Machine Learning

Introduction

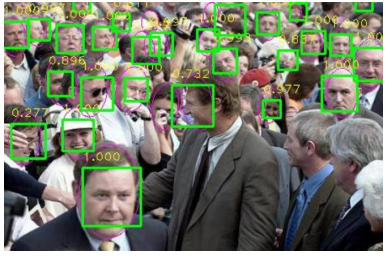
Kia Nazarpour

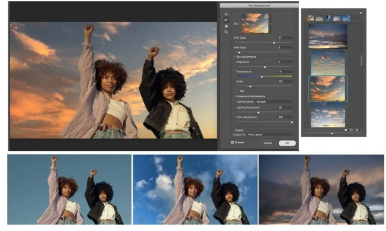
https://homepages.inf.ed.ac.uk/htang2/mlg2024/

Based on Hao Tang's slides

KN recognises the importance of Image Credits. They have been removed to avoid cluttering the slides.

Context: Image Processing





Context: Recommender systems

NETFLIX

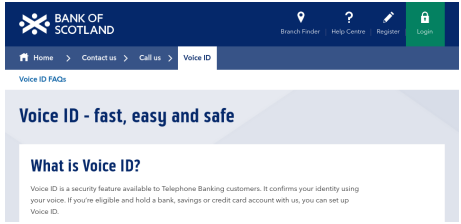
Spotify®

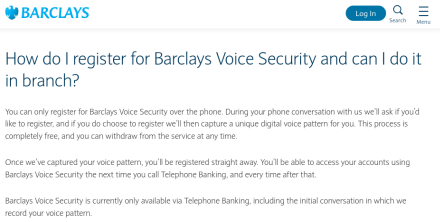
TikTok

YouTube

Context: Speech recognition

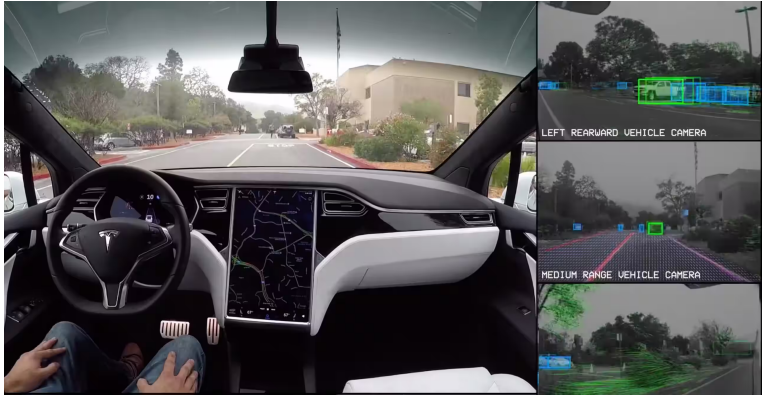
Context: Speech verification





Context: Robotic vacuum cleaner

Context: Autonomous driving



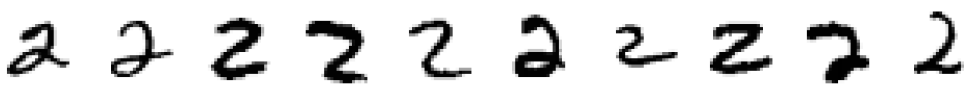
Context: Failures

Learning Outcomes

1. Define machine learning, identify its main types
2. Understand the core components of a machine learning systems
3. Learn about what we will do in this course

What is machine learning?

How would you write a program to recognise hand-written 2s?

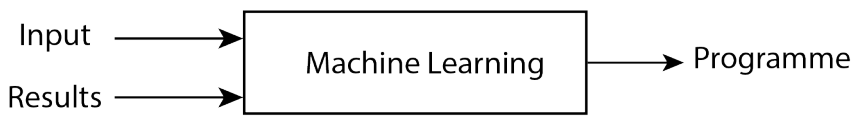


We "produce" a programme that recognises a 2 using the examples of 2s.

What is machine learning?

Traditional Computing

Machine Learning



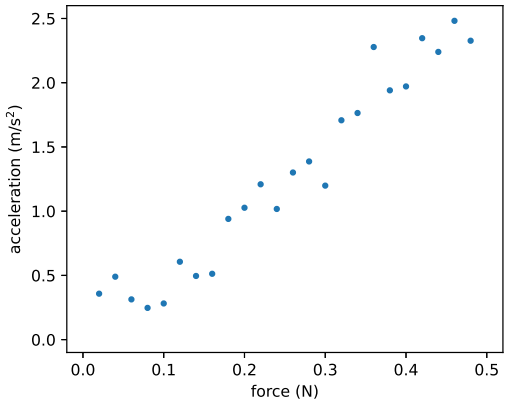
Example 1

In an experiment we measure force and acceleration. Do you see a *pattern*?

force (N)	acceleration (m/s^2)
0.02	0.358
0.04	0.490
0.06	0.313
0.08	0.247
0.10	0.282
0.12	0.606

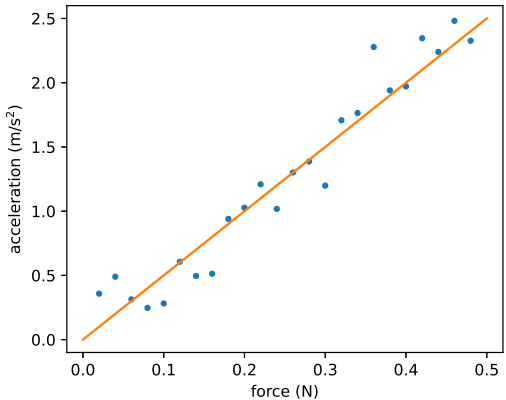
Example 1 - Linear Regression

Let's plot the data to *explore* it.

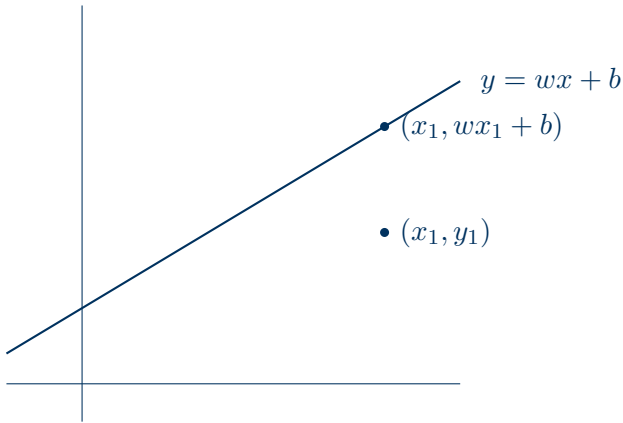


Example 1 - Linear Regression

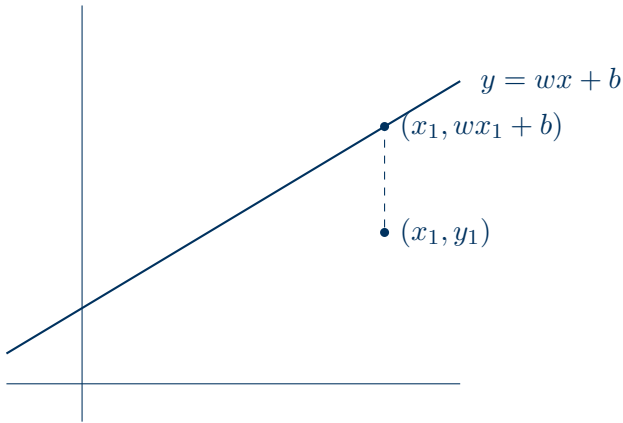
Let's plot the data to *explore* it.



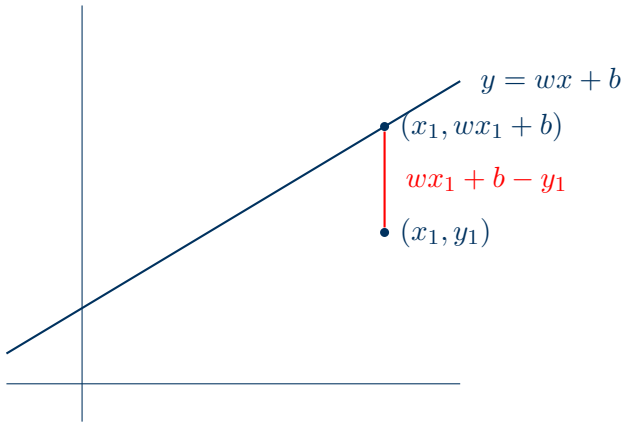
Linear Algebra - A quick reminder



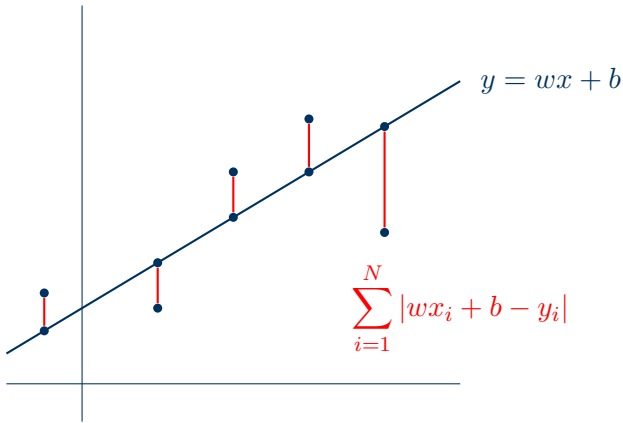
Linear Algebra - A quick reminder



Linear Algebra - A quick reminder



Linear Algebra - A quick reminder



Linear Regression in the Machine Learning Language

- Given N points $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$,

$$\text{sum of absolute errors} \qquad L = \sum_{i=1}^{N} |wx_i + b - y_i|. \tag{1}$$

- Find w and b that minimises L.

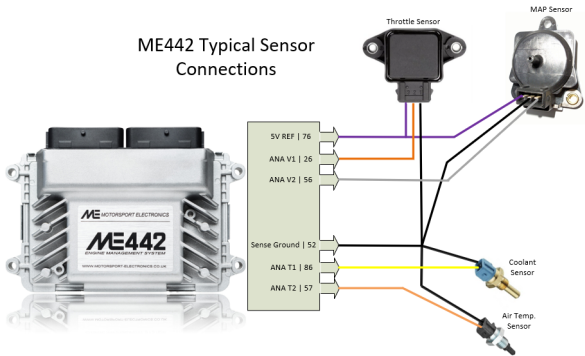
- Find a function $f(x) = y = wx + b$ that minimises L.

Machine Learning Jargon

- The N points $\{(x_1, y_1), \ldots, (x_N, y_N)\}$ constitute a **data set**.
- The value w and b are called **parameters**.
- The function L is called a **loss function**.
- The act of finding w and b that minimizes L is called **training**.
- The $x_1, x_2, \ldots, x_N$ are called (input) **features**.
- The $y_1, y_2, \ldots, y_N$ are called (output) **labels**.
- Specifying the above gives us a **task**.

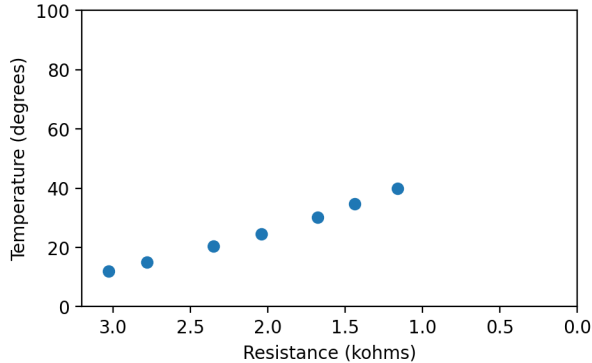
Example 2 - Temperature calibration

How much data is enough?



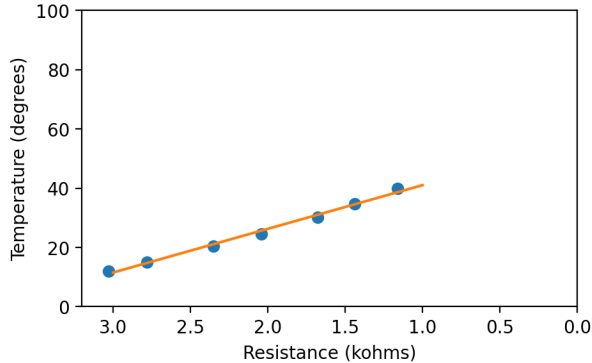
Example 2 - Temperature calibration

How much data is enough?



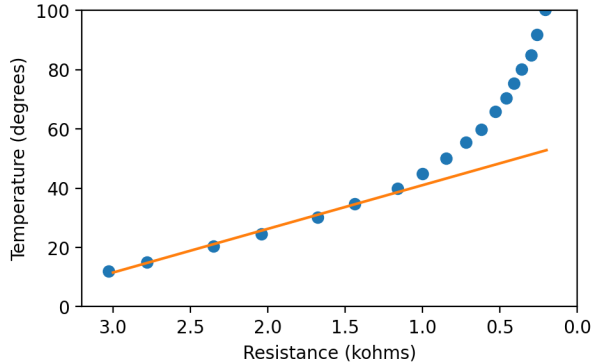
Example 2 - Temperature calibration

How much data is enough?



Example 2 - Temperature calibration

How much data is enough?



Generalisation

- A programme is **correct** if it produces the desired output on all input in the input domain.
- The fact that we use machine learning means that we do not have a good characterisation of the input. (If we do, we likely don't need machine learning.)
- Since the input domain is infinitely large, we only develop our programme on samples from the input domain.
- A programme **generalises** if it is developed with samples from the input domain but is able to produce the desired output on the entire input domain.

Our MLG Teaching Philosophy

We follow a fairly standard learning experience!

You need:

- Calculus
- Linear Algebra and Probability
- python, numpy, matplotlib, and Jupiter notebook

Logistics

- Course website: `https://homepages.inf.ed.ac.uk/htang2/mlg2024/`
- Textbooks (See MLG Calendar)
  1. Bishop, Pattern Recognition and Machine Learning, 2006
  2. Deisenroth, Faisal, and Ong, Mathematics for Machine Learning, 2020
  3. Lindholm, Wahlström, Linsten, and Schön, Machine Learning - A First Course for Engineers and Scientists, 2022
  4. Shalev-Schwartz and Ben-David, Understanding Machine Learning, 2014
- Exercises and notes
- We encourage you to review **Informatics 2 - Foundations of Data Science**.