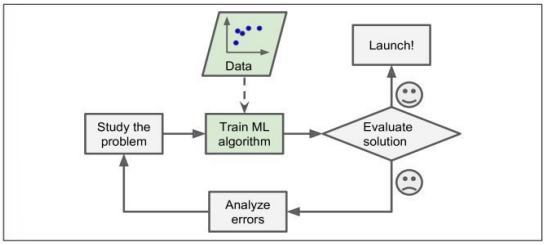
Traditional Approach



Funded by UK Government Wedi ei ariannu gan Llywodraeth y DU

institute of CODING in wales technocamps

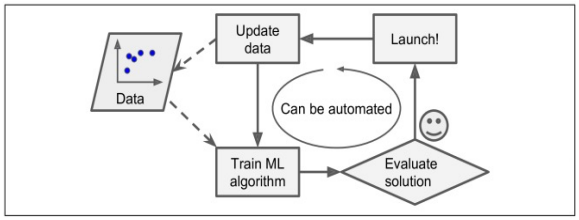
Machine Learning Approach



Funded by UK Government Wedi ei ariannu gan Llywodraeth y DU

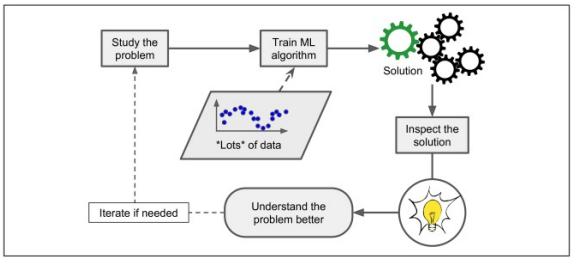
institute of CODING in wales technocamps

Automated Learning



Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Learning from Learning



Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Machine Learning: Types

Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Distinguishing Criteria

Three criteria to separate machine learning algorithms:

- Human involvement during training
  The machine learning *paradigm*
- When data is used to learn.
  Batch vs *Online* learning
- How data is used to learn
  Model-based vs *Instance-based* learning.

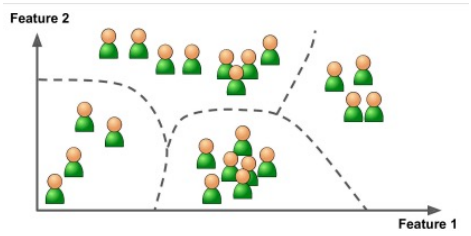
Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Learning Paradigms

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

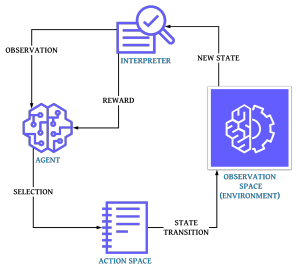
Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Unsupervised Learning



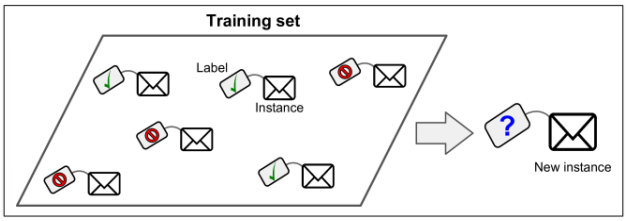
Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Reinforcement Learning



Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Supervised Learning: Classification



Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Supervised Learning: Regression

Funded by UK Government | Wedi ei ariannu gan Llywodraeth y DU | institute of CODING in wales | technocamps

Supervised Learning: Training and Evaluation