Introduction to Machine Learning: History, Key Concepts, and Real-World Applications

Lida Kanari

Mathematical Institute
University of Oxford

Introduction to Machine Learning, October 2025

Oxford Mathematics

University of Oxford Mathematical Institute

Practical Information

Mathematical Institute

When: October 13 – December 7 @ 14:00
Where: L4 - room

Instructor: Lida Kanari
Email: lida.kanari@math.ox.ac.uk
Office: S3.37
Office hours: By appointment (please send me an email)
Special Topics Report: Machine Learning [C]
Project deadline: January 19

Table of Contents

In this lecture, we will cover four main topics:

- General introduction to machine learning problems
- Basic principles of ML
- A brief history of machine learning
- Real-world applications and examples
- What to expect from this course

What is Machine Learning and Why Does It Matter?

Oxford Mathematical Institute

Definition 1

Machine learning refers to algorithms that automatically learn patterns from data and improve their performance with exposure to larger datasets.

Machine learning can be categorized into main categories based on the learning algorithms: **supervised learning**, **unsupervised learning**, **self-supervised learning** and **reinforcement learning**.

Unlike traditional programming, which is rule-based, machine learning is primarily **data-driven**, allowing models to adapt and improve as they are exposed to more data.

What is Machine Learning and Why Does It Matter?

The core objectives of machine learning are the **automation of tasks**, the **prediction**, and the **discovery of new concepts**.

ML tasks include:

- classification
- regression
- clustering
- dimensionality reduction

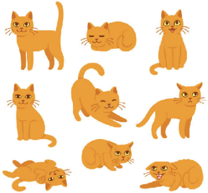
Machine Learning Model

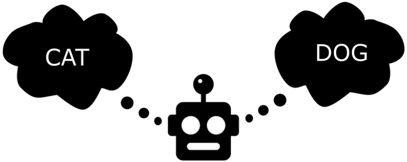
UNIVERSITY OF OXFORD
Mathematical Institute

- Data: input space $\mathcal{X}$, output space $\mathcal{Y}$.
- Model: a function $f_\theta : \mathcal{X} \to \mathcal{Y}$ with parameters θ.
- Training: find θ that minimize a loss function $L(f_\theta(x), y)$.
- Goal: generalize to unseen data, not just memorize training examples.

$$\theta^* = \arg\min_\theta \, \mathbb{E}_{(x,y)}\left[L(f_\theta(x), y)\right]$$

Supervised Learning





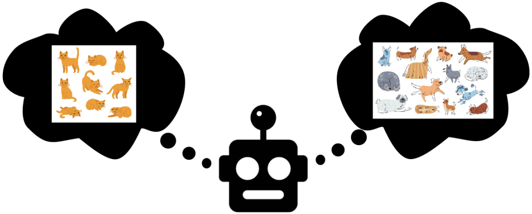
Example of supervised - classification

Supervised Learning

- Training data: $\{(x_i, y_i)\}_{i=1}^{n}$ with known labels y_i.
- Objective: learn $f_\theta(x)$ that predicts y from x.
- Examples: classification, regression.

$$\theta^* = \arg\min_\theta \frac{1}{n} \sum_{i=1}^{n} L(f_\theta(x_i), y_i)$$

Unsupervised Learning



Example of unsupervised - clustering

Unsupervised Learning

- Data: $\{x_i\}$ (no labels).
- Goal: discover hidden structure or distribution.
- Examples: clustering, dimensionality reduction, density estimation.

Find $g(x)$ that captures some intrinsic structure in the data $\{x_i\}$.

Reinforcement Learning

Example of reinforcement learning.

Reinforcement Learning

- A model interacts with the environment $\mathcal{E}$.
- At time t: observes state s_t, takes action a_t, receives reward r_t.
- Policy $\pi_\theta(a|s)$: probability of taking action a in state s.
- Goal: maximize expected reward.

$$\theta^* = \arg\max_\theta \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

ML vs Traditional Programming

- **Traditional:** Rules (coded) + Data $\rightarrow$ Output.
- **Machine Learning:** Data + Outputs (labels) $\rightarrow$ Rules (model).
- ML adapts automatically when more data are provided.

$$\text{Programming: } f(x) \to y$$

$$\text{Machine Learning: } (x, y) \to f$$

History of AI / ML: Key Discoveries

Some important landmarks we'll discuss:

- 1837 – First programmable machine (Babbage & Lovelace)
- 1943 – The first formal artificial neuron (McCulloch & Pitts)
- 1950 – Turing's theory of machine intelligence (Turing Test)
- 1957 – The Perceptron (Rosenblatt)
- 1986 – Backpropagation (Rumelhart, Hinton, Williams)
- 2022 – ChatGPT revolutionlizes LLMs
- 2024 – Nobel Prizes recognize AI / ML

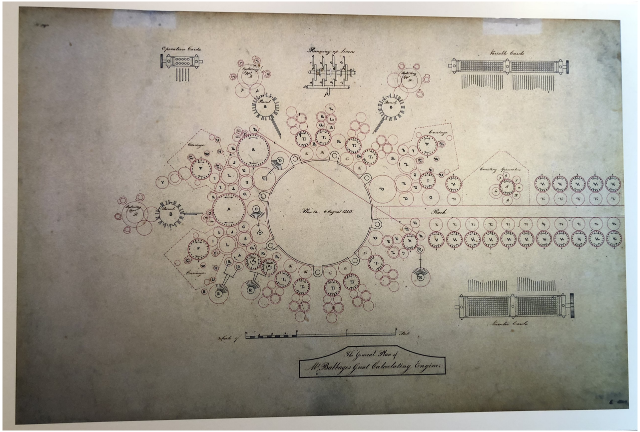
1837: First Programmable Machine

Who: Charles Babbage & Adda Lovelace

Why important:

- Charles developed the first concept of a stored-program mechanical machine with a memory ("store"), an arithmetic unit ("mill"), and control flows (conditional branching / loops).
- This laid the foundational ideas for what a computer is, even though it was never built fully in his lifetime.
- This influenced the theoretical work about computation, programming, and machine architecture.
- Adda created the first algorithm for the machine.

1837: First Programmable Machine



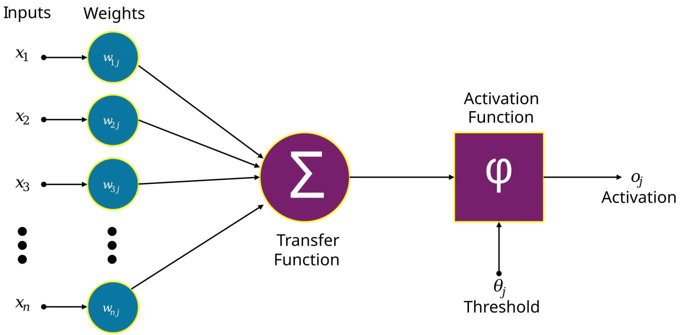
https://en.wikipedia.org/wiki/Analytical_engine

1943: Artificial Neuron Model

By whom: Warren S. McCulloch & Walter Pitts
What / Why:

- Published “A Logical Calculus of the Ideas Immanent in Nervous Activity.”
- Proposed the McCulloch-Pitts neuron: input signals, threshold activation, binary output.
- Showed that networks of such neurons could compute logical functions.
- Foundation for neural networks and later deep learning.

1943: Artificial Neuron Model



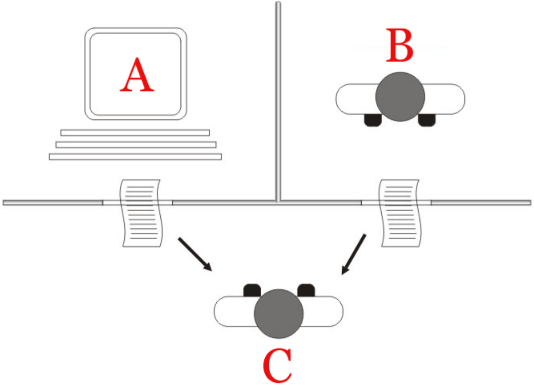
https://en.wikipedia.org/wiki/Artificial_neuron

1950: Turing's Theory of Machine Intelligence

Who: Alan Turing
What / Why:

- Published "Computing Machinery and Intelligence."
- Proposed the "Imitation Game" (later known as the Turing Test).
- Asked whether machines can think; framed rigorous discussion about intelligence.
- Influenced decades of AI research and philosophy.

1950: Turing's Theory



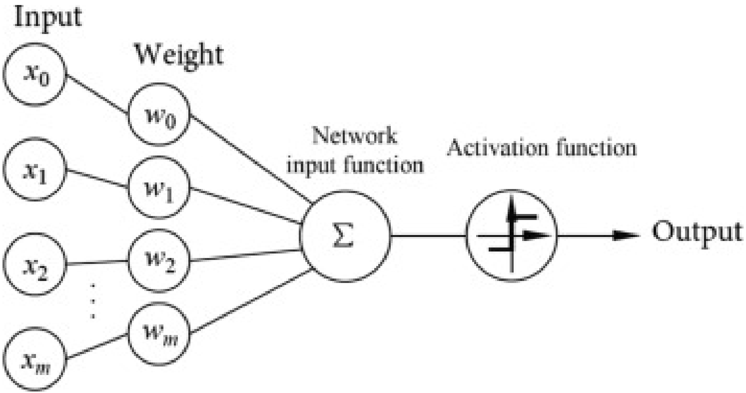
https://en.wikipedia.org/wiki/Turing_test

1957: The Perceptron

Who: Frank Rosenblatt
What / Why:

- Invented the Perceptron, an early neural network model for binary classification.
- Demonstrated learning from data with adjustable weights.
- Limitations: could only solve linearly separable problems.
- Set stage for multilayer networks and modern ML.

1957: The Perceptron



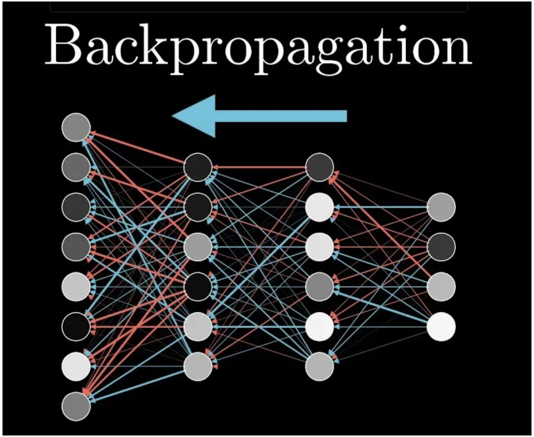
https://www.sciencedirect.com/topics/engineering/perceptron

1986: Backpropagation

Who: David E. Rumelhart, Geoffrey E. Hinton, Ronald J. Williams
What / Why:

- Introduced the backpropagation algorithm for training multilayer networks.
- Enabled hidden layers to learn via gradient descent.
- Crucial for the deep learning revolution (CNNs, RNNs, etc.).

1986: Backpropagation



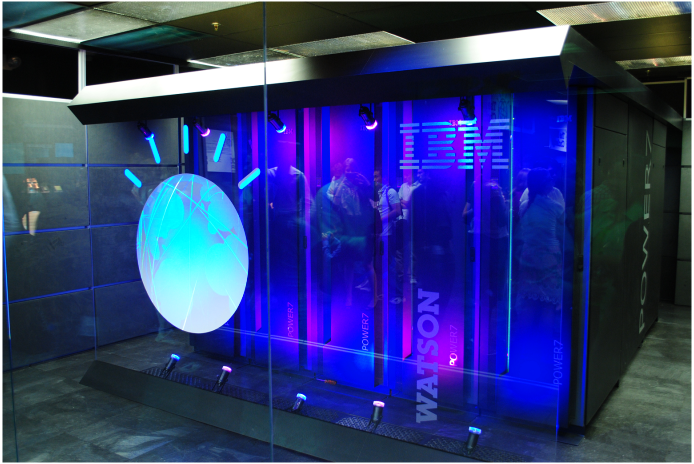
https://medium.com/@sallyrobotics.blog/backpropagation-and-its-alternatives-c09d306aae4c

2011: IBM Watson

Who: IBM (David Ferrucci's team (IBM))
What / Why:

- IBM Watson won the quiz show *Jeopardy!*, beating human champions.
- Combined NLP, information retrieval, statistical learning at scale.
- Shifted public perception of AI capabilities in language and reasoning.

2011: IBM Watson



https://en.wikipedia.org/wiki/IBM_Watson

Some other ML advances

- 1997: Deep Blue defeated Gasparov
- 2011: IBM Watson won in Jeopardy
- 2016: AlphaGo defeated world champion
- 2020: GoogleHealth detecting breastcancer
- 2022: ChatGPT3

2024: Nobel Prizes Recognising AI / ML

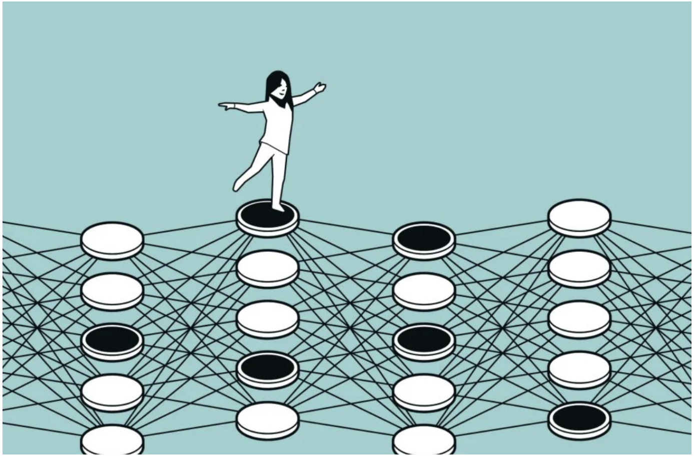
Who:

- Physics: John J. Hopfield & Geoffrey E. Hinton – for foundational discoveries in neural networks.
- Chemistry: David Baker, Demis Hassabis, John M. Jumper – for breakthroughs in protein structure prediction (AlphaFold).

Why important:

- Recognition of AI/ML as central to science.
- Physics: theory and learning in neural networks.
- Chemistry: solved protein structure prediction, impacting medicine and biology.

2024: Nobel Prizes



https://www.nobelprize.org/all-nobel-prizes-2024/

Real world applications

- Structure Prediction: Predicting Protein Structures with AlphaFold
- Image recognition: Deep Learning for Radiology and Pathology
- Large Language Models: Medical question answering with LLMs
- Drug Discovery: A Deep Learning Approach to Antibiotic Discovery

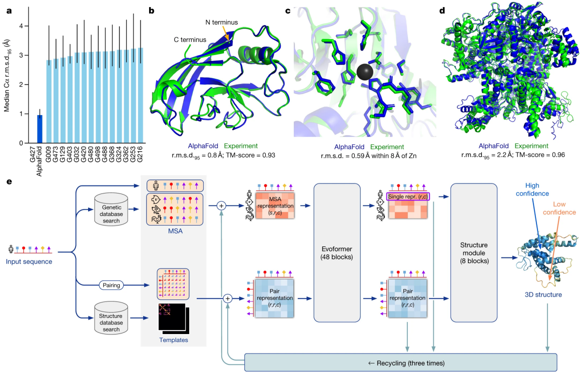
Predicting Protein Structures with AlphaFold

Mathematical Institute

Jumper, John M. et al. "Highly accurate protein structure prediction with AlphaFold." Nature 596 (2021): 583 - 589.

- Unprecedented accuracy in protein structure prediction
- Novel neural network architecture integrating evolutionary and geometric features
- Massive coverage of the protein universe via open-access database

Predicting Protein Structures with AlphaFold



Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function.

Predicting the 3D structure of a protein based on its amino acid sequence has been an important open research problem.

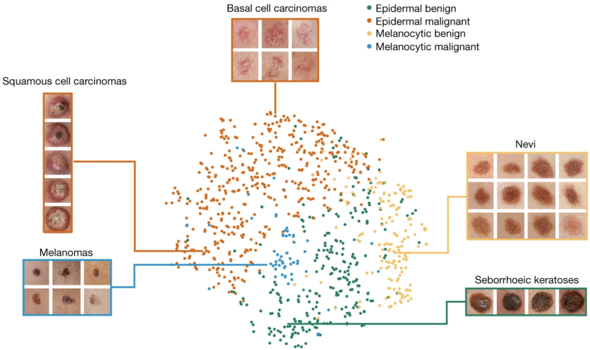
AlphaFold produces highly accurate structures.

Deep Learning for Radiology and Pathology

Esteva, Andre et al. "Dermatologist-level classification of skin cancer with deep neural networks." Nature 542 (2017): 115-118.

- A deep convolutional neural network matched the diagnostic performance of board-certified dermatologists.
- The model was trained on an exceptionally large and varied dataset.
- End-to-end training using only pixel data and labels.
- Potential for widespread deployment, including mobile devices.

Deep Learning for Radiology and Pathology



Skin cancer, the most common human malignancy is primarily diagnosed visually.

The full taxonomy contains 2,032 diseases and is organized based on visual and clinical similarity of diseases.

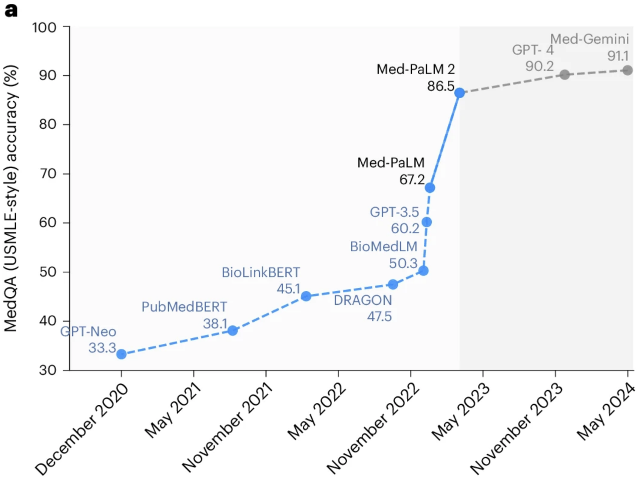
The CNN achieves performance equivalent to experts tested across several diagnostic tasks.

Medical question answering with LLMs

Singhal, Karan et al. "Toward expert-level medical question answering with large language models." Nature Medicine 31 (2025): 943 - 950.

- Dramatic improvement in accuracy on medical benchmarks.
- In pairwise evaluations on nine clinically relevant dimensions, physicians preferred Med-PaLM2 answers over other physicians.
- The model answers were judged to be as safe as the one generated by the physician, highlighting its potential as a practical support tool in healthcare settings.

Medical question answering with LLMs

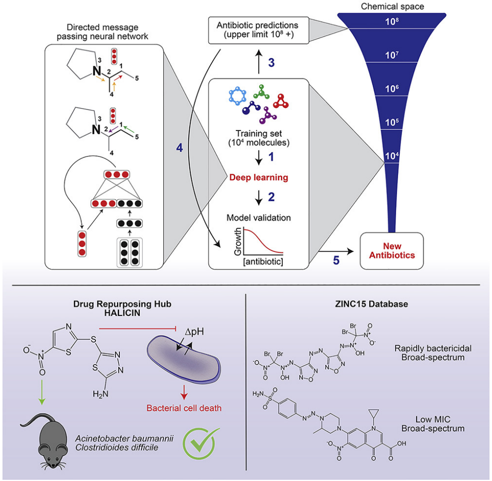


A Deep Learning Approach to Antibiotic Discovery

Stokes, Jonathan M. et al. "A Deep Learning Approach to Antibiotic Discovery." Cell 181 (2020): 475-483.

- A deep learning model is trained to predict antibiotics based on structure
- Halicin is predicted as an antibacterial molecule from the Drug Repurposing Hub
- Halicin shows broad-spectrum antibiotic activities in mice
- More antibiotics with distinct structures are predicted from the ZINC15 database

A Deep Learning Approach to Antibiotic Discovery



Due to **antibiotic-resistant bacteria**, we need to discover new antibiotics.

Halicin was identified as structurally divergent from conventional antibiotics and could be used against a wide phylogenetic spectrum of pathogens.

This work highlights the **utility of deep learning** approaches to expand our antibiotic arsenal.

Course overview

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Accuracy measures
- Practical Applications
- Neural Networks
- Transformers
- Interesting papers

Grades will be based on projects

Project topic: Select a dataset! Apply what you learned through the course.
Critical thinking is important!
Project deadline: January 19

Questions for the practical part

- How many use Python?
- Are you familiar with Python notebooks?
- Which operating system are you using?

Thank you! Questions?