



University of  
**Nottingham**  
UK | CHINA | MALAYSIA

# COMP3055

# Machine Learning

## Topic 1 – Introduction

Ying Weng  
2024 Autumn

# COMP3055 Machine Learning

## Ying WENG

**Teaching for 6 weeks**

Office: PMB-341

Office hour: Monday 14:00 – 16:00

Email: [ying.weng@nottingham.edu.cn](mailto:ying.weng@nottingham.edu.cn)

## Zheng LU

**Teaching for 6 weeks (Module Convenor)**

Office: PMB-426

Office hour: TBC

Email: [zheng.lu@nottingham.edu.cn](mailto:zheng.lu@nottingham.edu.cn)

- Lecture
- Lab
  - Tuesday 16:00-18:00, IAMET-406
  - For the 1<sup>st</sup> Teaching Week, there is no lab on Tuesday 24<sup>th</sup> Sep

# Lecture Schedule

- Plan ahead
- All material on Moodle
- Regular update

# Lab Schedule

- Lab sessions will provide you hands on experiences what you learned during the lecture sessions
- Lab materials will be published on Moodle

# Summary of Content

- Provide you with an introduction to **machine learning**, **pattern recognition**, and **data mining** techniques
- Enable you to consider both systems which are able to develop their own rules from trial-and-error experience to solve problems
- Find patterns in data without any supervision. Data mining techniques will make generation of new knowledge possible, including big datasets. This is now fashionably termed 'Big Data' science
- Cover a range of topics including: machine learning foundations; pattern recognition foundations; applications of machine learning; data mining techniques and evaluating hypotheses; artificial neural networks; deep learning
- You'll spend **six hours** each week (on average) **in lectures and computer practice** for this module

# Assessments

## ❖ Exam: 70%

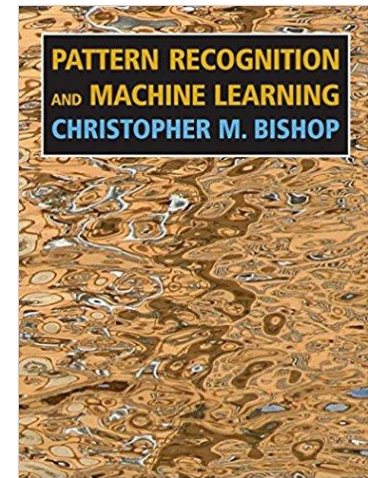
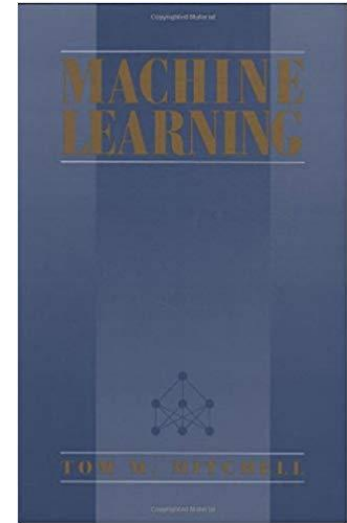
- 2 hour written examination

## ❖ Coursework: 30%

- 1 piece of individual programming assignment

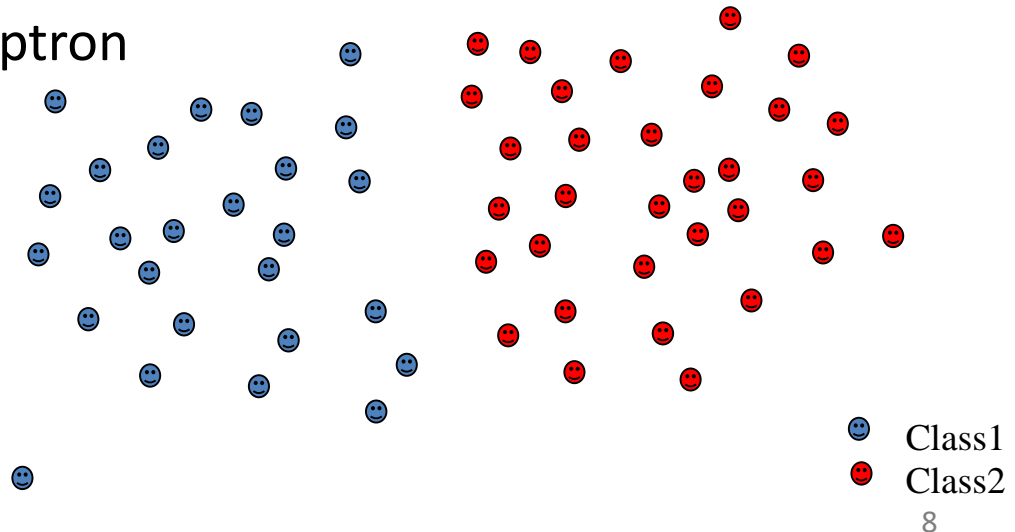
# Textbooks

- Tom M. Mitchell, **Machine Learning**, McGraw-Hill
- Chris Bishop, **Pattern Recognition and Machine Learning**, Springer
- Some of recent technical papers



# Topics

- Design learning systems
- Data collection
- Learning theory practice
- Instance based learning, KNN etc.
- Bayesian learning
- Data processing representation
- Unsupervised learning, clustering, K-Means etc.
- Decision tree, random forest
- Perceptron, multilayer perceptron
- Support vector machine
- Deep learning, CNN etc.
- ...
- (to be continued)





# How to Get 70+

- **Studying...**

- You are recommended to study the relevant notes before attending the lecture or lab.
- Review as soon as possible to maximize retention.



- **Practice...**

- **Do the lab exercise yourself** and repeat the practice for better learning.
- If you get help on the labs, don't just blindly accept it, but try to understand what each part of the code is doing.
- Do the Math in the lecture for better understanding.

- **Assignments...**

- Start work on the assignment **when they are released**, and come up with a good plan to finish it.
- It will take **longer** than you expect to fix problems in your program, so make sure you have plenty of time to complete.

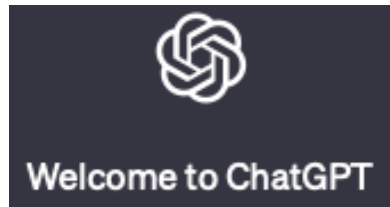
Do the thing right, Do the right thing



# A Few Quotes

- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, Director, DARPA)
- Machine learning is the hot new thing” (John Hennessy, President, Stanford)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)
- “Machine learning is today’s discontinuity” (Jerry Yang, CEO, Yahoo)

# ChatGPT

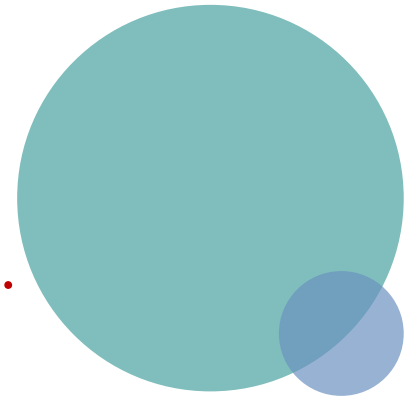


- ❖ An artificial intelligence (AI) chatbot developed by OpenAI and release in November 2022
- ❖ Built on top of GPT-4 foundational large language model (LLM)
- ❖ Fine-tuned to transfer learning

incredible, powerful, helpful

ChatGPT has rapidly immersed itself into our lives.  
I think everyone has read articles related to it.

ChatGPT is a milestone in the era of large models.



# What is Machine Learning?

- It is very hard to write programs that solve problems like recognizing a face.
  - We don't know what program to write because we don't know how our brain does it.
  - Even if we had a good idea about how to do it, the program might be horrendously complicated.
- Instead of writing a program by hand, we collect lots of examples that specify the correct output for a given input.
- A machine learning algorithm then takes these examples and produces a program that does the job.
  - The program produced by the learning algorithm may look very different from a typical hand-written program. It may contain millions of numbers.
  - If we do it right, the program works for new cases as well as the ones we trained it on.

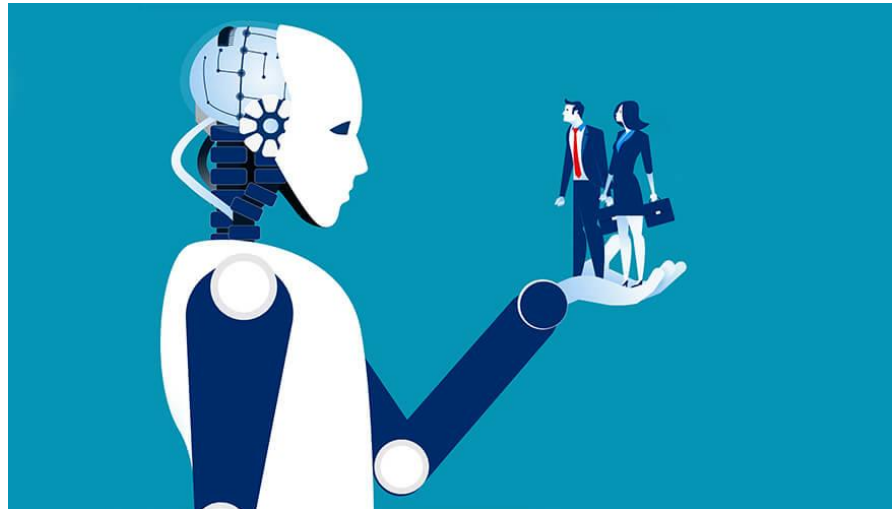
# What is Machine Learning?

- Automating automation
- Getting computers to program themselves
- Writing software is the bottleneck
- Let the data do the work instead!

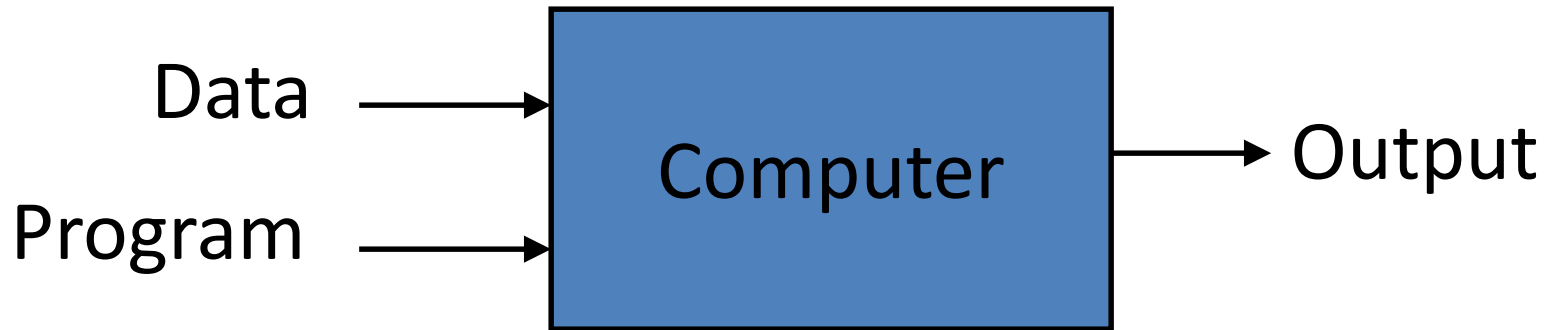


# What is Machine Learning?

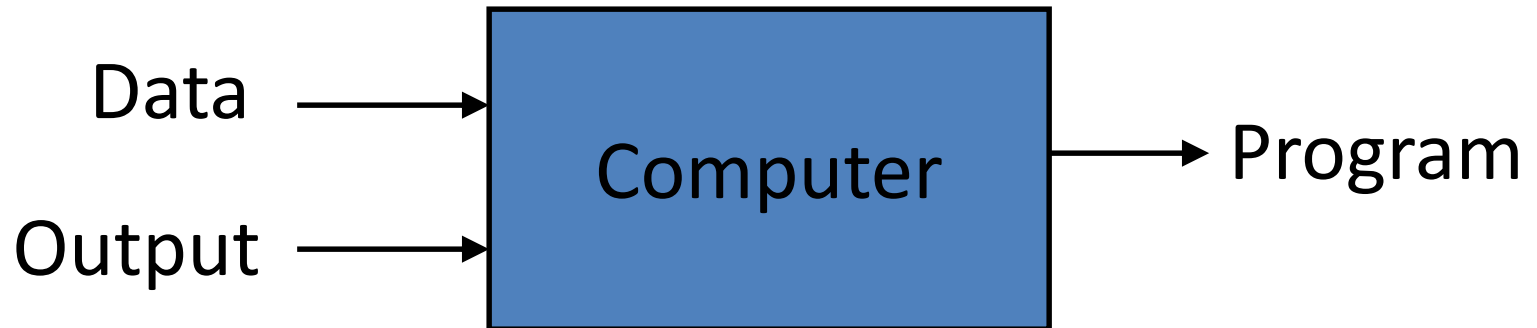
- Machine Learning is making great strides
  - Large, good data sets
  - Computer power
  - Progress in algorithms
- Many interesting applications
  - Commercial
  - Scientific



## Traditional Programming



## Machine Learning





# Magic?

No, more like gardening

- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs



# Task: Handwriting Recognition

It is very hard to say what makes a 2

0 0 0 1 1 1 1 1 1 2

2 2 2 2 2 2 2 3 3 3

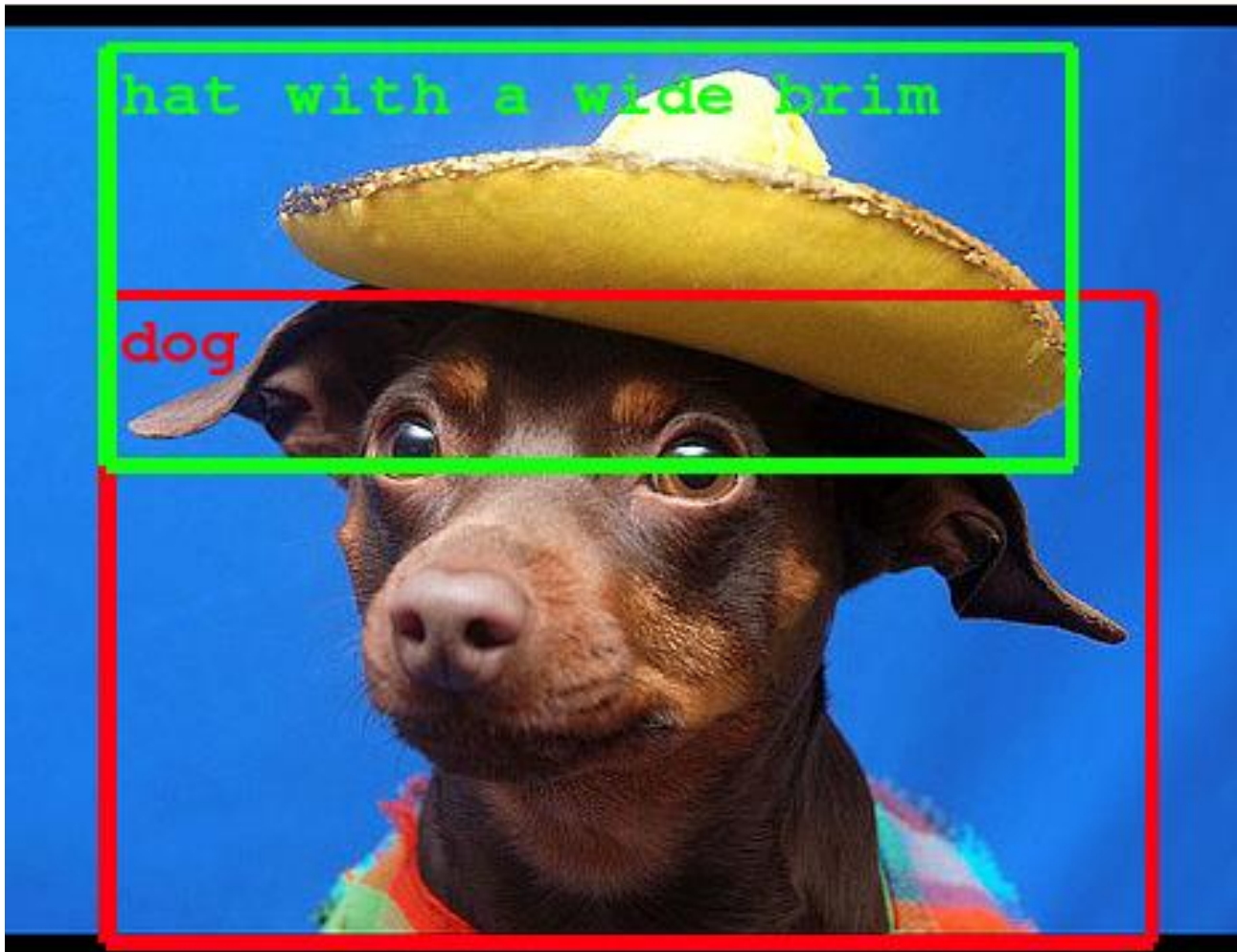
3 4 4 4 4 4 5 5 5 5

6 6 7 7 7 7 7 8 8 8

9 9 9 9 9 9 9 9 9

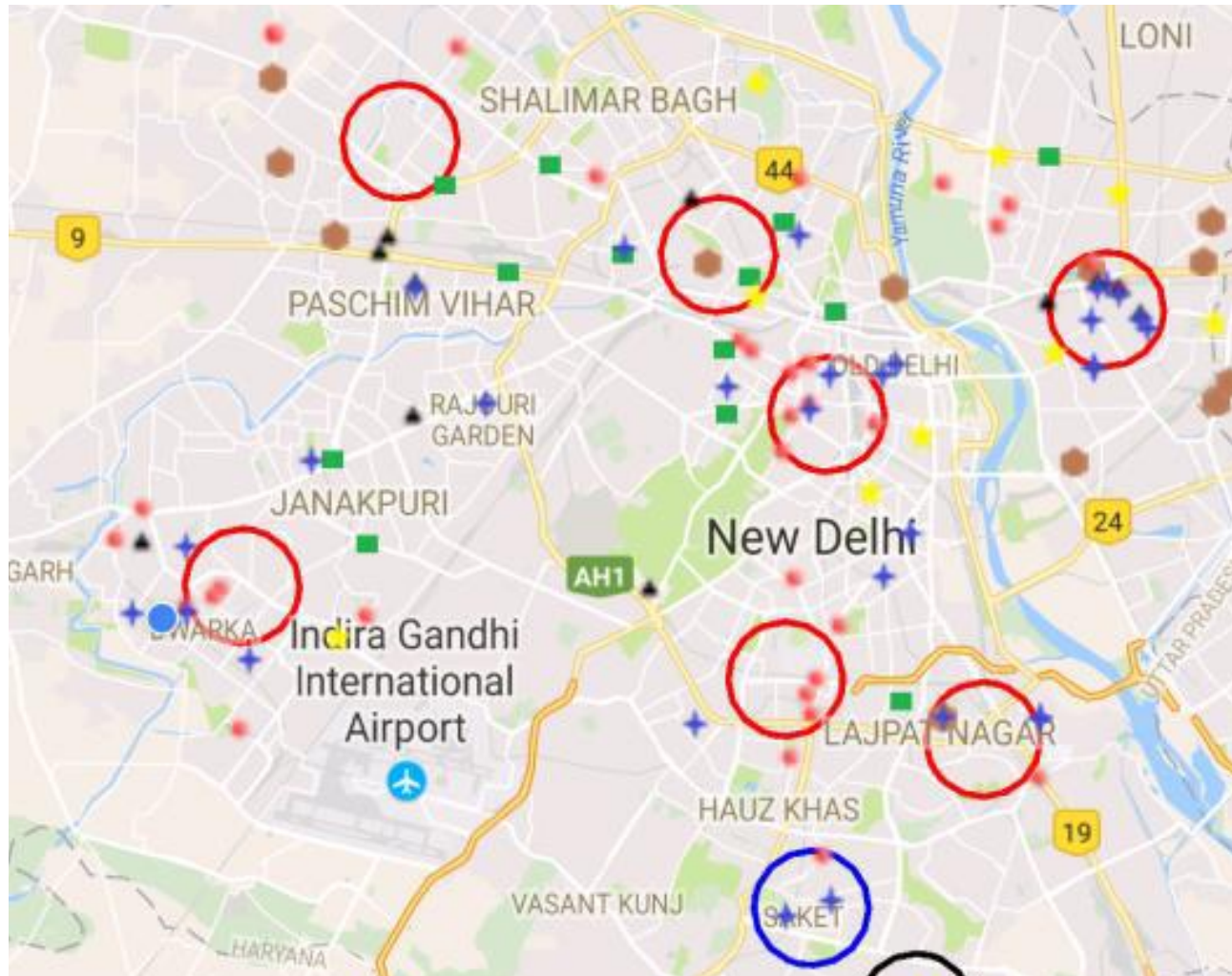
# Task: Classification

## Object recognition



# Task: Clustering

## Crime prediction using k-means clustering

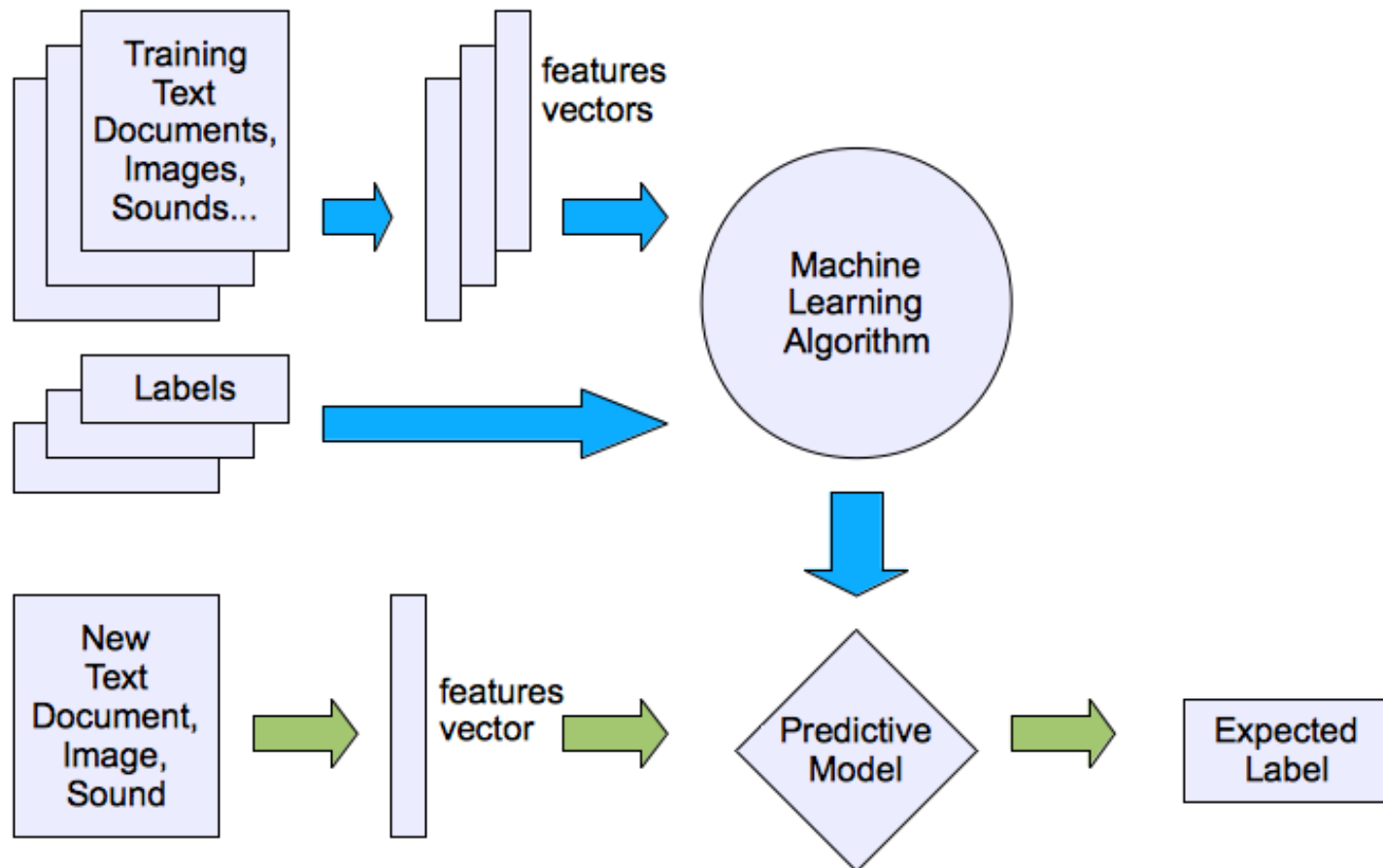


# More Sample Applications

- Web search
- Healthcare
- Robotics
- Information extraction
- Finance
- E-commerce
- Computational biology
- Space exploration
- Social networks
- [Your favorite area]

# Types of Learning

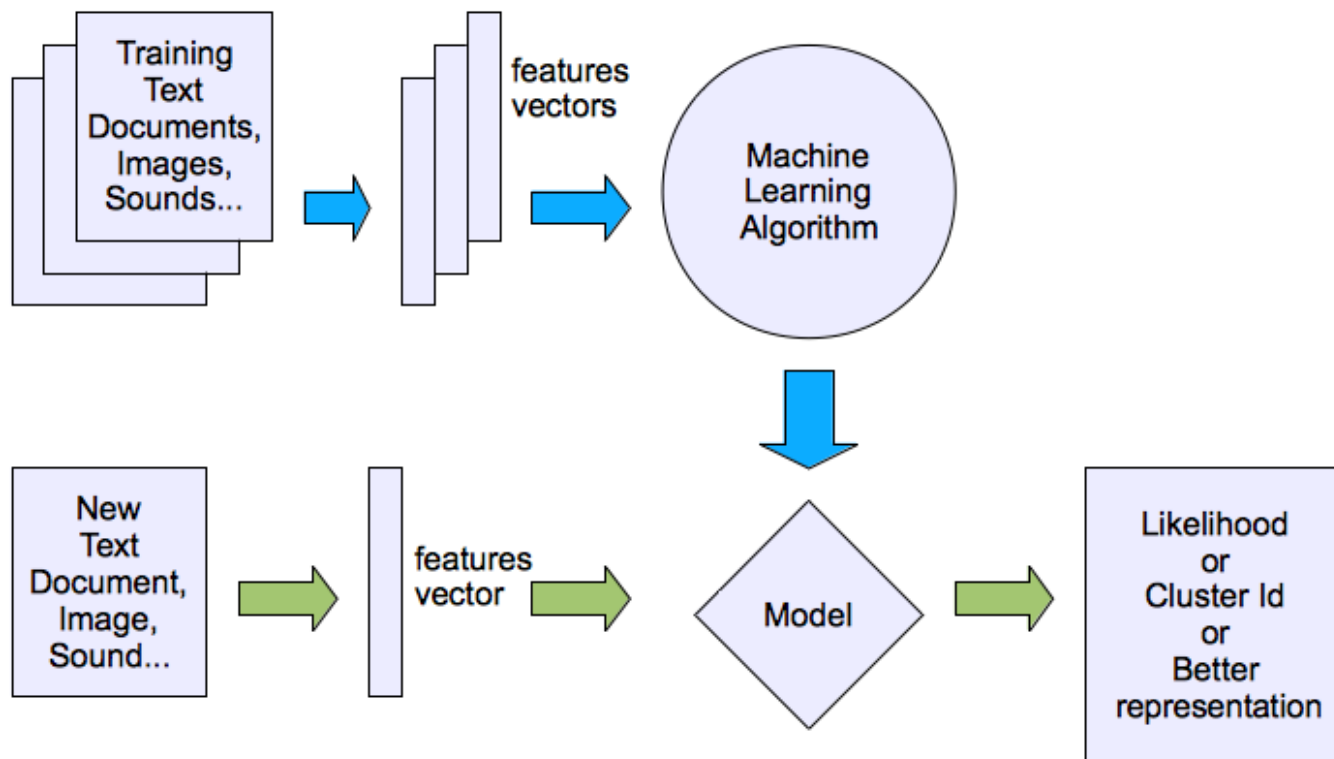
- Supervised learning





# Types of Learning

- Unsupervised learning



# Types of Learning

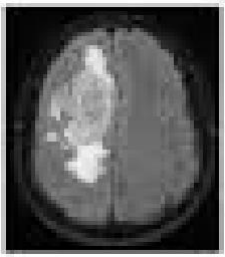
- **Supervised (inductive) learning**
  - Training data includes desired outputs
- **Unsupervised learning**
  - Training data does not include desired outputs
  - This is the new frontier of machine learning because most big datasets do not come with labels
- **Semi-supervised learning**
  - Training data includes a few desired outputs



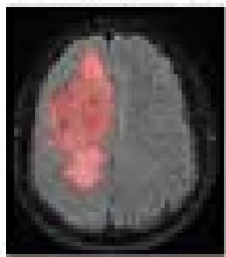
# Semi-supervised Learning

Motivation: labeled data are expensive

unlabeled data are free



unlabeled



labeled

Medical field:

- The process of labeling medical images requires **specialized expertise** and is often **time-consuming** and **expensive**.
- This is due to the **complexity** of medical images, the need for **high-quality annotations**, and **the limited availability of trained medical professionals**.

# Framework

- **Programming languages**
  - Python
  - C++
  - ...
- **Many libraries**
  - PyTorch
  - TensorFlow
  - ...

# Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Entropy
- etc.

# Issues

- Many machine learning/AI projects fail (Gartner claims 85%)



- Ethics



# Reasons for Failure

- Asking the wrong question
- Trying to solve the wrong problem
- Not having enough data
- Not having the right data
- Hiring the wrong people
- Using the wrong tools
- Not having the right model
- Not having the right yardstick



# Any Questions?

