01.112/50.007 Machine Learning

Lecture 1

Introduction

Who are we?

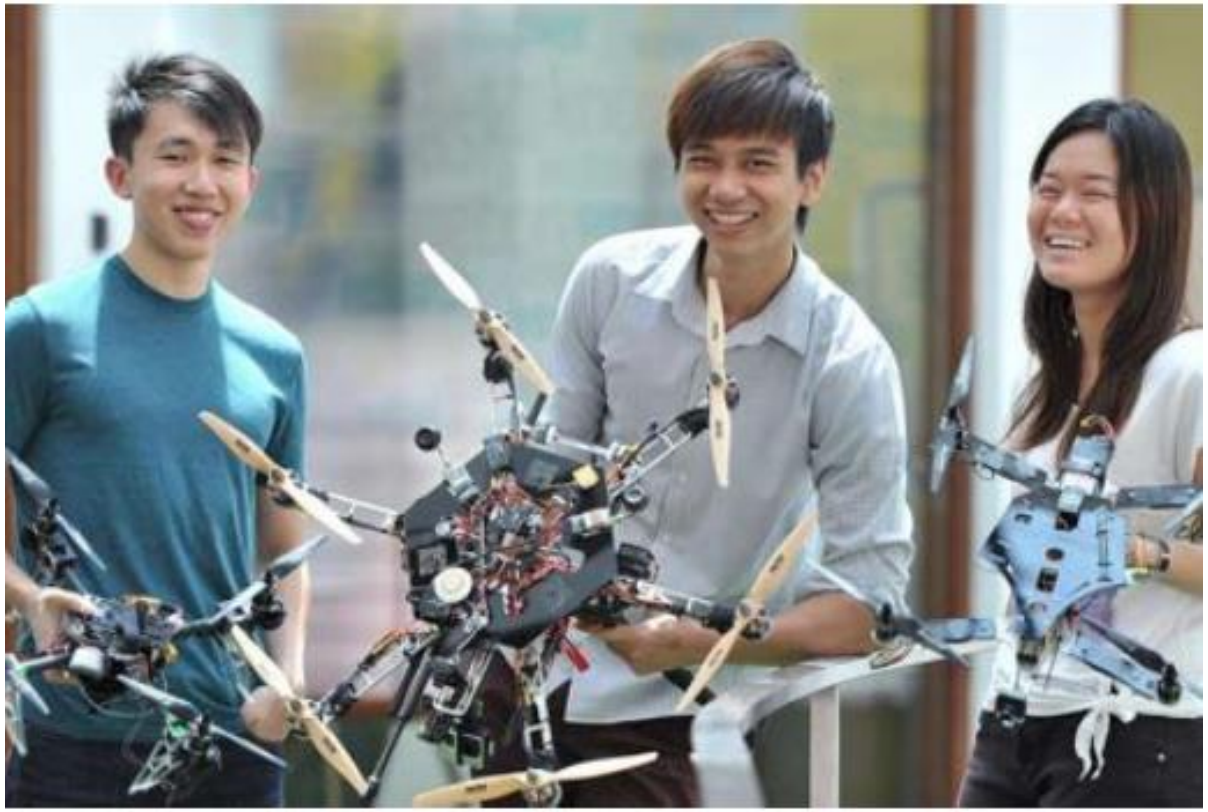
Prof. Malika Meghjani
Instructor
Weeks 1-4

Prof. Wei Lu
Instructor
Weeks 5-14

Teaching Assistants

- Chen Zihan, Fri 14:00-15:00, 2.716-S03
- Sun Xiaobing, Fri 14:00-15:00, 1.417
- Joel Ong, Fri 17:00-18:00, 1.417

Who are you?



Outline

- Administrative details
- What is machine learning?
- Types of machine learning
- A case study for supervised learning
- Linear Classification

Administrative details

- **Class materials:**
  - No required textbook
  - Recommended reading (from books or research papers) posted on eDimension
  - Class notes: posed on eDimension

- **Pre-requisite:**
  - Linear Algebra
  - Probability/statistics
  - Knowledge of Algorithms
  - Python Programming

Evaluation

- Homework (28%)
    - Programming and theory
    - Honour Code
        - Form study groups to work on homework
        - You can discuss with other classmates
        - Write-up solutions on your own
        - List names of anyone you talked to
- Project (20%)
- Midterm Exam (25%)
- Final Exam (25%)
- Participation (2%)

Course Goals

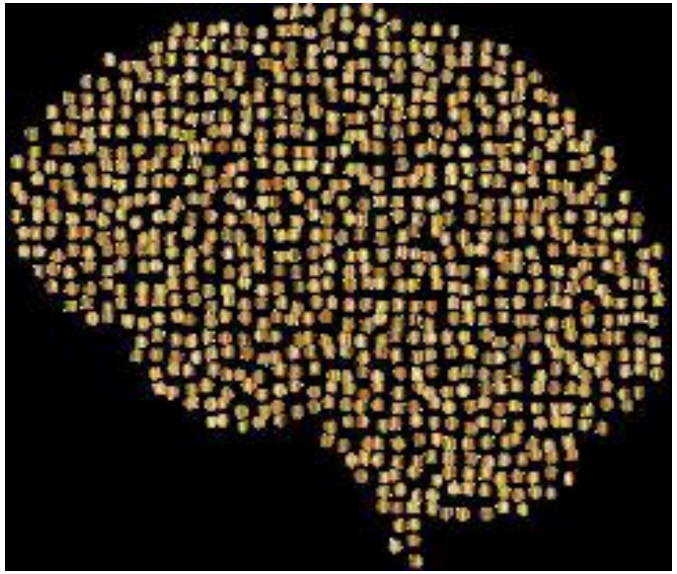
- Curious to discover more
- Confident of doing it yourself
- Contemplative of the theory
- Cautious of the danger

Acknowledgement

- MIT 6.036 Introduction to Machine Learning
- SUTD 50.007 Machine Learning (Prof. Liang Zheng)
- Stanford CS229 Machine Learning
- McGill COMP-652 Machine Learning

What is Machine Learning?

Hardcoded



Trained

- Giving computers the **ability to learn** without being explicitly programmed – Arthur Samuel (1959)

What is Machine Learning?

Task

Performance

Experience

- Algorithms that improve their **performance** at some **task** with **experience** – Tom Mitchell (1998)

What is Machine Learning?

- A branch of **artificial intelligence**, concerned with the design and development of algorithms that allow computers to evolve behaviours based on empirical data.

- As intelligence requires knowledge, it is necessary for the computers to first **acquire knowledge through learning**.

- Machine learning is programming computers to optimize a **performance criterion** using example data or **past experience**.

Why study Machine Learning?

Engineering reasons:

- Easier to build a learning system than to hand-code a working program!
  - Robot that learns a map of the environment by exploring
  - Programs that learn to play games by playing against themselves

- Improving on existing programs,
  - Instruction scheduling and register allocation in compilers
  - Combinatorial optimization problems

- Solving tasks that require a system to be adaptive
  - Speech and handwriting recognition
  - “Intelligent” user interfaces

Why study Machine Learning?

Scientific reasons:

- Discover knowledge and patterns in highly dimensional, complex data
  - Sky surveys
  - High-energy physics data
  - Sequence analysis in bioinformatics
  - Social network analysis
  - Ecosystem analysis
- Understanding animal and human learning
  - How do we learn language?
  - How do we recognize faces?
- Creating real AI!

  "If an expert system–brilliantly designed, engineered and implemented– cannot learn not to repeat its mistakes, it is not as intelligent as a worm or a sea anemone or a kitten." (Oliver Selfridge).

Very brief history

- Studied ever since computers were invented (e.g. Samuel’s checkers player)
- Very active in 1960s (neural networks)
- Died down in the 1970s
- Revival in early 1980s (decision trees, backpropagation, temporal difference learning) - coined as “machine learning”
- Exploded since the 1990s
- Now: very active research field, several yearly conferences (e.g., ICML, NIPS), major journals (e.g., Machine Learning, Journal of Machine Learning Research), rapidly growing number of researchers
- The time is right to study in the field!
  - Lots of recent progress in algorithms and theory
  - Flood of data to be analyzed
  - Computational power is available
  - Growing demand for industrial applications

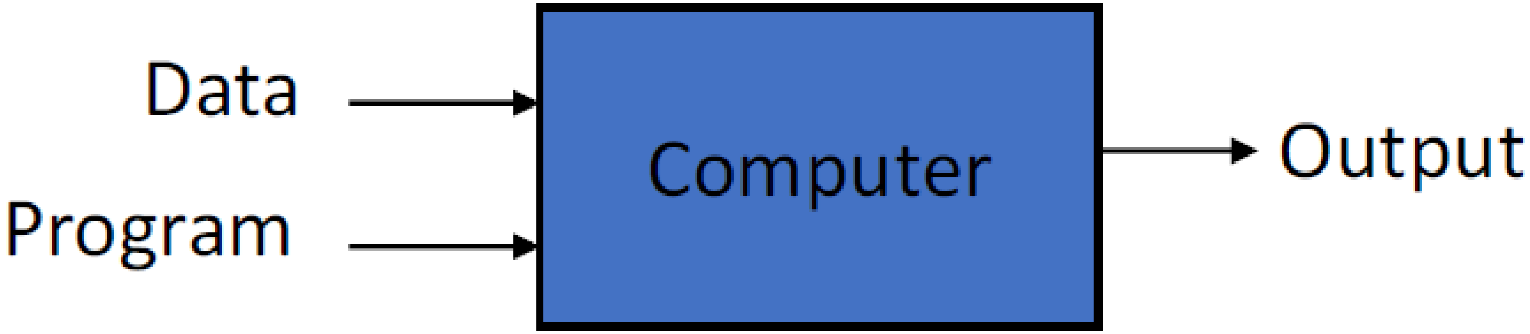
What are good Machine Learning tasks?

- There is no human expert
  E.g., DNA analysis
- Humans can perform the task but cannot explain how
  E.g., character recognition
- Desired function changes frequently
  E.g., predicting stock prices based on recent trading data
- Each user needs a customized function
  E.g., news filtering

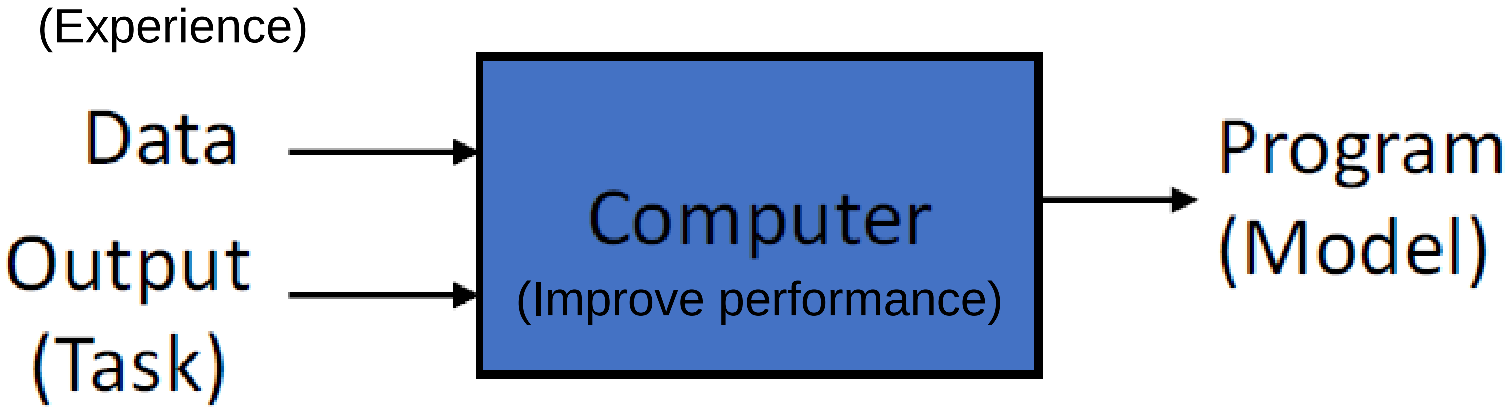
Important application areas

- **Bioinformatics:** sequence alignment, analyzing microarray data, information integration, ...
- **Computer vision:** object recognition, tracking, segmentation, active vision, …
- **Robotics:** state estimation, map building, decision making
- **Graphics:** building realistic simulations
- **Speech:** recognition, speaker identification
- **Financial analysis:** option pricing, portfolio allocation
- **E-commerce:** automated trading agents, data mining, spam, ...
- **Medicine:** diagnosis, treatment, drug design,...
- **Computer games:** building adaptive opponents
- **Multimedia:** retrieval across diverse databases

Traditional Programming

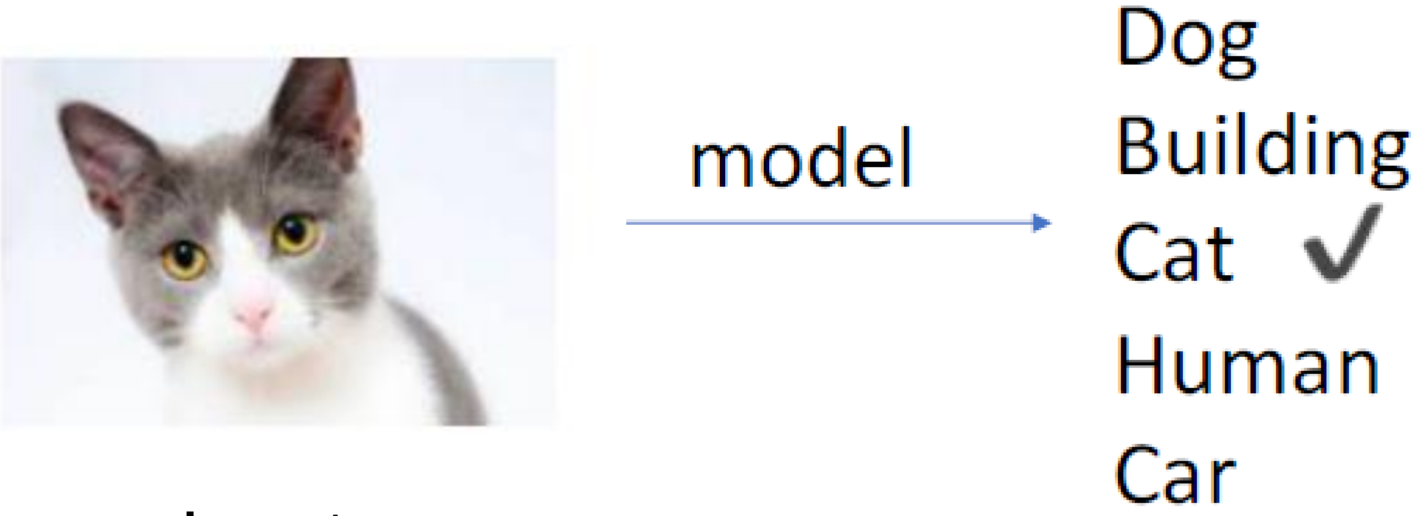


Machine Learning



What is Machine Learning?

- We have a model which we can use to predict new data
- E.g. Image classification

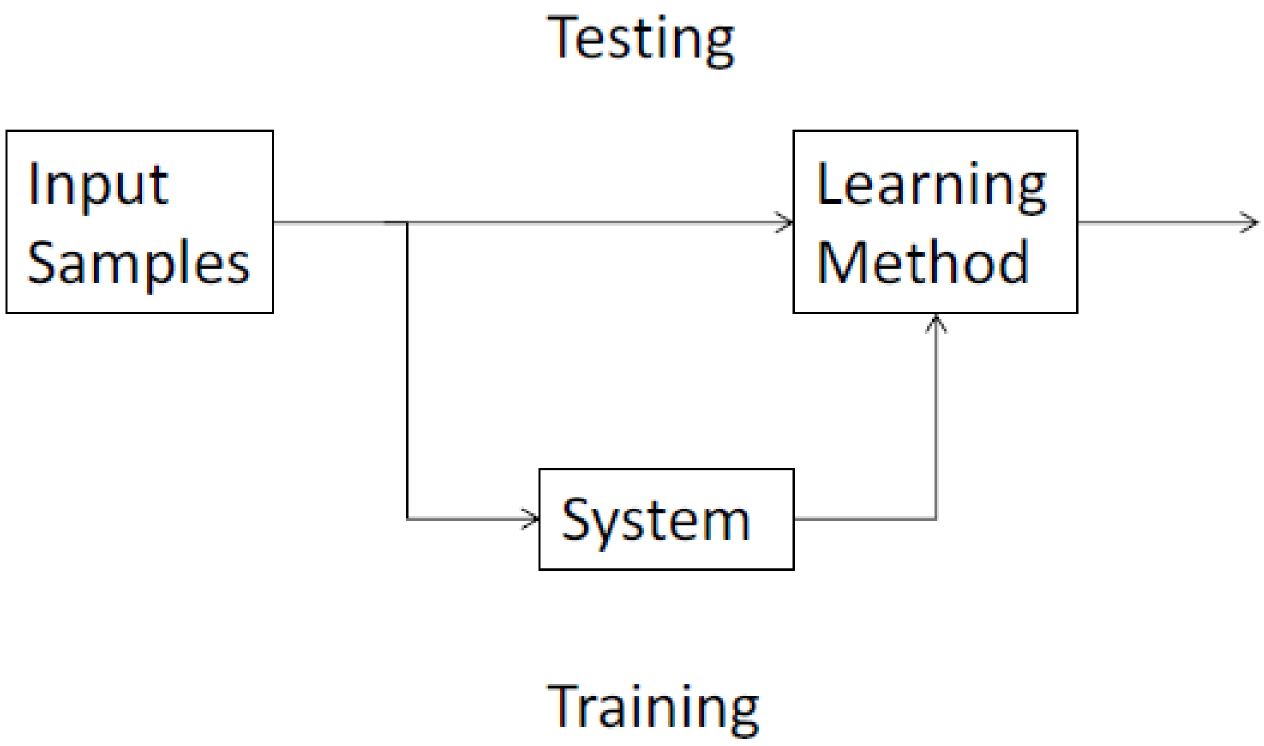


Input

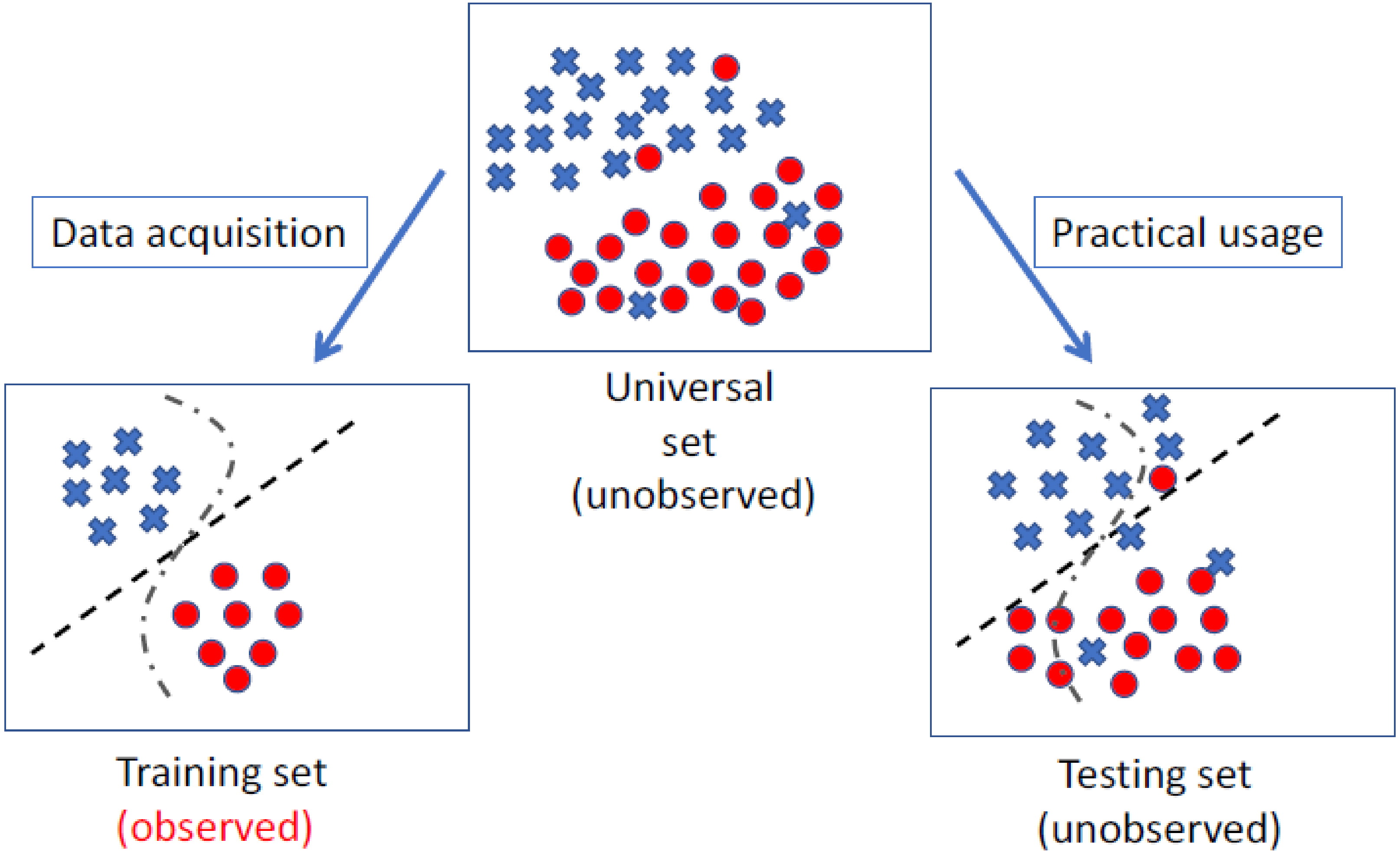
What is Machine Learning?

- Learning general models from a data of particular examples
- Data is cheap (?) and abundant (data warehouses, data mars); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behaviour:
  - *People who bought "Da Vinci Code" also bought "The Five People You Meet in Heaven" (www.amazon.com)*
- Build a model that is *a good and useful approximation* to the data.

Learning system model

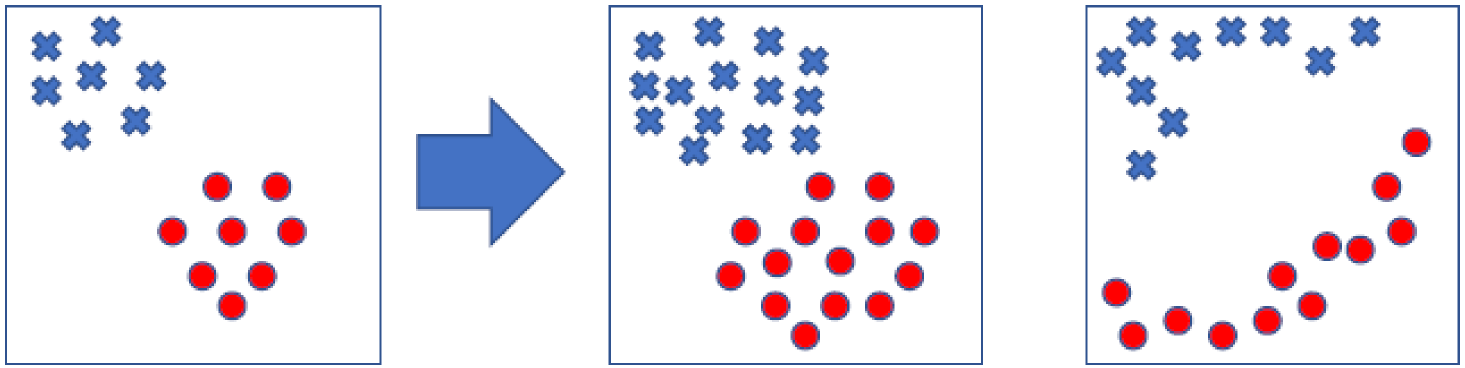


Training and Testing



Training and Testing

- Training is the process of making system able to learn.
- No free lunch rule:
  - Training set and testing set come from the same distribution
  - Need to make some assumption or bias



Training data Testing data

Performance

- There are several factors affecting the performance:
  - **Quality of training data** provided
  - The form and extent of any initial **background knowledge**
  - The **type of feedback** provided
  - The **learning algorithm** used

- Two important factors:
  - Modelling
  - Optimization

Algorithms

- The success of machine learning system also depends on the algorithms.
- The algorithms control the search to find and build the knowledge structures.
- The learning algorithms should extract useful information from training examples.

Types of Machine Learning

Based on information available

- **Supervised learning ($\{x_n \in R^d, y_n \in R\}_{n=1}^N$)**
  - Classification (discrete labels)
  - Regression (real values)

- **Unsupervised learning ($\{x_n \in R^d\}_{n=1}^N$)**
  - Clustering
  - Probability distribution estimation
  - Finding association (in features)
  - Dimension reduction

- **Reinforcement learning**
  - Decision making (robotics, board games)

- **Semi-supervised learning**

Based on learner's role

- **Passive learning**
- **Active learning**

Passive and active learning

- Traditionally, learning algorithms have been **passive learners**, which take a given batch of data and process it to produce a hypothesis or model.

  Data → Learner → Model

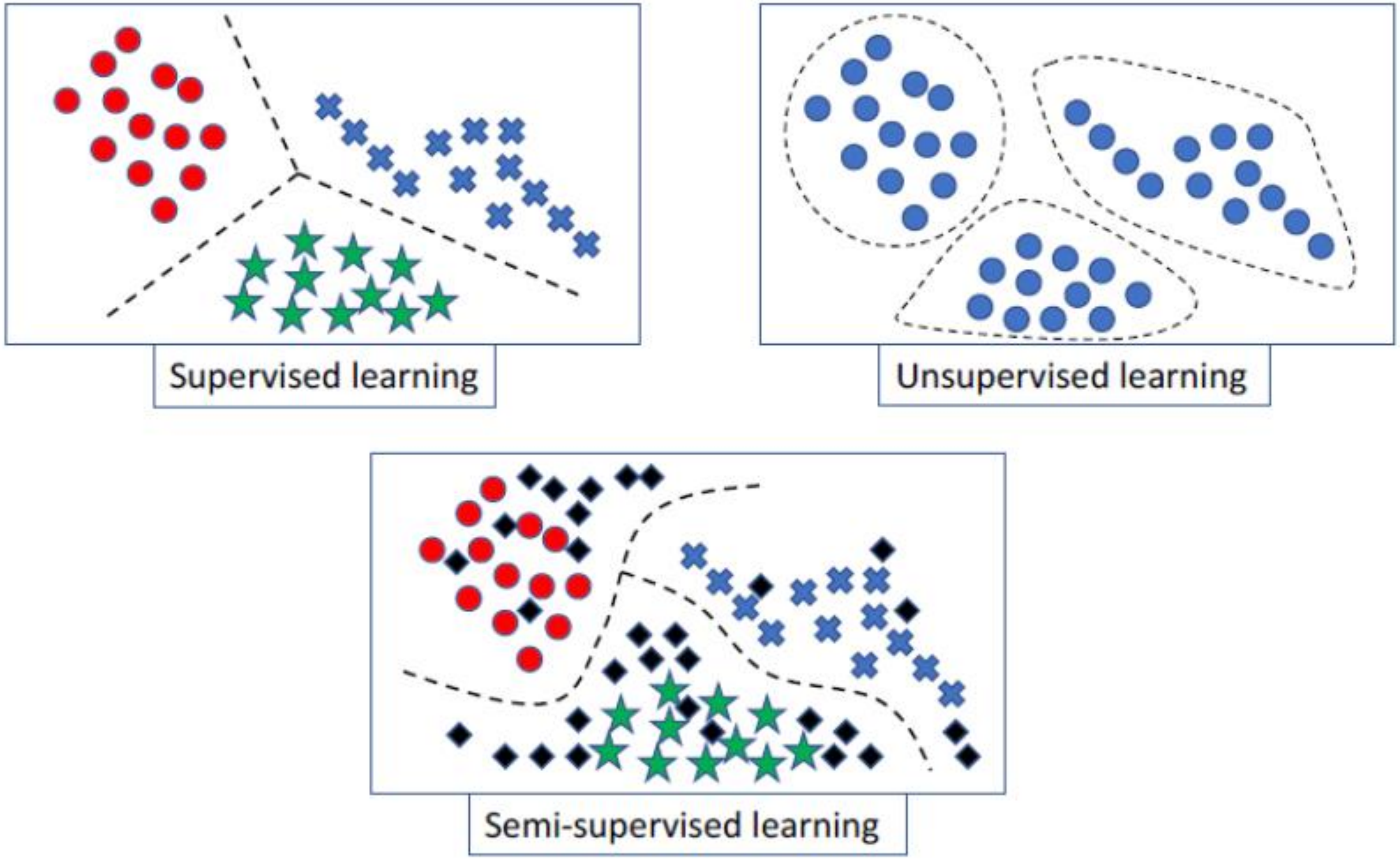
- **Active learners** are instead allowed to query the environment

– Ask questions

– Perform experiments

- Open issues: how to query the environment optimally? how to account for the cost of queries?

Types of Machine Learning

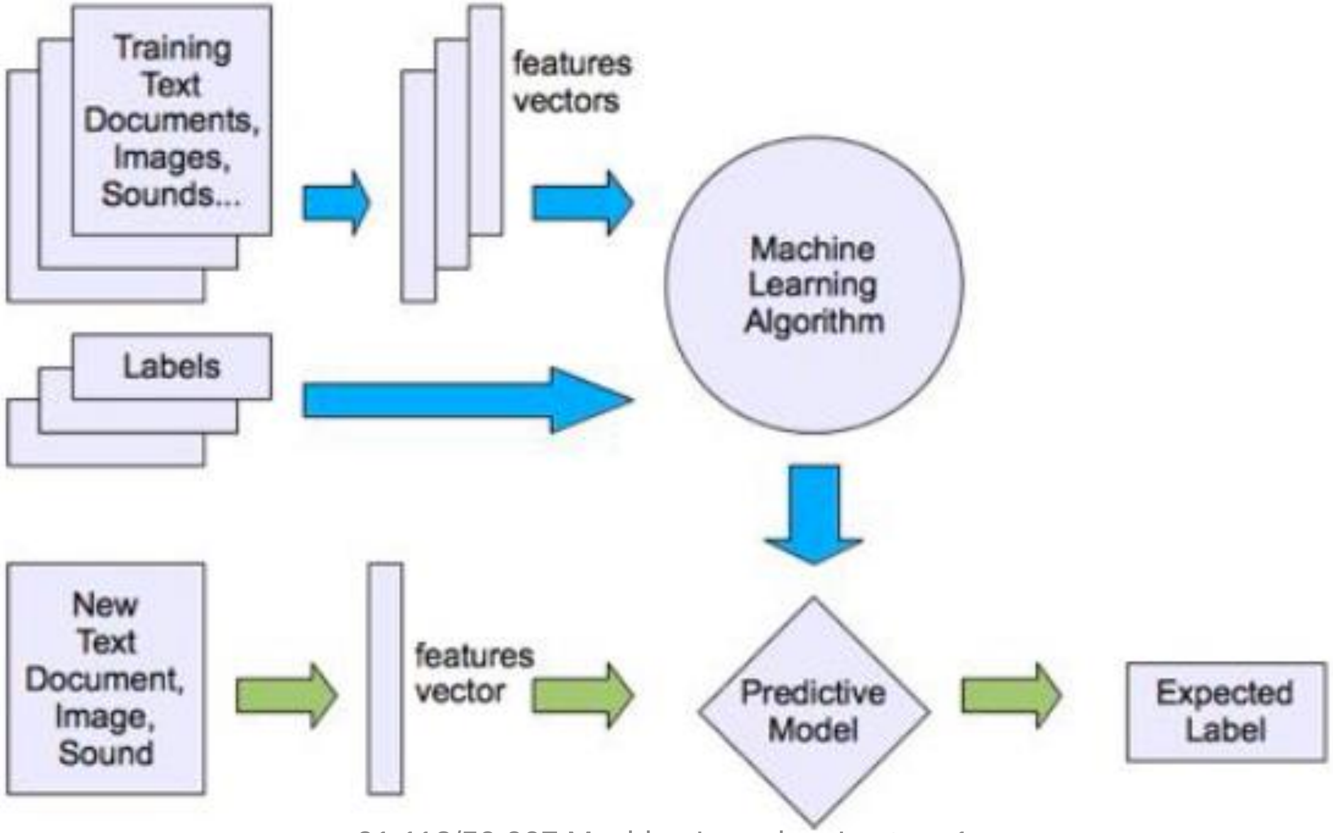


Types of Machine Learning

- **Supervised Learning**

Types of Machine Learning

- **Supervised Learning**

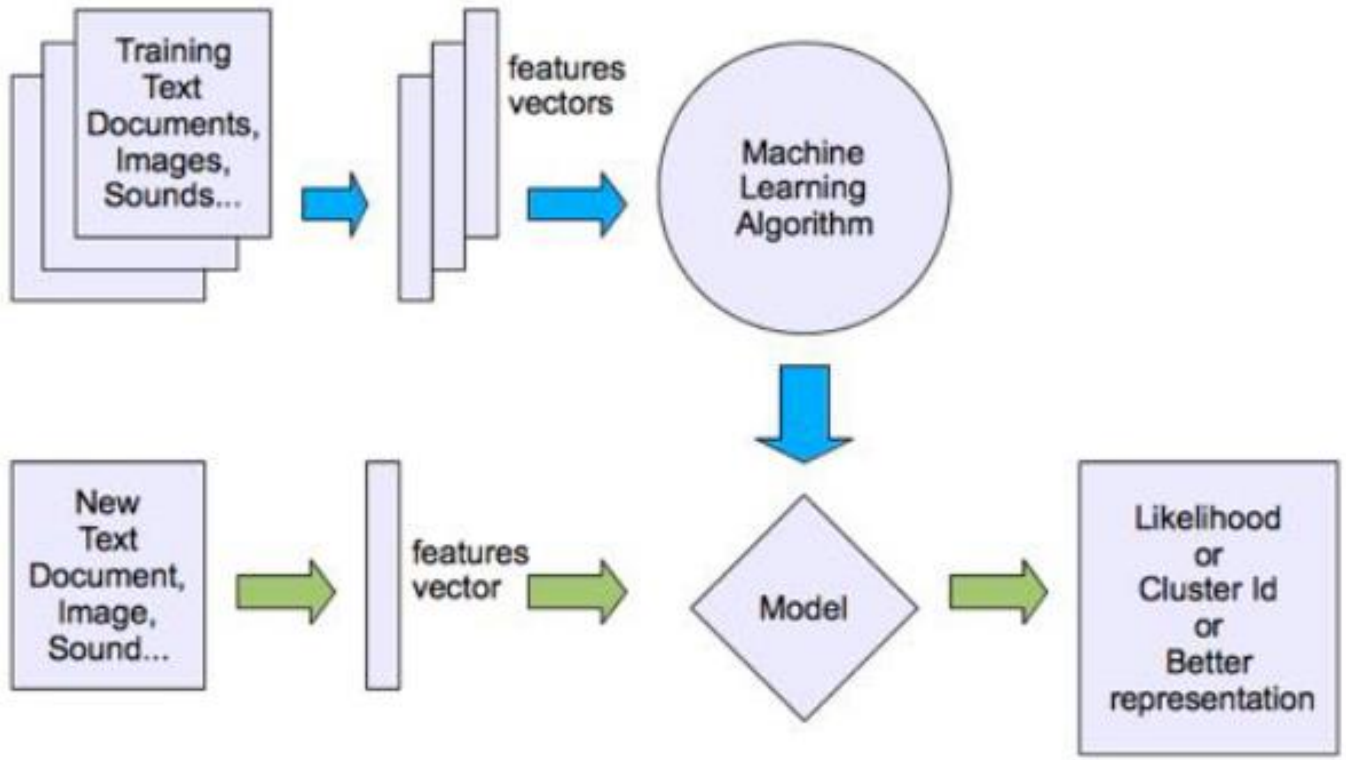


Types of Machine Learning

- **Unsupervised Learning**

Types of Machine Learning

- **Unsupervised Learning**

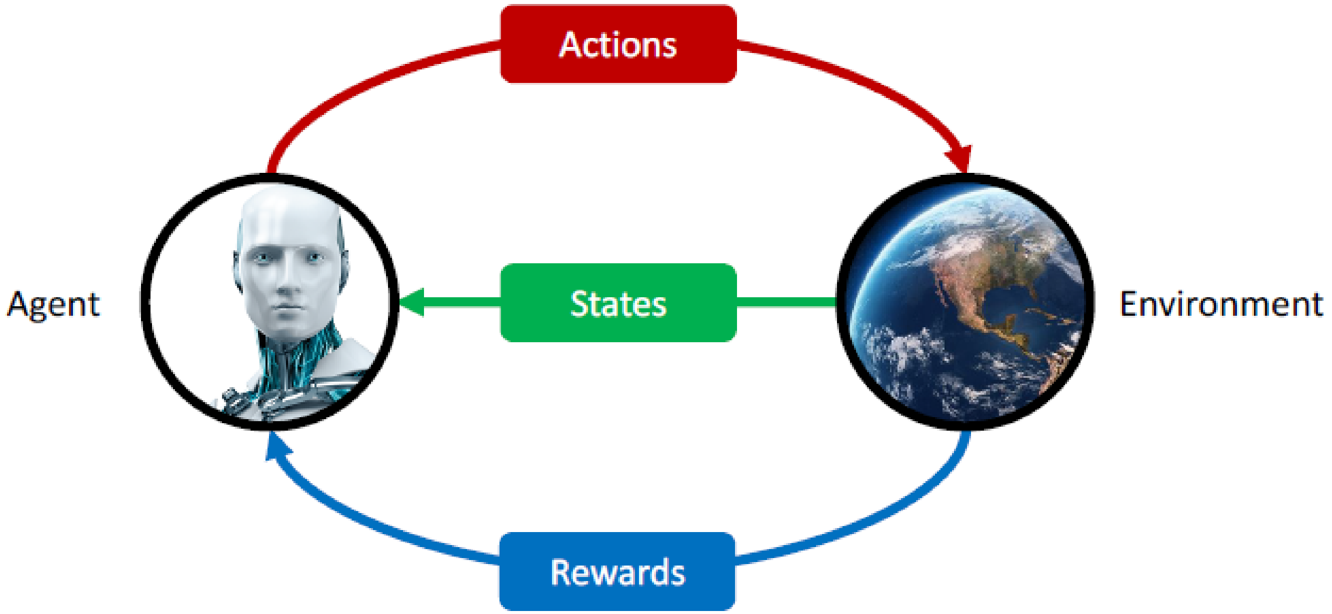


Types of Machine Learning

- **Reinforcement Learning:** Rewards from a sequence of actions

Types of Machine Learning

- **Reinforcement Learning**

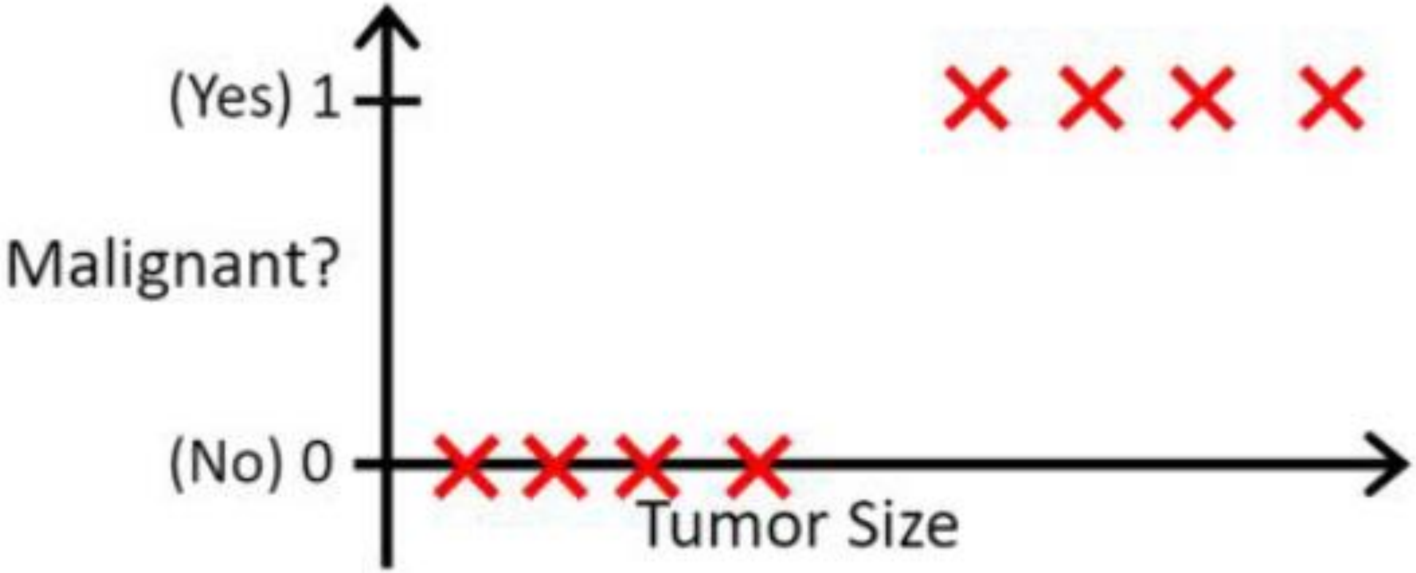


The state space can be discrete or continuous. In case of continuous states, you would use a function approximator to represent your state.

Supervised Learning

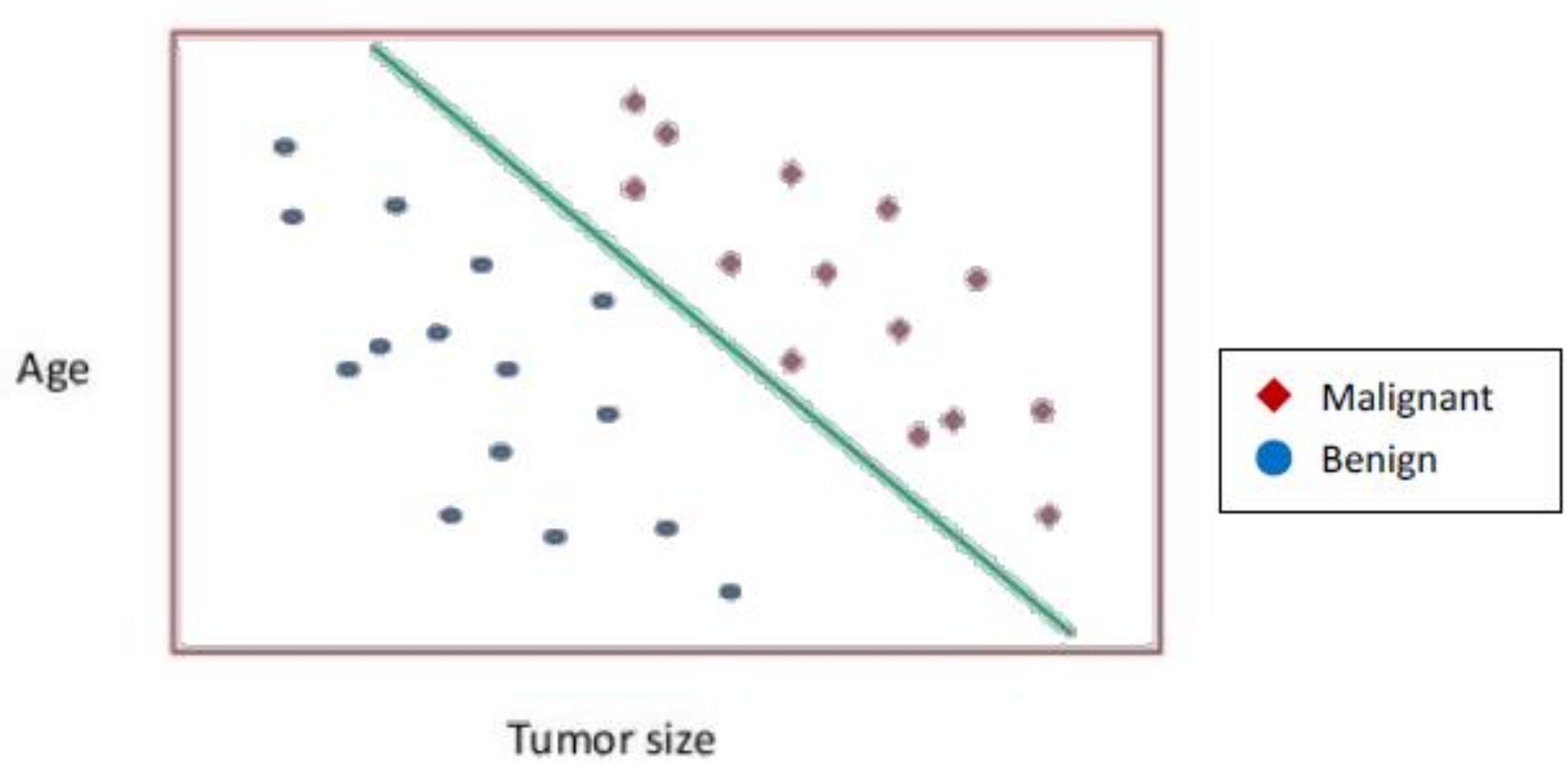
- **Classification (1-d features)**

Learning a function
$y = f(x)$
$x \in \mathbb{R}$
$y \in \{1, 2, \ldots, k\}$



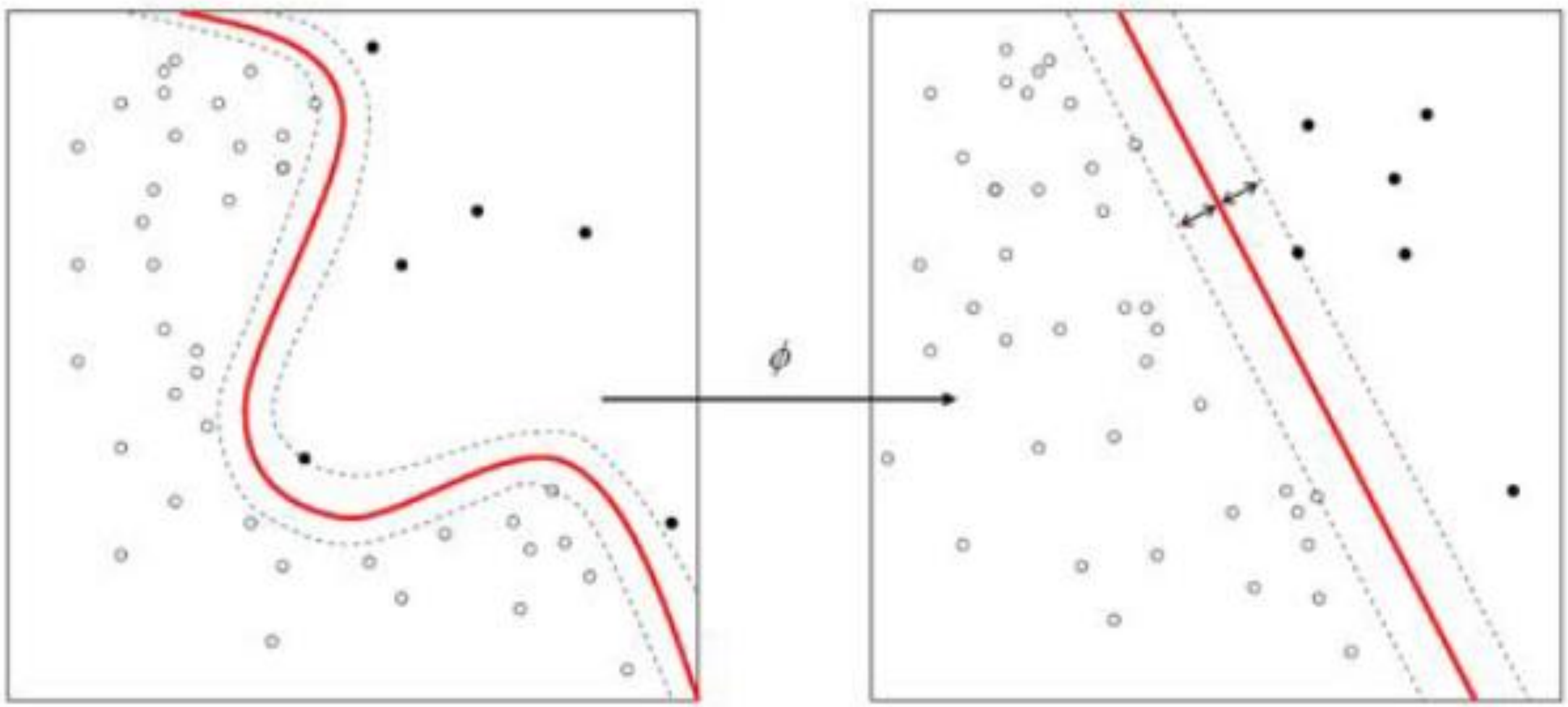
Supervised Learning

- **Classification (2-d features) - Linear**



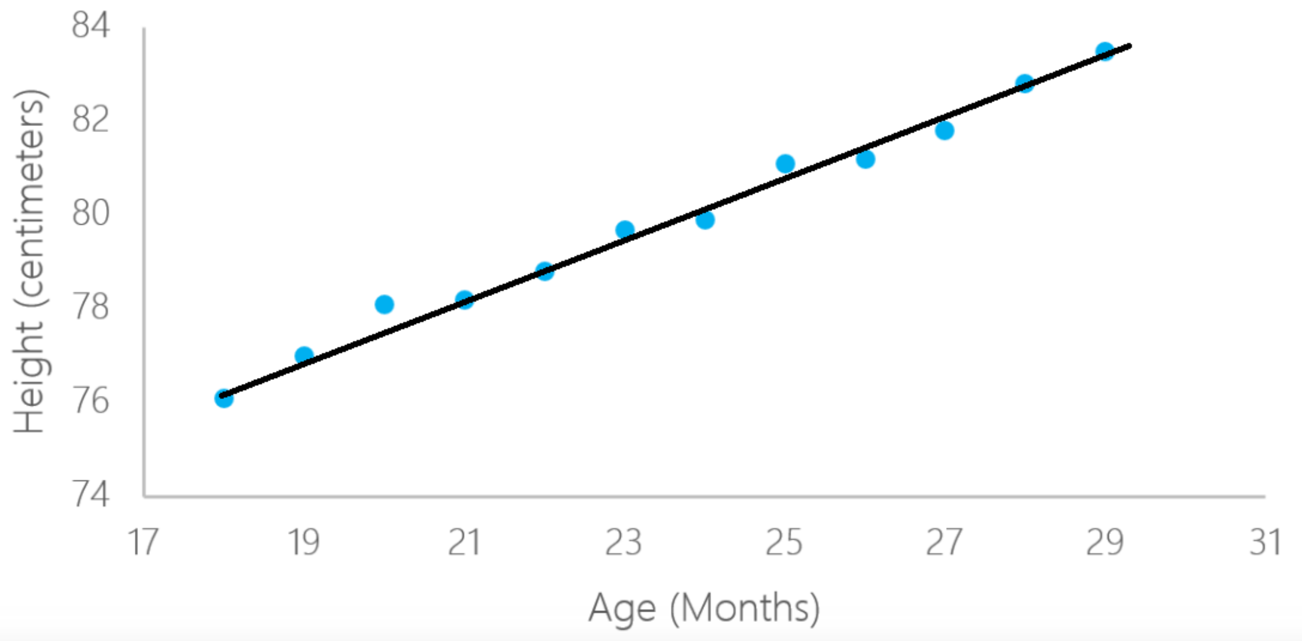
Supervised Learning

- **Classification (Non-Linear)**



Supervised Learning

- **Regression (Linear)**



Learning a function

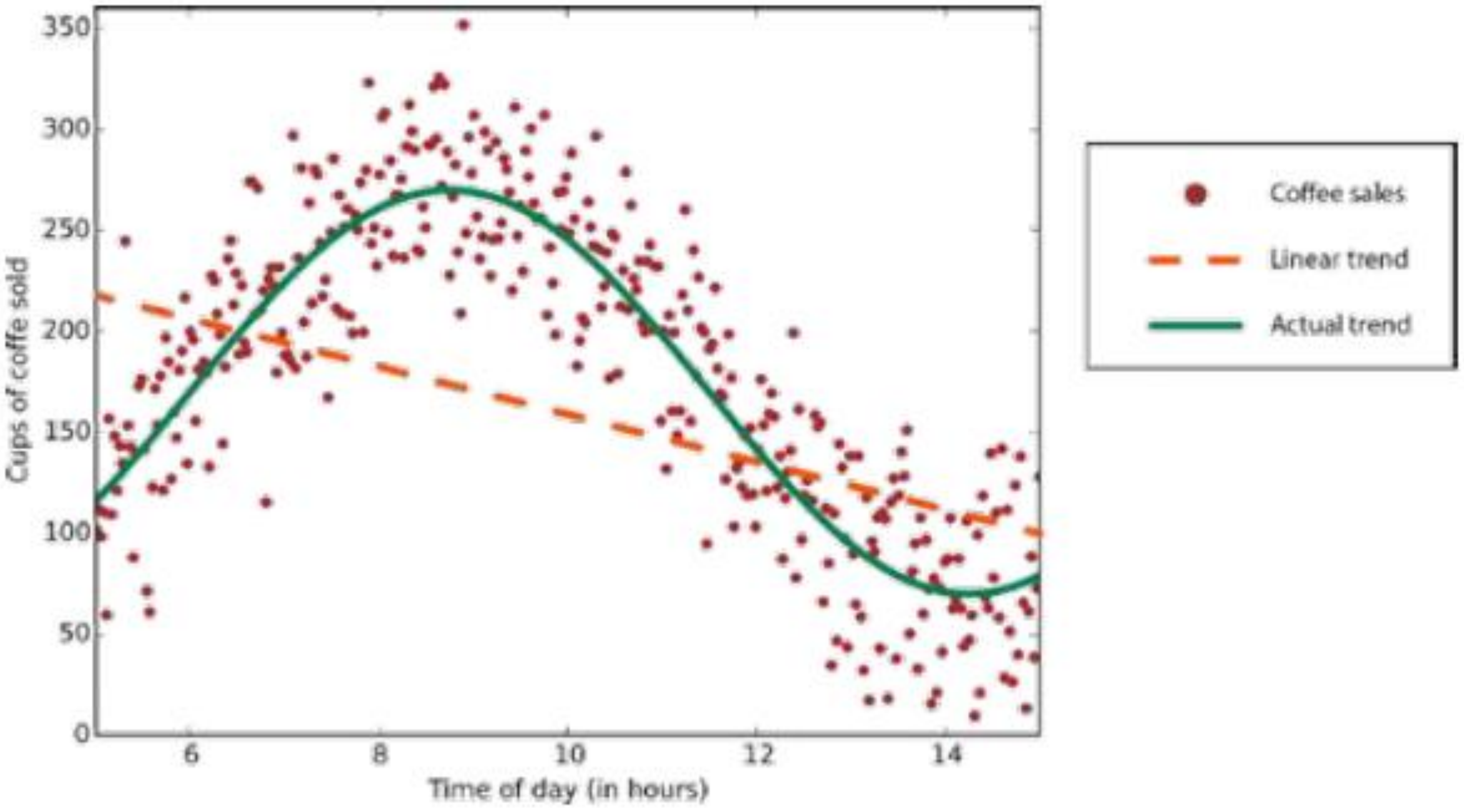
$y = f(x)$

$x \in \mathbb{R}$

$y \in \mathbb{R}$

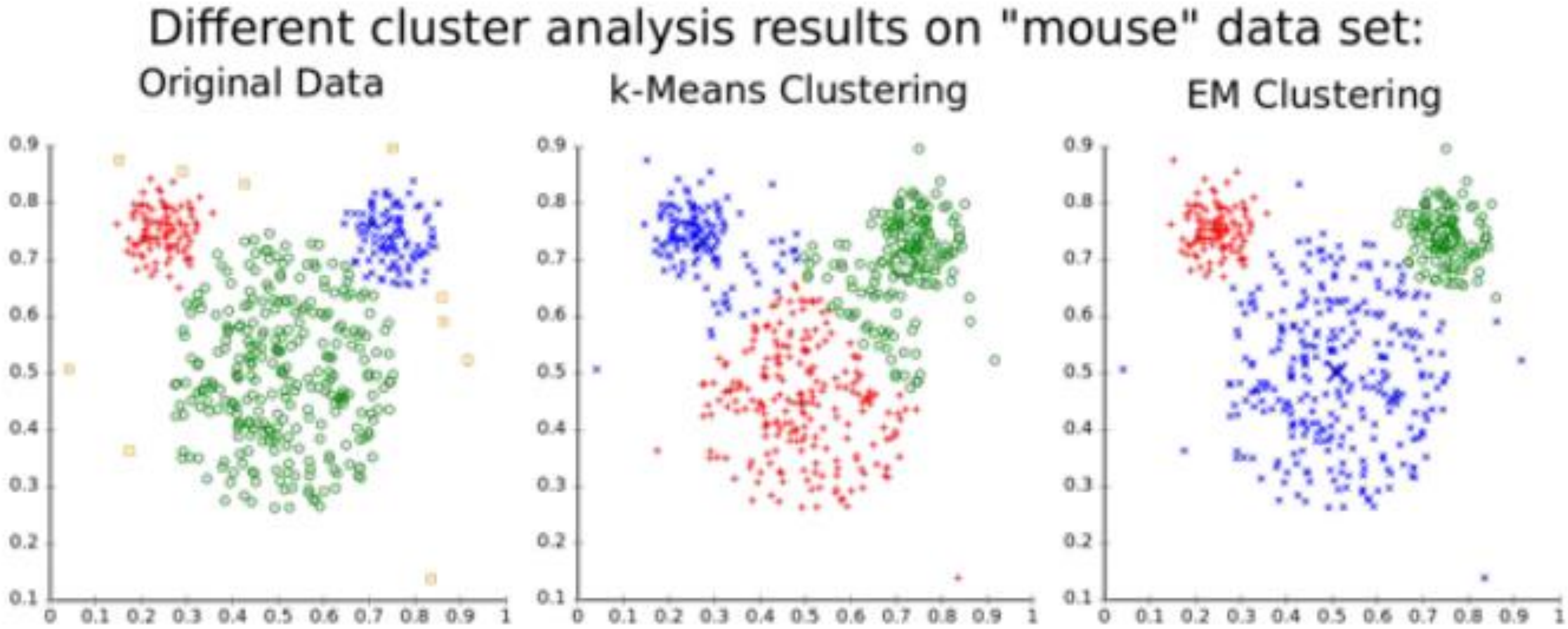
Supervised Learning

- **Regression (Non-Linear)**



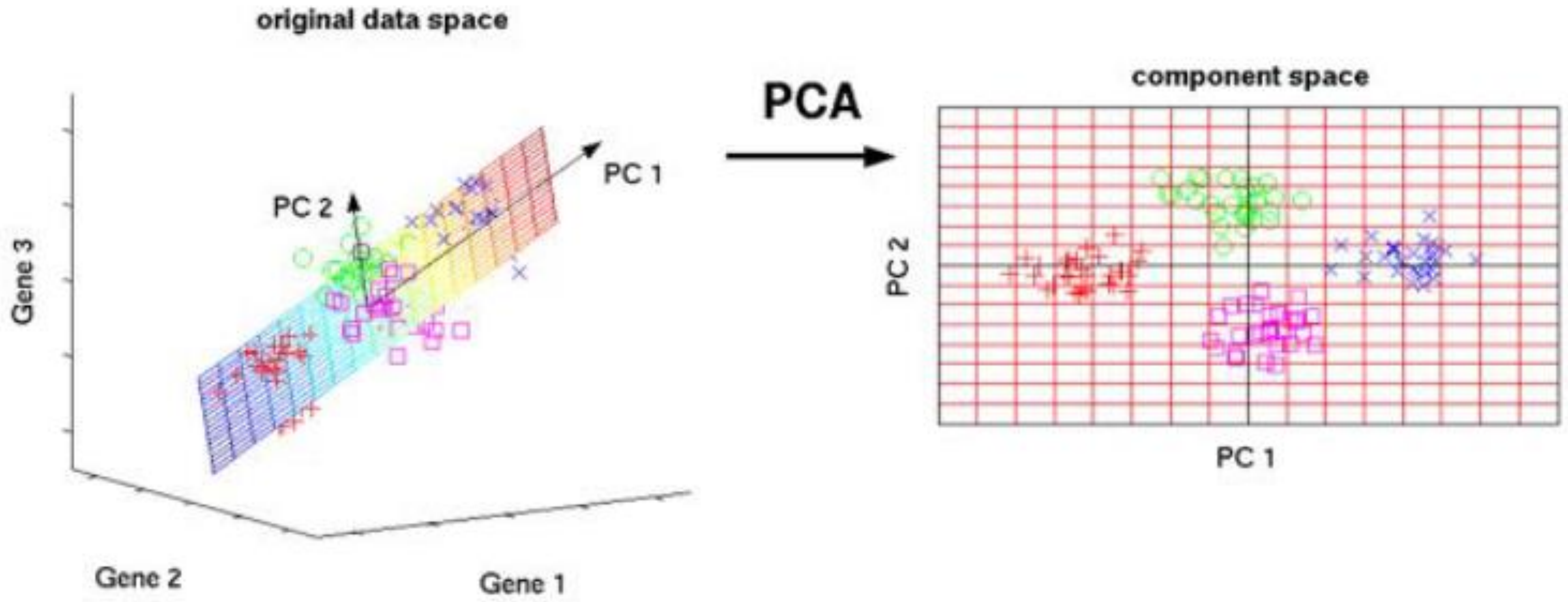
Unsupervised Learning

- **Clustering**



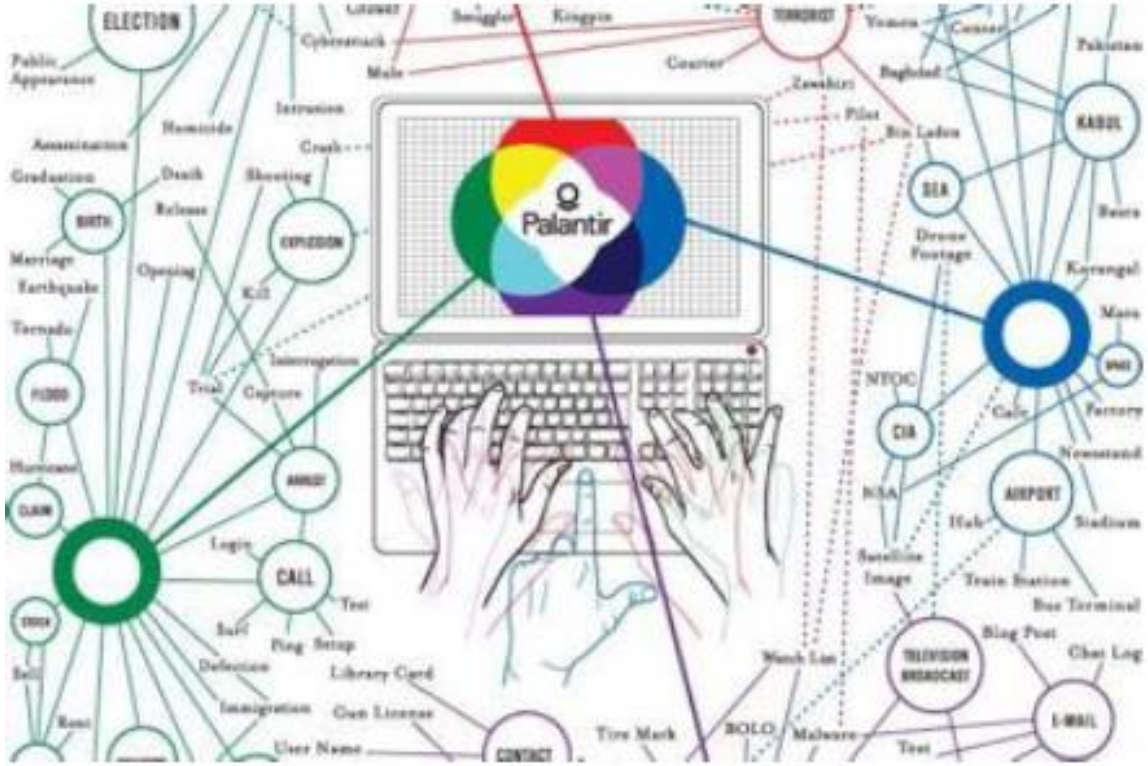
Unsupervised Learning

- **Dimensionality Reduction: Sub-space Learning**



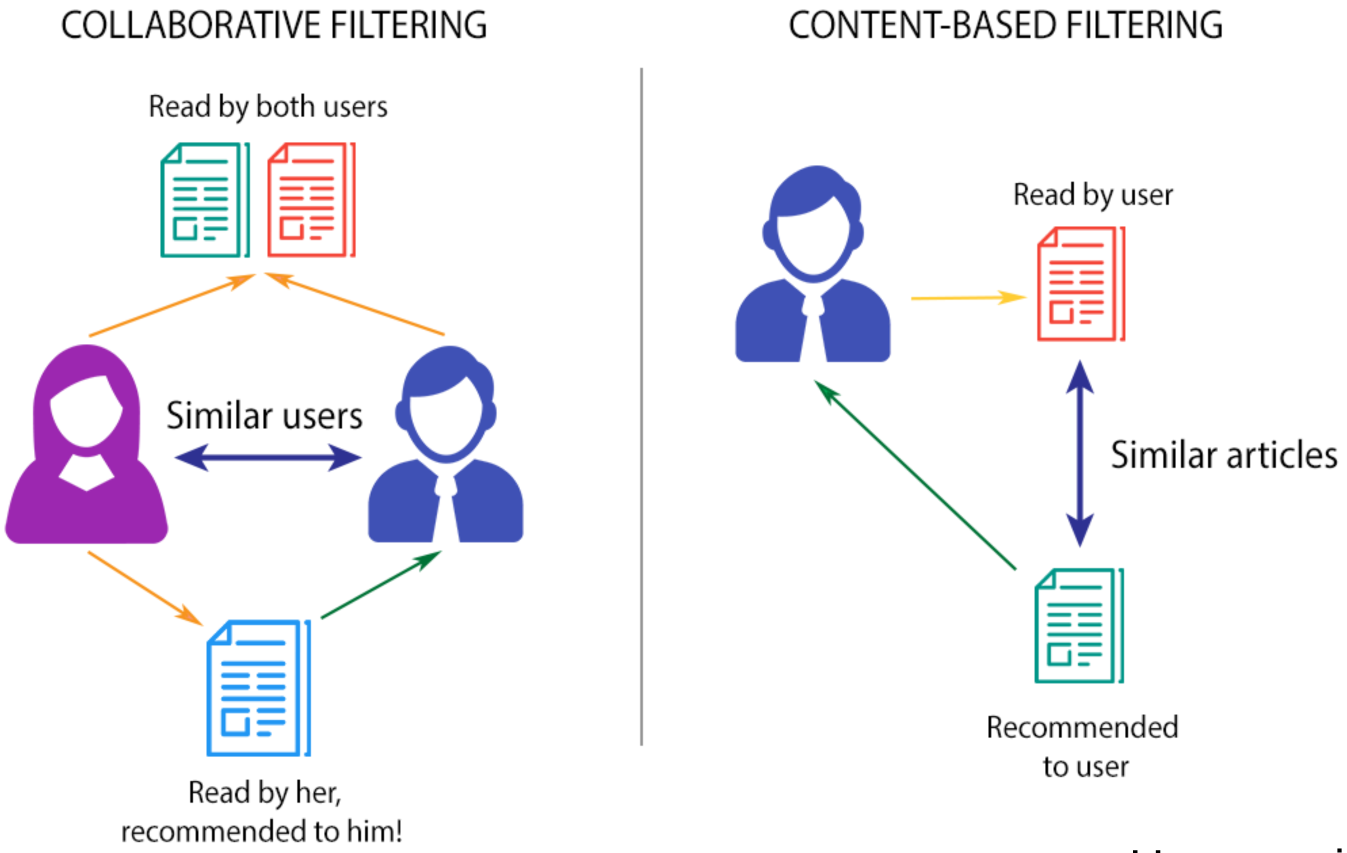
Examples

Fraud detection



Supervised learning
https://www.washingtonian.com/2012/01/31/killer-app/

Recommender Systems



Unsupervised learning

https://towardsdatascience.com/brief-on-recommender-systems-b86a1068a4dd

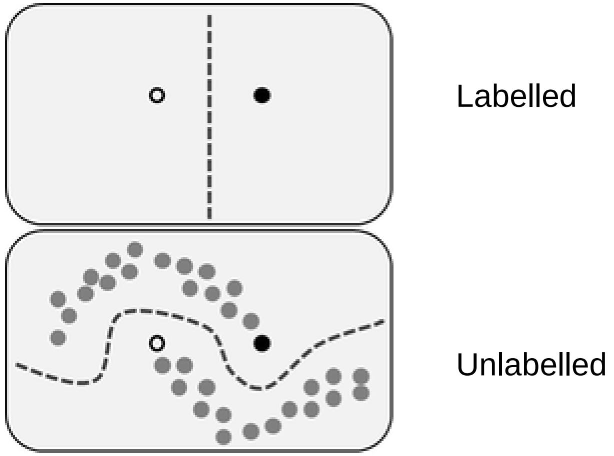
Spam Filters

Bayesian Networks

Supervised learning

Spam Filters

With some labelled emails (spam/not spam) and unlabelled emails in your inbox, we can create a **customized spam filter** for new emails using semi-supervised learning.



Performing clustering and then labelling clusters with labelled data

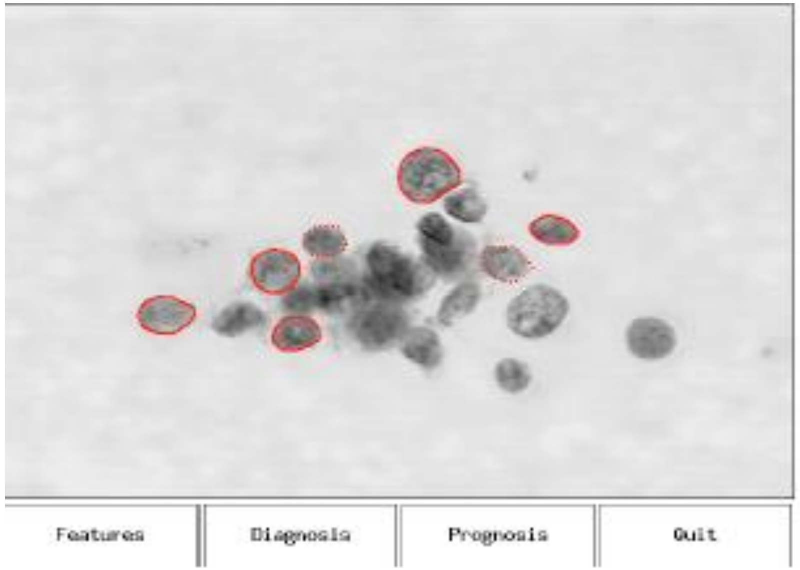
Semi-Supervised learning

https://en.wikipedia.org/wiki/Semi-supervised_learning

A case study

Supervised Learning

Example: A dataset



- Cell samples were taken from tumors in breast cancer patients before surgery, and imaged
- Tumors were excised
- Patients were followed to determine whether or not the cancer recurred, and how long until recurrence or disease free

Example: A dataset

- 30 real-valued variables per tumour
- Two variables that can be predicted:
  - Outcome (R = recurrent, N = non-recurrent)
  - Time (until recurrence, for R, time healthy for N).

tumor size	texture	perimeter	...	outcome	time
18.02	27.6	117.5		N	31
17.99	10.38	122.8		N	61
20.29	14.34	135.1		R	27

Terminology

tumor size	texture	perimeter	. . .	outcome	time
18.02	27.6	117.5		N	31
17.99	10.38	122.8		N	61
20.29	14.34	135.1		R	27

- Columns are called *input variables* or *features* or *attributes*
- The outcome and time (which we are trying to predict) are called *output variables* or *targets or responses*.
- A row in the table is called *training example* or *instance*
- The whole table is called (*training) data set*.
- The problem of predicting the recurrence is called *(binary) classification*.
- The problem of predicting the time is called *regression*.

More formally

tumor size	texture	perimeter	...	outcome	time
18.02	27.6	117.5		N	31
17.99	10.38	122.8		N	61
20.29	14.34	135.1		R	27

Training data

$$\mathcal{S}_n = \left\{ \left(x^{(i)}, y^{(i)}\right) \,\middle|\, i = 1, \dots, n \right\}$$

- Features/Inputs $x^{(i)} = \left(x_1^{(i)}, \dots, x_d^{(i)}\right)^{\top} \in \mathbb{R}^d$
- Response/Output $y^{(i)} \in \mathbb{R}$ or $y \in \{1, 2, \dots, k\}$

Supervised learning problem

- Let $\mathcal{X}$ denote the space of input values
- Let $\mathcal{Y}$ denote the space of output values
- Given a data set $\mathcal{S}_n \subset \mathcal{X} \times \mathcal{Y}$, find a function:

$$h : \mathcal{X} \to \mathcal{Y}$$

  such that $h(\mathbf{x})$ is a *"good predictor"* for the value of y.
- h is called a *hypothesis*
- Problems are categorized by the type of output domain
  - If $\mathcal{Y} = \mathbb{R}$, this problem is called *regression*
  - If $\mathcal{Y}$ is a categorical variable (i.e., part of a finite discrete set), the problem is called *classification*
  - If $\mathcal{Y}$ is a more complex structure (eg graph) the problem is called *structured prediction*

Key aspects of learning problems

- **Set of classifiers H:** modelling
- **Learning algorithm / Criterion:** optimizing
- **Generalization**
  - Choice of H
  - Training data S_n
  - Learning algorithm

Steps to solving a supervised learning problem

1. Decide what the input-output pairs are.
2. Decide how to encode inputs and outputs.
   - This defines the input space X and the output space Y.
3. Choose a class of **hypotheses/representations H (modeling)**.
4. Choose an **error function** (**cost function**) to define the best hypothesis.
5. Choose an **algorithm for searching** efficiently through the space of hypotheses (**optimizing**).

Linear Classification

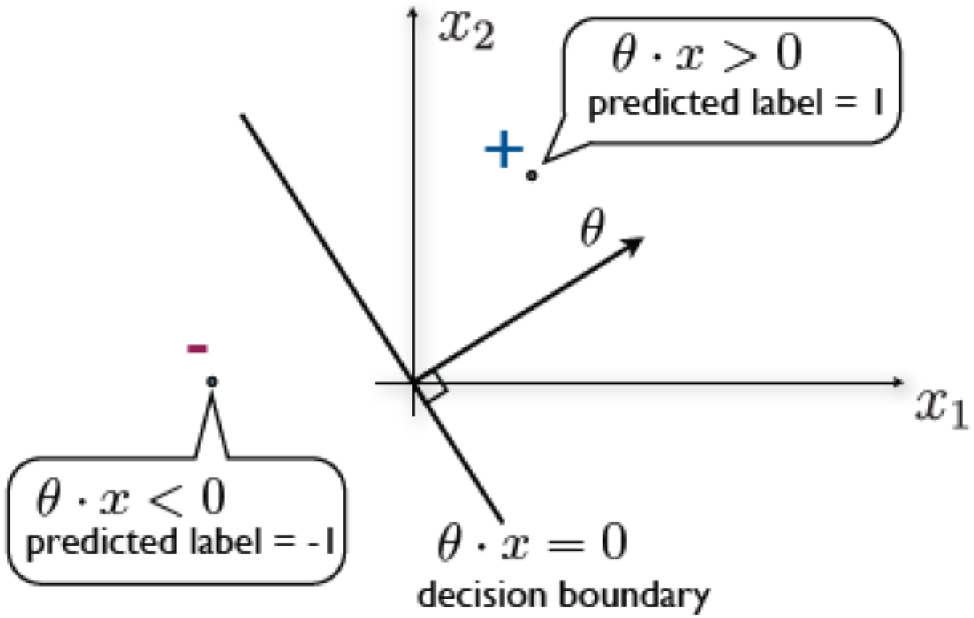
- Let's consider a particular constrained set of classifiers

$$h(x;\theta) = \text{sign}(\theta_1 x_1 + \ldots + \theta_d x_d) = \text{sign}(\theta \cdot x) = \begin{cases} +1, & \theta \cdot x \geq 0 \\ -1, & \theta \cdot x < 0 \end{cases}$$

- $\theta \cdot x = \theta^T x$ and $\theta = [\theta_1, \ldots, \theta_d]^T$ is a column vector of real valued parameters or weights

Linear Classification

$$h(x;\theta) = \text{sign}(\theta_1 x_1 + \ldots + \theta_d x_d) = \text{sign}(\theta \cdot x) = \begin{cases} +1, & \theta \cdot x \geq 0 \\ -1, & \theta \cdot x < 0 \end{cases}$$



Intended Learning Outcomes

- Define machine learning in terms of **algorithms**, **tasks**, **performance** and **experience**.

- List four main types of machine learning, e.g.. **supervised, unsupervised, reinforcement learning, semi-supervised learning**

- Describe some potential dangers in machine learning, e.g. applying an **algorithm without understanding its assumptions**, **forgetting that the training data could be biased.**