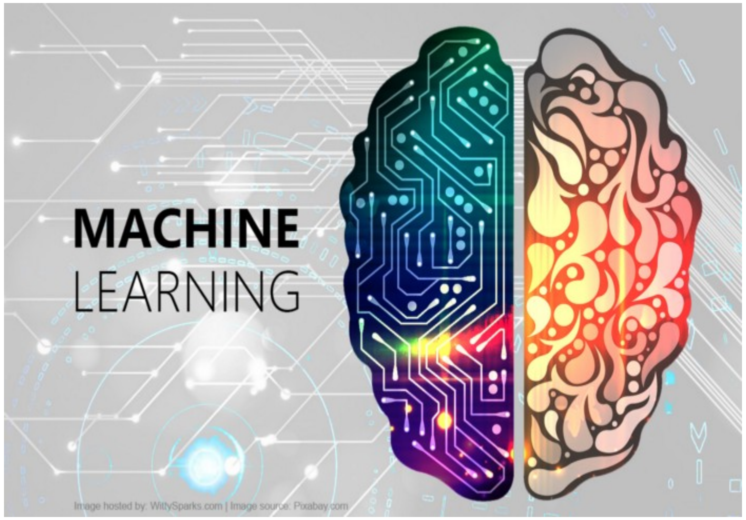
Introduction to Machine Learning

(A gentle approach – no maths!)

Dr. Vassilis S. Kontogiannis
Reader in Computational Intelligence
Email: V.Kodogiannis@westminster.ac.uk
https://scholar.google.co.uk/citations?user=meTTcLAAAAAJ&hl=en&oi=ao



What is Machine Learning?

"The subfield of computer science that "gives computers the ability to learn without being explicitly programmed"
(Arthur Samuel, 1959)

"***Machine learning (ML)*** *is concerned with the design and development of algorithms and techniques that* ***allow computers to "learn"****. The* ***major focus*** *of ML research is to* ***extract information from data automatically****, by computational and statistical methods. It is thus closely related to data mining and statistics*".
(Svensson and Söderberg, 2008)

What is Machine Learning?

Definition by Tom Mitchell (1998):

Machine Learning is the study of algorithms that

- improve their performance P
- at some task T
- with experience E

A well-defined learning task is given by <P, T, E>

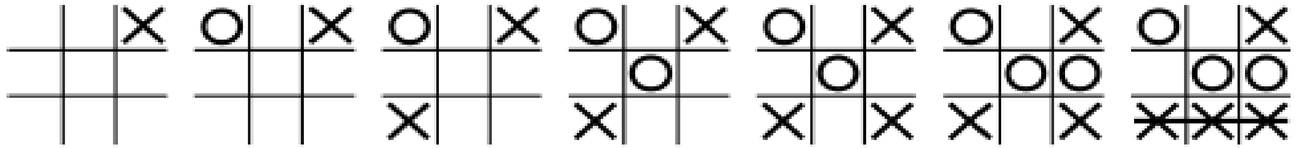
Examples:

T: Playing Chess (or Go)
P: Percent games won against an opponent
E: Playing games against itself

T: Classify emails as legitimate or spam
P: Percentage of emails labeled correctly
E: Repository of emails, some with human-specified labels

Example: tic-tac-toe

- How to program the computer to play tic-tac-toe?



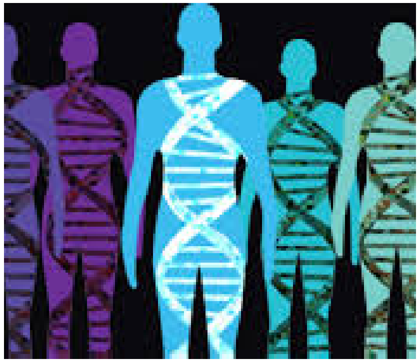
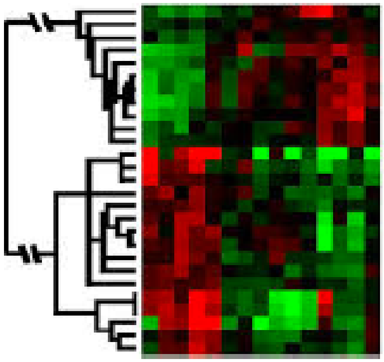
- Option A: The programmer writes explicit rules, e.g. 'if the opponent has two in a row, and the third is free, stop it by placing your mark there', etc (lots of work, difficult, not at all scalable!)
- Option B: Go through the game tree, choose optimally (for non-trivial games, must be combined with some heuristics to restrict tree size)
- Option C: Let the computer try out various strategies by playing against itself and others, and noting which strategies lead to winning and which to losing (='machine learning')

When Do We Use Machine Learning?

ML is used when:

- Human expertise does not exist (navigating on Mars)
- Humans can’t explain their expertise (speech recognition)
- Models must be customized (personalized medicine)
- Models are based on huge amounts of data (genomics)



Learning isn’t always useful:

- There is no need to “learn” to calculate payroll

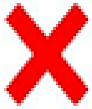
Why "Learn" ?

- *"The goal of machine learning is to make a computer learn just like a baby — it should get better at tasks with experience."*

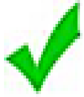
- A machine learning system can be used to
  - *Automate a process*
  - *Automate decision making*
  - *Extract knowledge from data*
  - *Predict future event*
  - *Adapt systems dynamically to enable better user experiences*
  - …

- How do we build a machine learning system?

Write code to explicitly do the above tasks



Write code to make the computer ***learn*** how to do the tasks

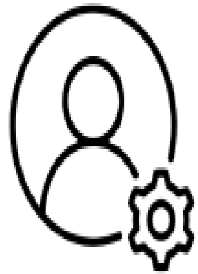


When We Need Machine Learning

Tasks involving big data

- Genomics
- Internet search
- Anomaly detection

Tasks requiring customization



- Email filters
- Personalized medicine
- Image inpainting

Tasks for which it is challenging to specify our knowledge

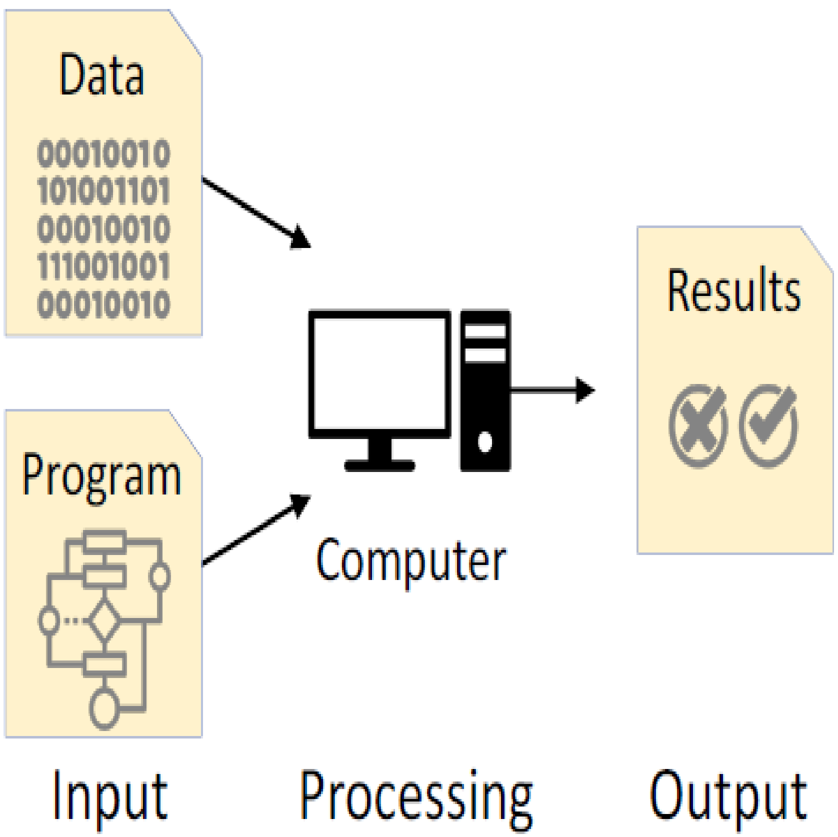
- Facial recognition
- Understanding speech
- Medical diagnosis

Tasks for which we don’t have human expertise

- Space exploration
- Undersea manipulation
- Cellular robotics

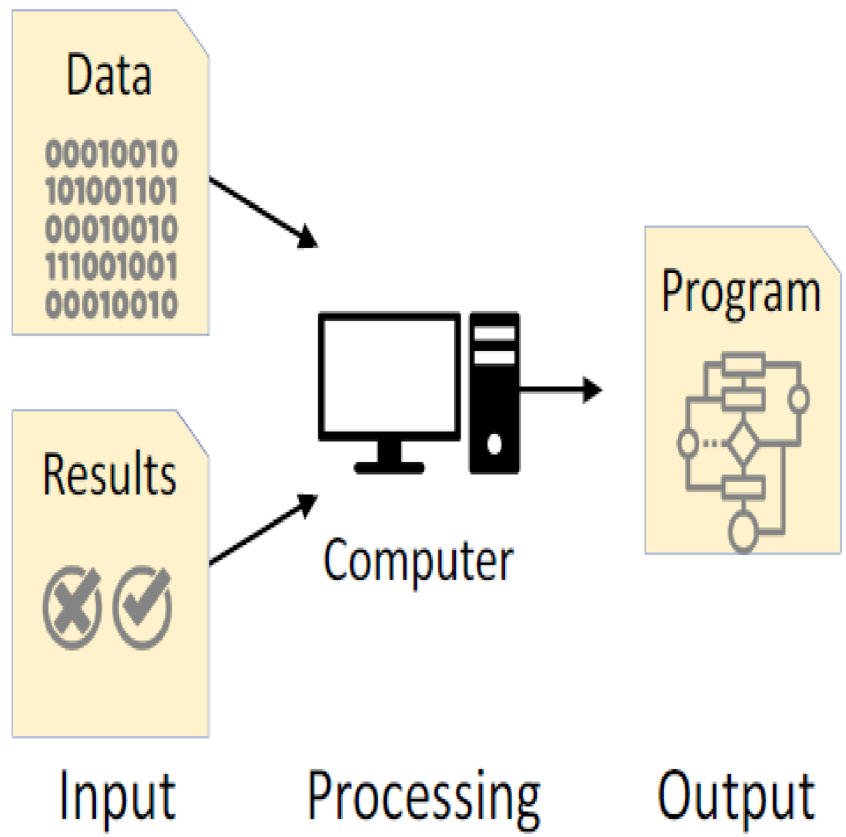
Traditional Programming

Works well when we know how to specify the program



Machine Learning

Needed when we <u>don't</u> know how to specify the program



History of Machine Learning - I

- 1940s, Human reasoning / logic first studied as a formal subject within mathematics (Claude Shannon, Kurt Godel et al).

- 1950s, The Turing Test is proposed: a test for true machine intelligence, expected to be passed by year 2000. Various game-playing programs built. 1956, Dartmouth conference coins the phrase artificial intelligence. 1959, Arthur Samuel wrote a program that learnt to play draughts (checkers if you are American).

- 1960s, A.I. funding increased (mainly military). Famous quote: Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved."

- 1970s, A.I. winter. Funding dries up as people realise it is hard. Limited computing power and dead-end frameworks.

- 1980s, Revival through bio-inspired algorithms: Neural networks, Genetic Algorithms. A.I. promises the world – lots of commercial investment – mostly fails. Rule based expert systems used in medical / legal professions.

History of Machine Learning - II

- 1990s, AI diverges into separate fields: Machine Learning, Computer Vision, Automated Reasoning, Planning systems, Natural Language processing… Machine Learning begins to overlap with statistics / probability theory.
- 2000s, ML merging with statistics continues. Other subfields continue in parallel. First commercial-strength applications: Google, Amazon, computer games, route-finding, credit card fraud detection, etc... Tools adopted as standard by other fields e.g. biology.
- 2010s, deep neural networks have led to significant performance improvement in speech recognition, reinforcement learning, image classification, machine translation, etc..
- Future?

Some links on machine learning history:
https://en.wikipedia.org/wiki/Timeline_of_machine_learning
https://cloud.withgoogle.com/build/data-analytics/explore-history-machine-learning/

Challenge

Recognize dogs in images

What a human sees

DOG

NOT A DOG

What a computer sees

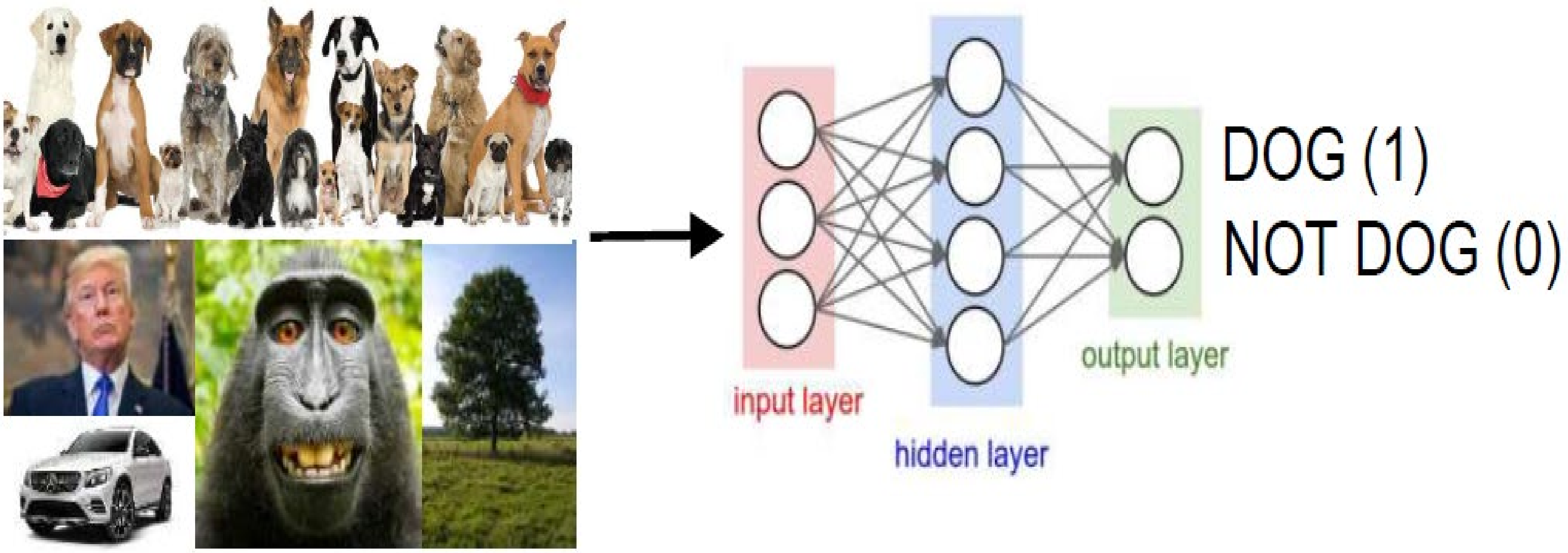
01101111	01101110	01101111	01101110
01100101	01110011	01100101	01110011
00100000	01100001	00100000	01100001
01101110	01100100	01101110	01100100
00100000	01111010	00100000	01111010
01100101	01110010	01100101	01110010
01101111	01110011	01101111	01110011

??? ???

What do we need?

TRAINING DATA

Training phase



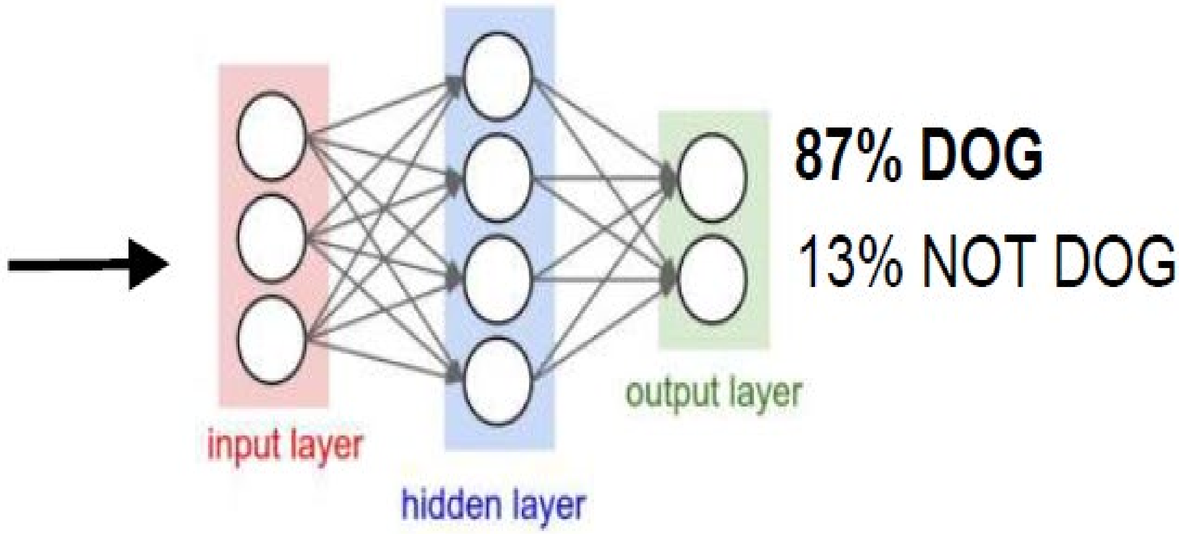
Labeled training set (dog/not dog)

> 1000 images

Untrained Neural Network

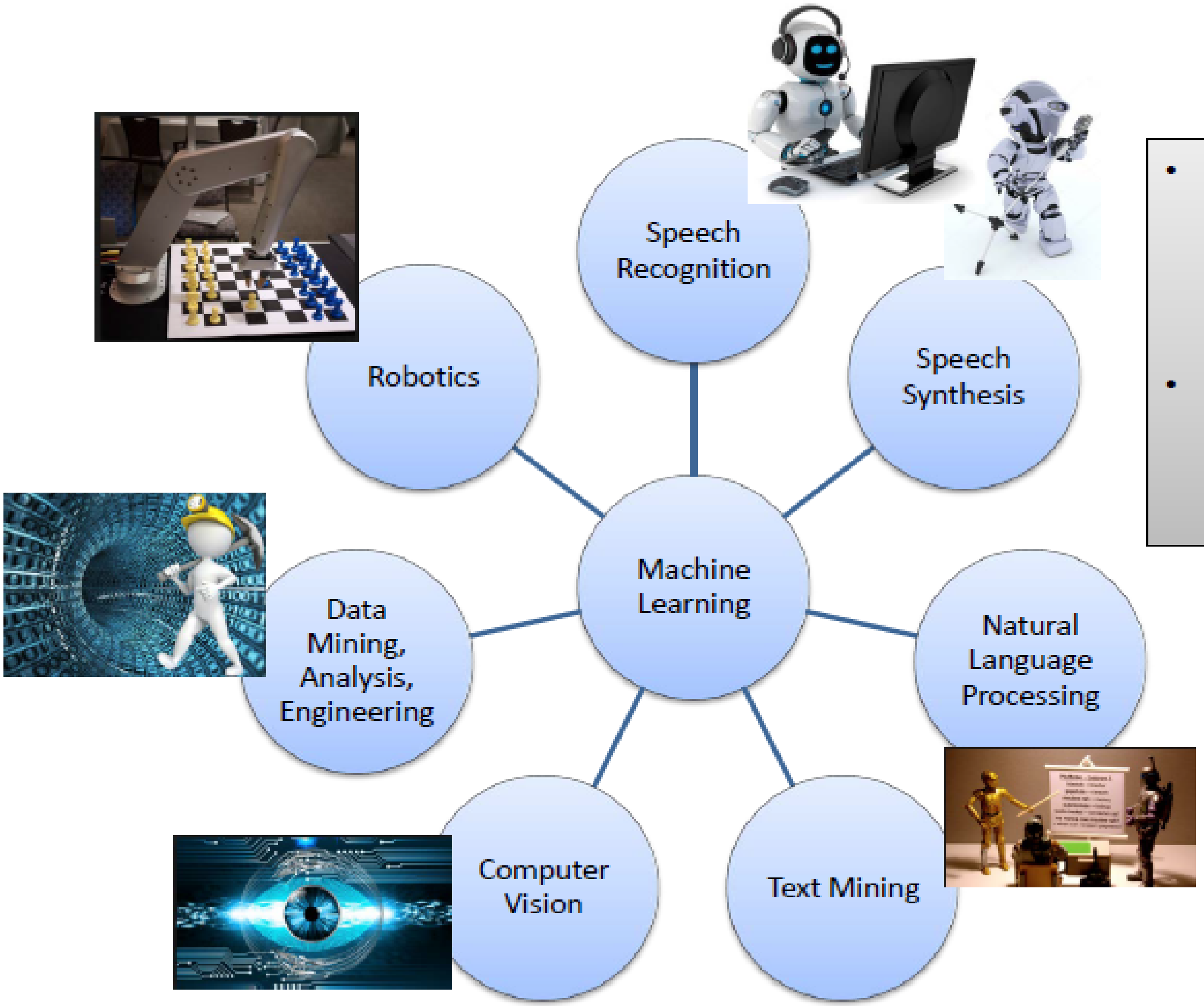
Prediction phase

Unlabeled image



Trained Neural Network

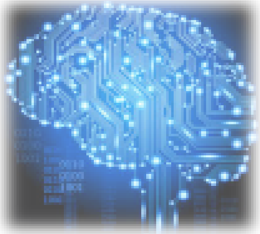
Machine learning in A.I.



- All of these are subfields of Artificial Intelligence (A.I.).
- Machine learning plays a significant role in A.I.

Machine learning Concept/Requirement

Machine Learning: Methodology



- Basic idea:
  - To represent experiences with **data**.
  - To covert a task to a **parametric model**.
  - To convert the learning quality to an **objective function**.
  - To determine the model through optimising an objective function.

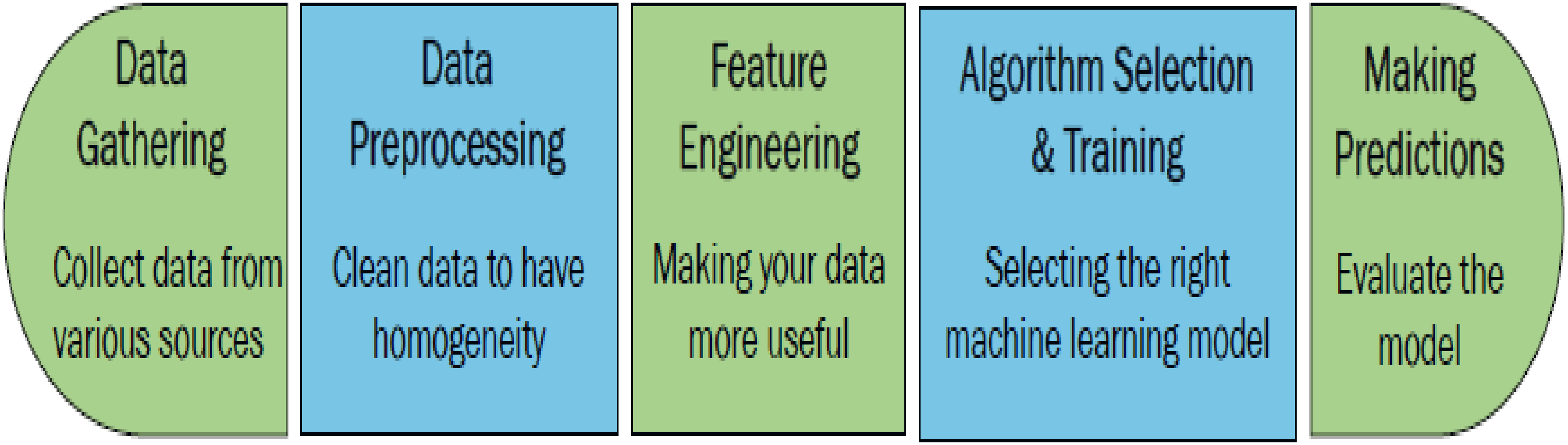
- Machine learning research builds on optimisaton theory, linear algebra, probability theory…

Maths Knowledge Overview

- Linear Algebra:
  - Concepts: vector, matrix, etc.
  - Operations: transpose, sum, multiplication, trace, inverse, etc.

- Calculus:
  - Derivative, partial derivative, gradient, etc.

Steps to Solve a Machine Learning Problem



Data Gathering

Might depend on human work

- Manual labelling for supervised learning.
- Domain knowledge. Maybe even experts.

The more the better: Some algorithms need large amounts of data to be useful (e.g., neural networks).

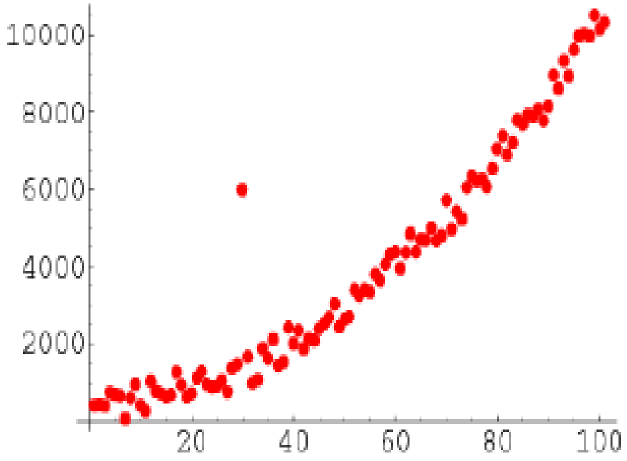
The quantity and quality of data dictate the model accuracy

Data Pre-processing

Is there anything **wrong** with the data?

- Missing values
- Outliers
- Bad encoding (for text)
- Wrongly-labeled examples
- Biased data
  - Do I have many more samples of one class than the rest?

Need to fix/remove data?



Feature Engineering

What is a feature?

A feature is an individual measurable property of a phenomenon being observed

Extract more information from existing data, not adding "new" data by itself

- Making it more useful
- With good features, most algorithms can learn faster

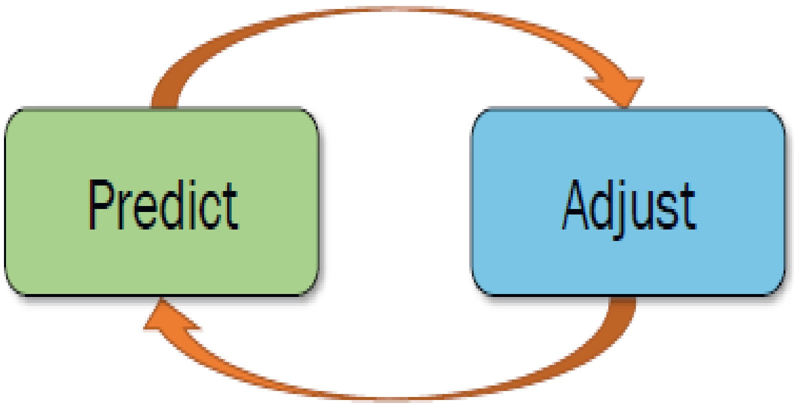
It can be an "art"

- Requires thought and knowledge of the data

Algorithm Selection & Training

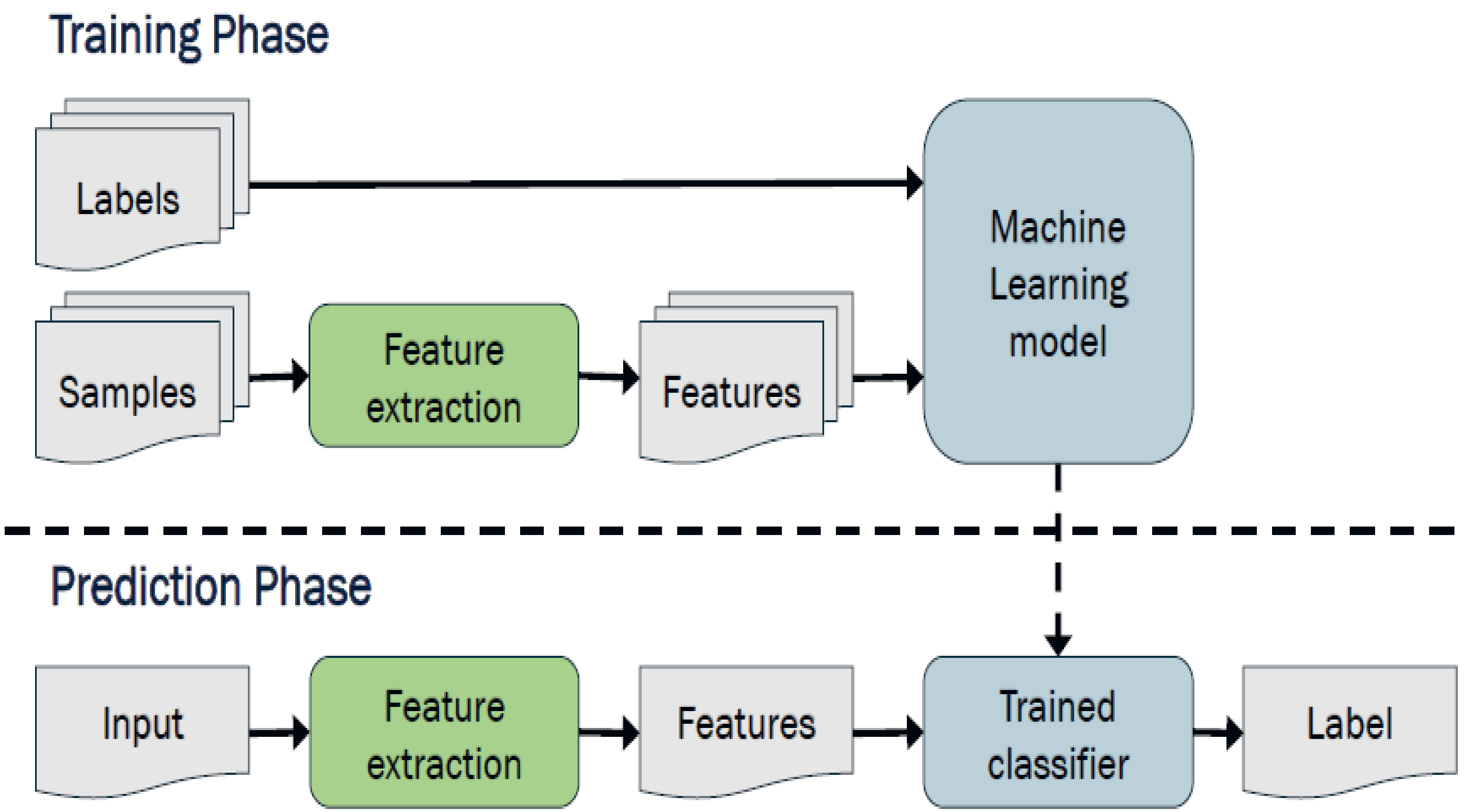
Goal of training: making the correct prediction as often as possible

- Incremental improvement:



- Use of metrics for **evaluating** performance and comparing solutions
- **Hyperparameter tuning**: more an art than a science

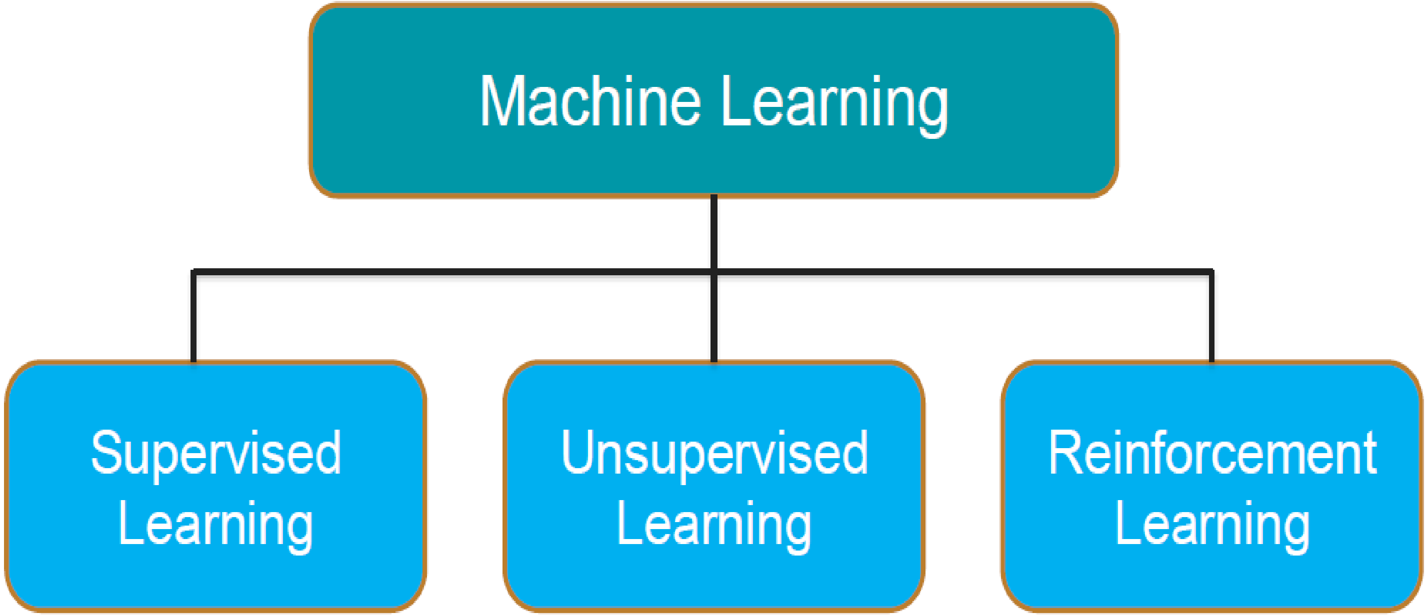
Making Predictions



Summary

- Machine Learning is intelligent use of data to answer questions
- Enabled by an **exponential** increase in computing power and data availability
- Three big types of problems: **supervised**, **unsupervised**, **reinforcement**
- 5 steps to every machine learning solution:
    1. Data Gathering
    2. Data Preprocessing
    3. Feature Engineering
    4. Algorithm Selection & Training
    5. Making Predictions

Types of Learning



Supervised learning

- Given: training data + desired outputs (labels)

Unsupervised learning

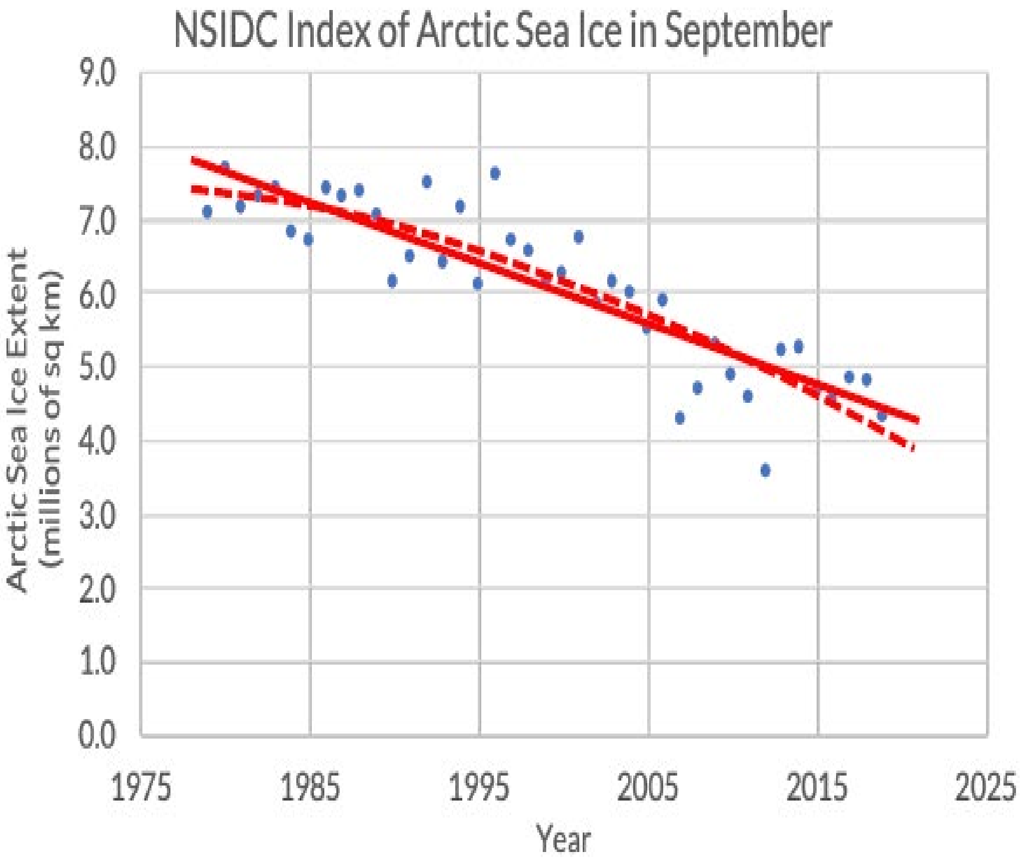
- Given: training data (without desired outputs)

Reinforcement learning

- Rewards from sequence of actions

Supervised Learning: Regression

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is numeric == regression



Supervised Learning: Regression Examples

The target output is a continuous number (or a set of such numbers).

- Finance: x=current market conditions and other possible side information, y=tomorrow's stock market price
- Social Media: x=videos the viewer is watching on YouTube, y=viewer's age
- Robotics: x=control signals sent to motors, y=the 3D location of a robot arm end effector
- Medical Health: x=a number of clinical measurements, y=the amount of prostate specific antigen in the body
- Environment: x=weather data, time, door sensors, etc., y=the temperature at any location inside a building

… … … this list can never end, applications of regression are vast and extremely active!

Example: Stock price prediction

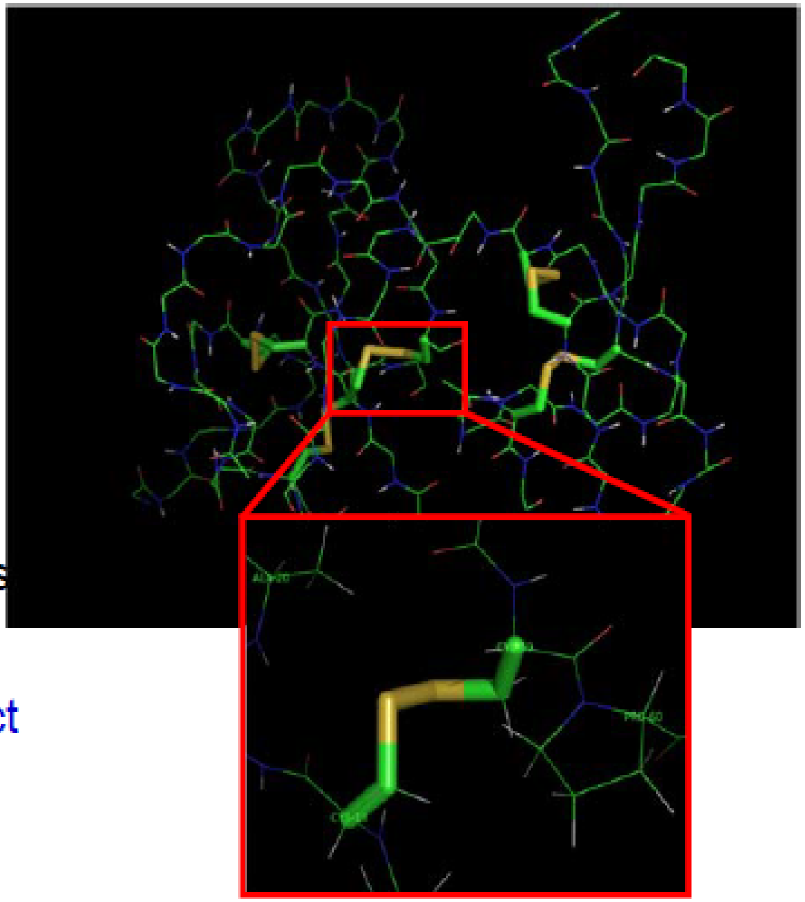
- Task is to predict stock price at future date
- This is a regression task, as the output is continuous

Example: Computational biology

x

y

AVITGACERDLQCG
KGTCCAVSLWIKSV
RVCTPVGTSGEDCH
PASHKIPFSGQRMH
HTCPCAPNLACVQT
SPKKFKCLSK



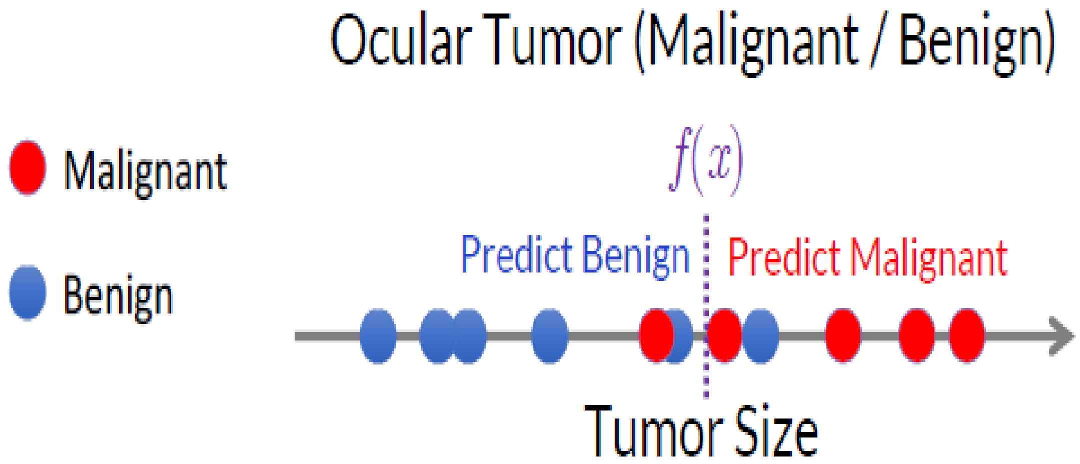
Protein Structure and Disulfide Bridges

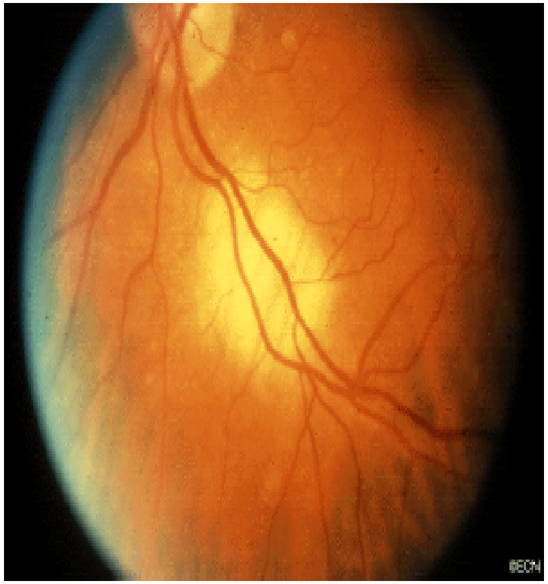
Regression task: given sequence predict 3D structure

Protein: 1IMT

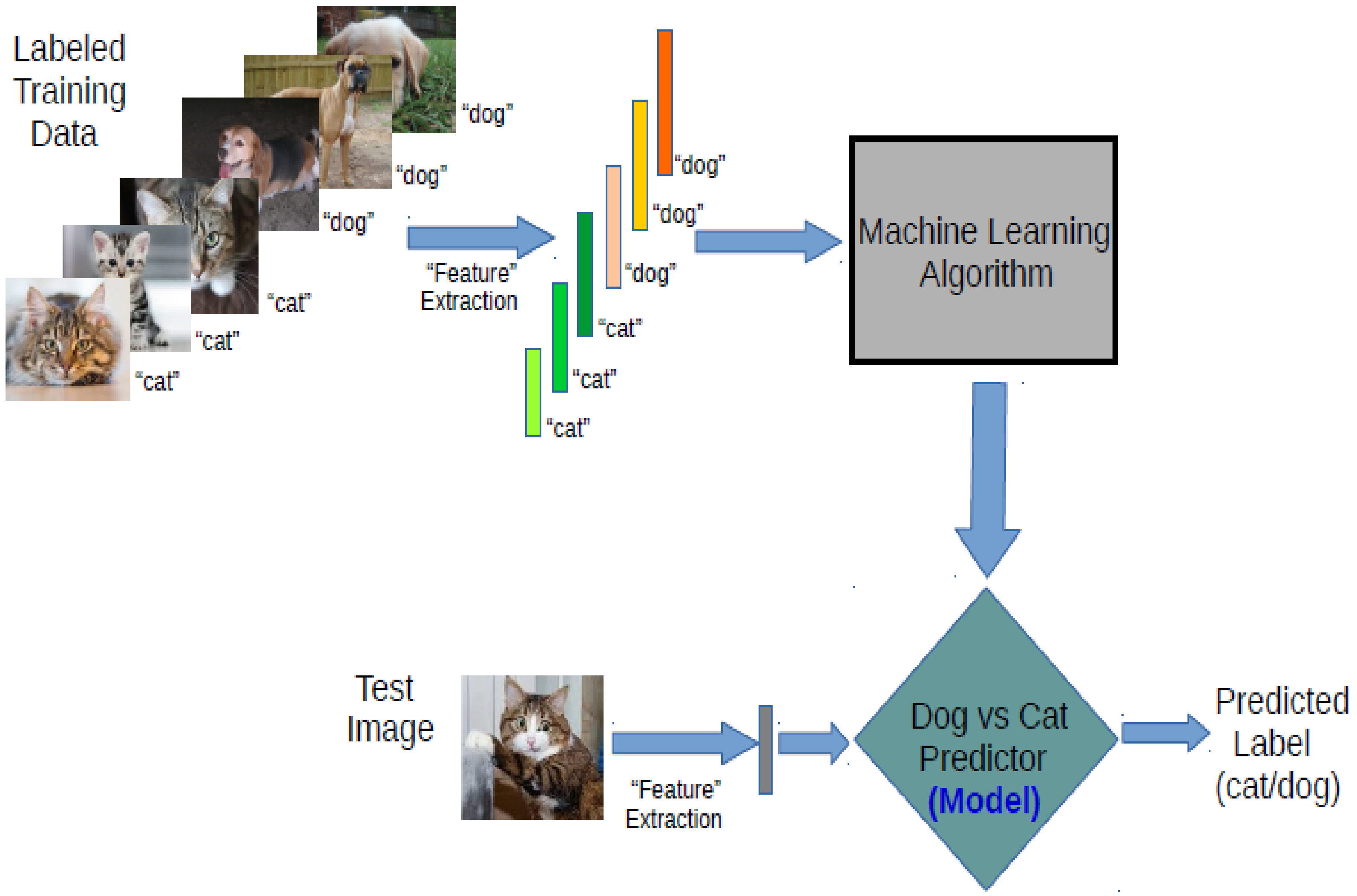
Supervised Learning: Classification

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification





Supervised Learning Procedure (Classification)



Supervised Learning: Classification Examples

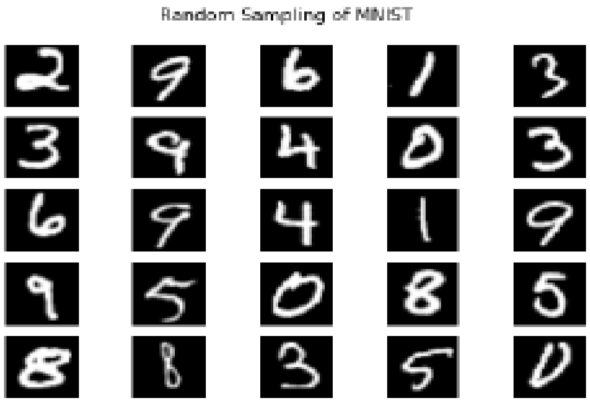
The target output is a category label.

- Medical diagnosis: x=patient data, y=positive/negative of some pathology
- Optical character recognition: x=pixel values and writing curves, y='A', 'B', 'C', …
- Image analysis: x=image pixel features, y=scene/objects contained in image
- Weather: x=current & previous conditions per location, y=tomorrow's weather

… … … this list can never end, applications of classification are vast and extremely active!

Supervised Classification. Example: Image classification

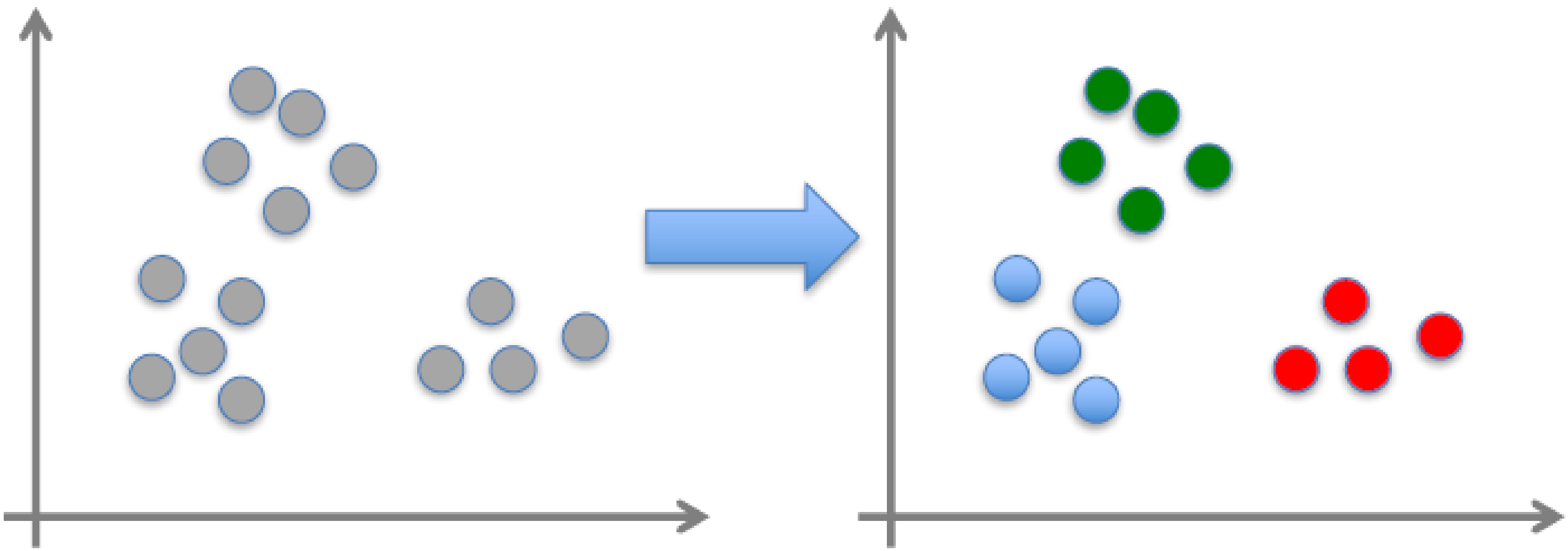
- Handwritten digit recognition (convert hand-written digits to characters 0..9)



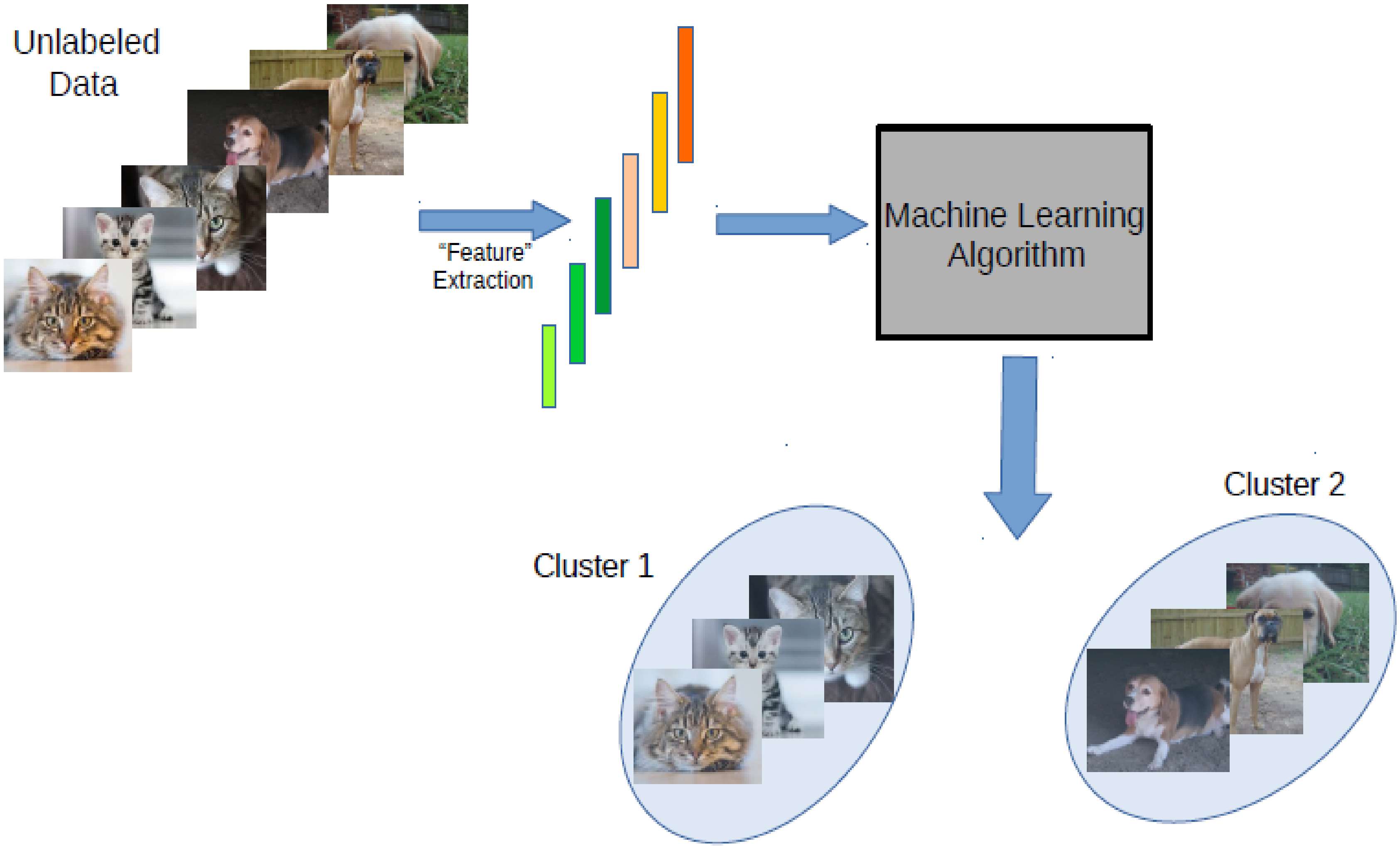
- Face Detection and Recognition

Unsupervised Learning

- Given x_1, x_2, …, x_n (without labels)
- Output hidden structure behind the x's
  - E.g., clustering

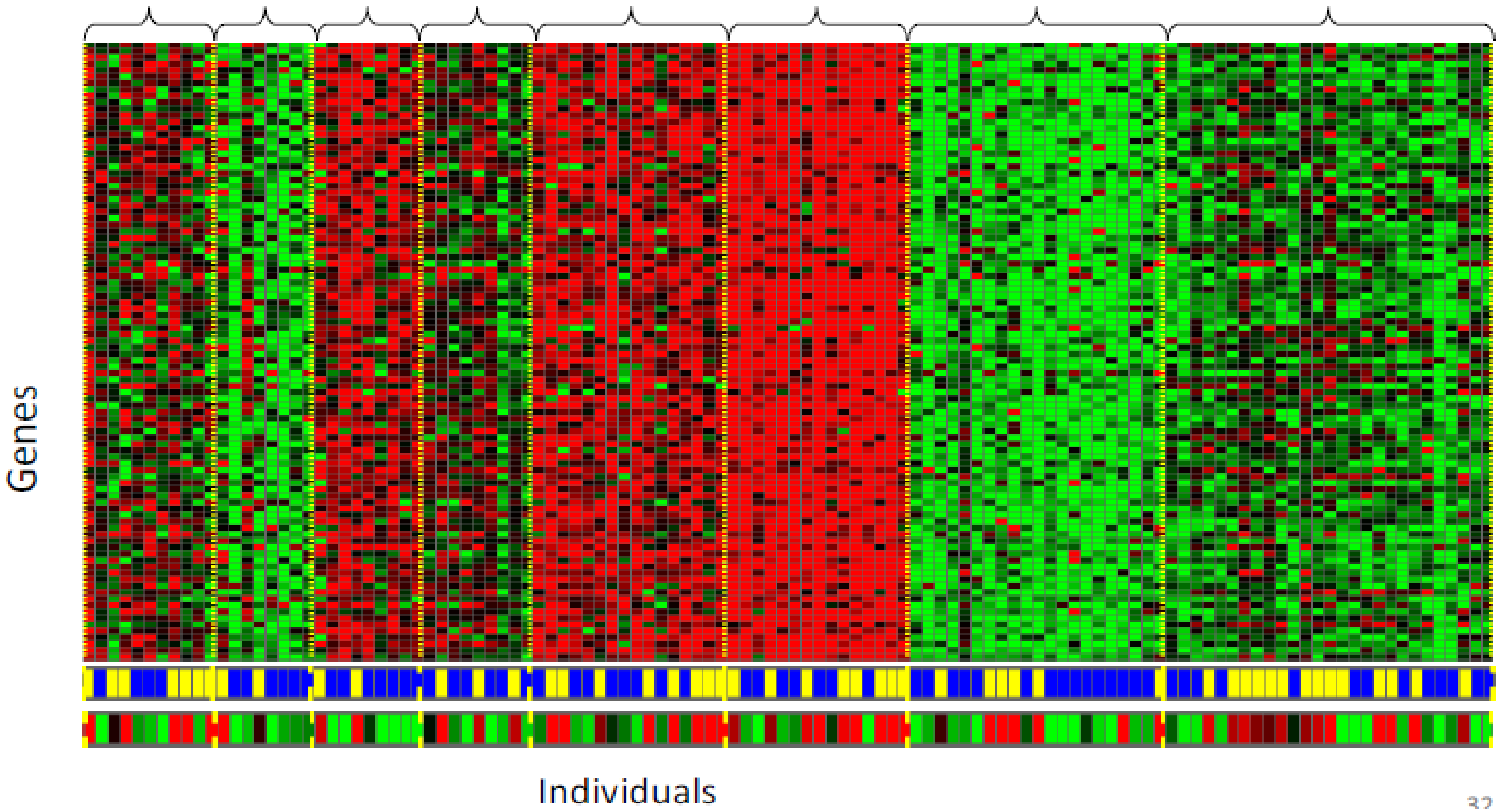


Unsupervised Learning Procedure (Clustering)



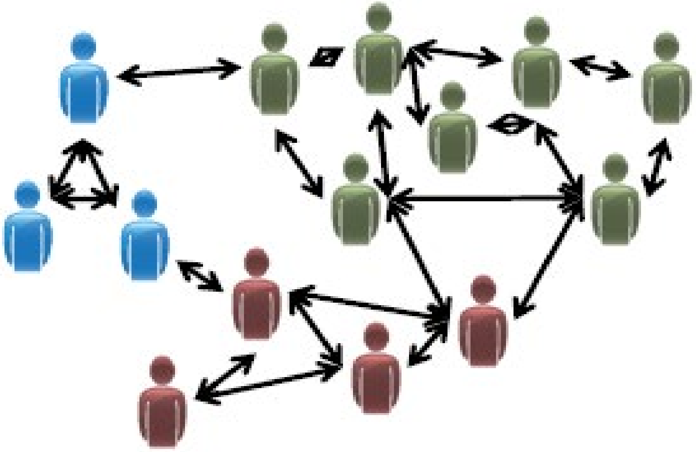
Unsupervised Learning

Genomics application: group individuals by genetic similarity



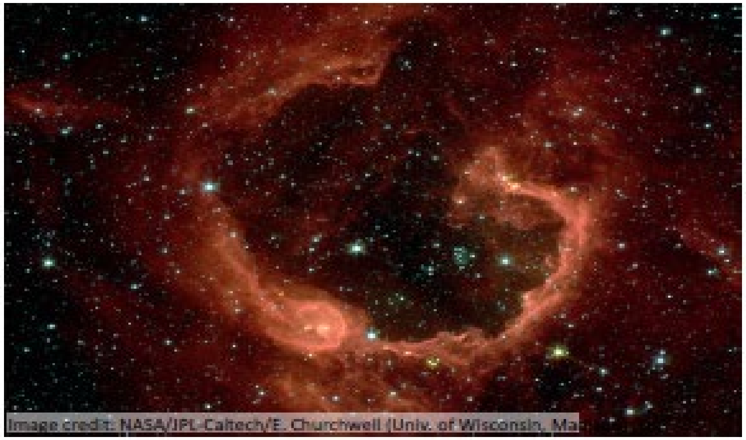
Unsupervised Learning

Organize computing clusters



Social network analysis

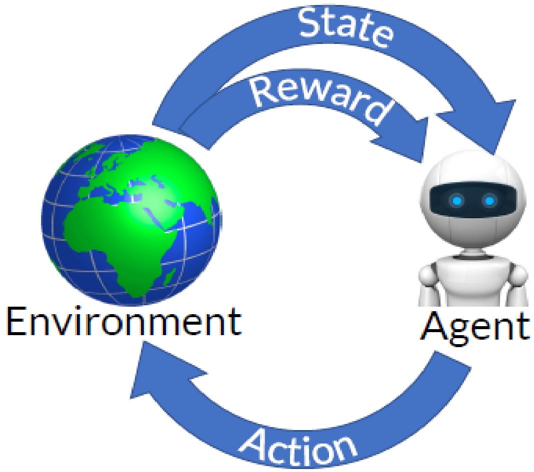
Market segmentation



Astronomical data analysis

Reinforcement Learning

- Given a sequence of states and actions with (delayed) rewards, output a policy
  - Policy is a mapping from states → actions that tells you what to do in a given state
- Examples:
  - Credit assignment problem
  - Game playing
  - Robot in a maze
  - Balance a pole on your hand



Reinforcement Learning

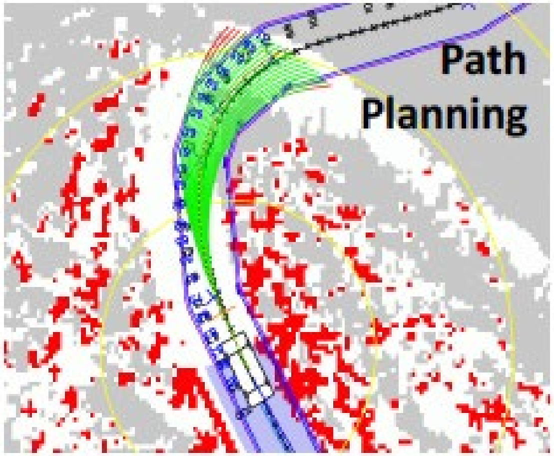
- Learn policy from user demonstrations

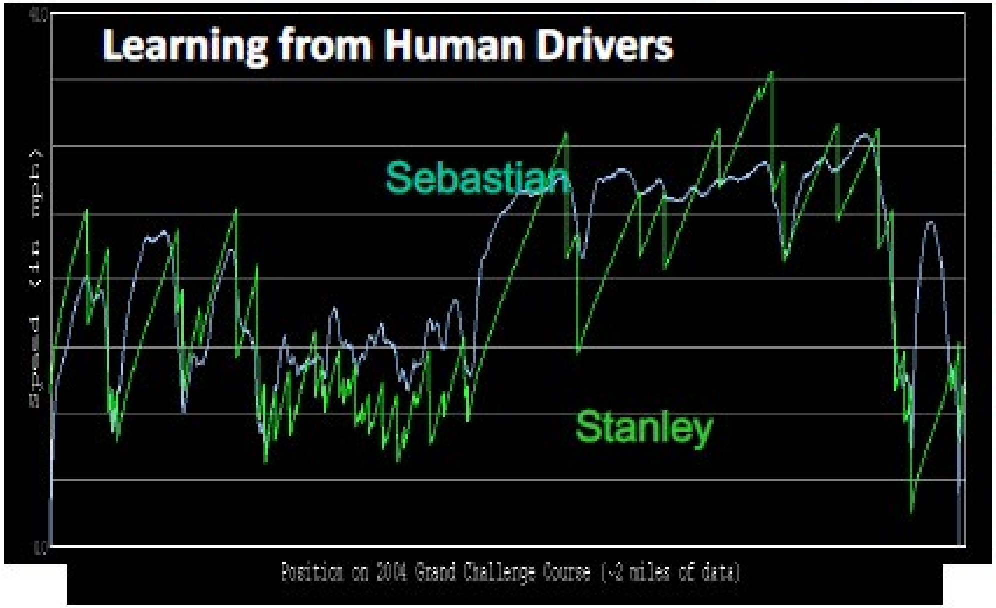
Stanford Autonomous Helicopter

http://heli.stanford.edu/

https://www.youtube.com/watch?v=VCdxqn0fcnE

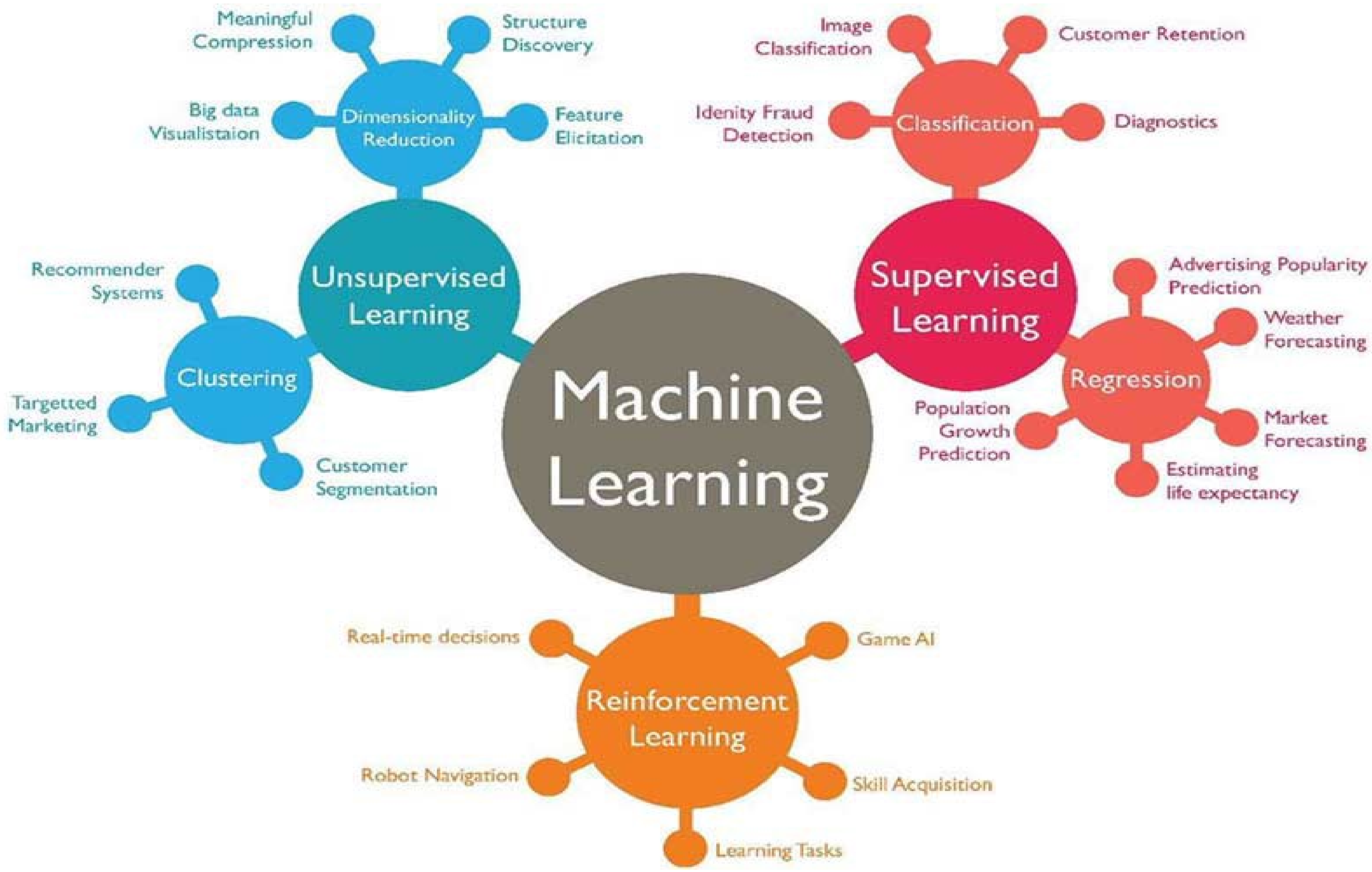
Autonomous Car Technology



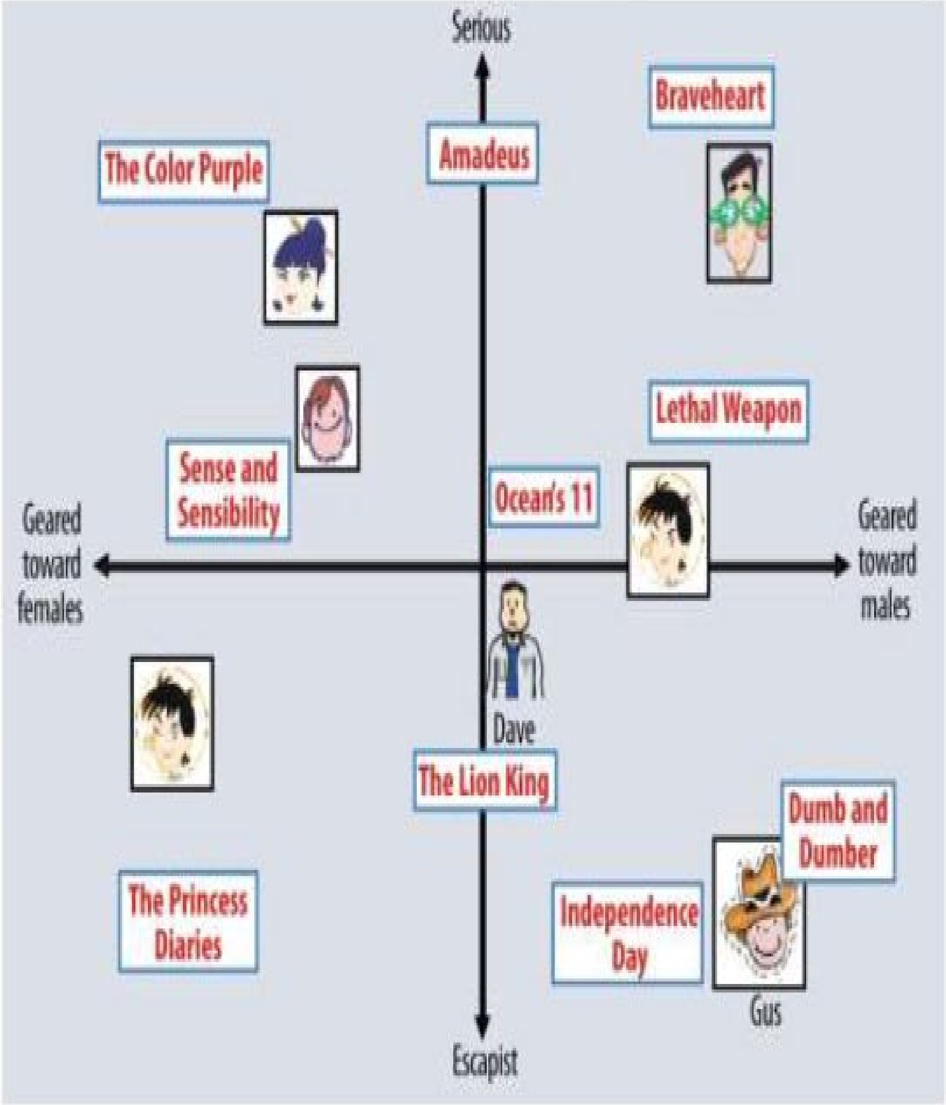


Images and movies taken from Sebastian Thrun's multimedia website.

Applications Overview

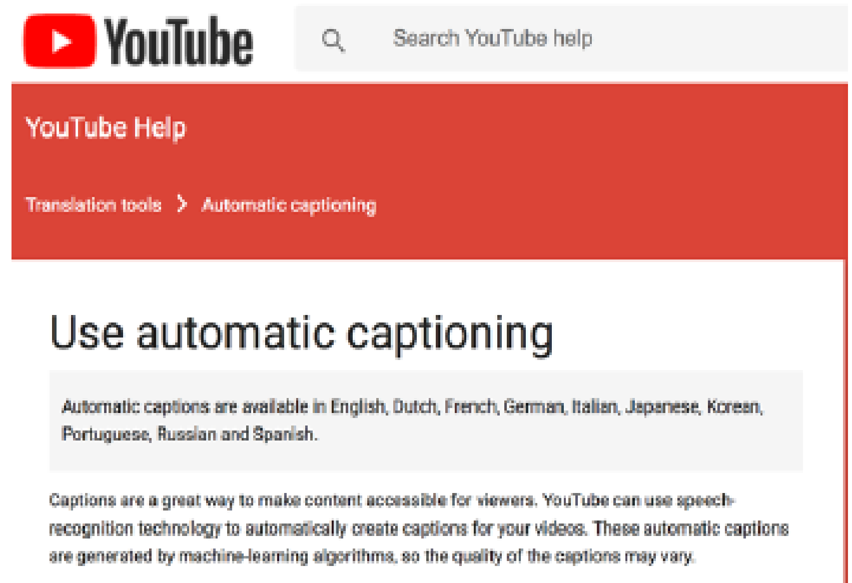


Unsupervised Learning – Recommended Systems



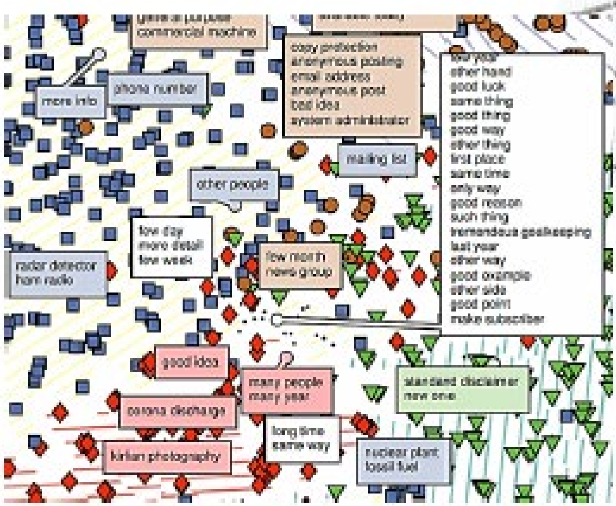
Successful (supervised) Applications

- Convert speech to text, translate from one language to the other.



Successful (unsupervised) Applications

- Document clustering and visualisation



Successful (reinforcement) Applications

- Game player, self-driving cars, trading strategy.

The story of AlphaGo so far

AlphaGo is the first computer program to defeat a professional human Go player, the first program to defeat a Go world champion, and arguably the strongest Go player in history.

AlphaGo's first formal match was against the reigning 3-times European Champion, Mr Fan Hui, in October 2015. Its 5-0 win was the first ever against a Go professional, and the results were published in full technical detail in the international journal, Nature. AlphaGo then went on to compete against legendary player Mr Lee Sedol, winner of 18 world titles and widely considered to be the greatest player of the past decade.

> More on The Future of Go Summit in this video

RE•WORK

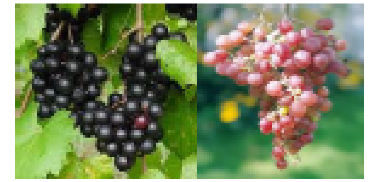
DEEP LEARNING IN FINANCE: LEARNING TO TRADE WITH Q-RL AND DQNS

Full Self-Driving Hardware on your Model S

All Tesla vehicles produced in our factory, including Model 3, have the hardware needed for full self-driving capability at a safety level substantially greater than that of a human driver.

Example: Wine Classification

- Wine experts identify the grape type by smelling and tasting the wine.



- The chemist says that wines derived from different grape types are different in terms of alcohol, malic acid, alcalinity of ash, magnesium, color intensity, etc.

- We get the measurements. But, too many numbers…

14.23,1.71,2.43,15.6,127,2.8,3.06,.28,2.29,5.64,1.04,3.92,1065

Can build a machine learning system
to automate grape type identification!

Example: Wine Classification

- Task: To identify the grape type of a wine sample based on the measured chemical quantities!

  - Collecting wine samples for each grape type.
  - Characterising each wine sample with 13 chemical features.

Feature Extraction

1) Alcohol
2) Malic acid
3) Ash
4) Alcalinity of ash
5) Magnesium
6) Total phenols
7) Flavanoids
8) Nonflavanoid phenols
9) Proanthocyanins
10)Color intensity
11)Hue
12)OD280/OD315 of diluted wines
13)Proline

Experiences

feature vectors

class labels

$$\mathbf{x}_1 = [x_{1,1}, x_{1,2}, x_{1,3}, \ldots, x_{1,12}, x_{1,13}], \quad y_1 = \text{grape type } 1$$

$$\mathbf{x}_2 = [x_{2,1}, x_{2,2}, x_{2,3}, \ldots, x_{2,12}, x_{2,13}], \quad y_2 = \text{grape type } 2$$

$$\mathbf{x}_3 = [x_{3,1}, x_{3,2}, x_{3,3}, \ldots, x_{3,12}, x_{3,13}], \quad y_3 = \text{grape type } 2$$

$$\vdots \qquad \qquad \vdots$$

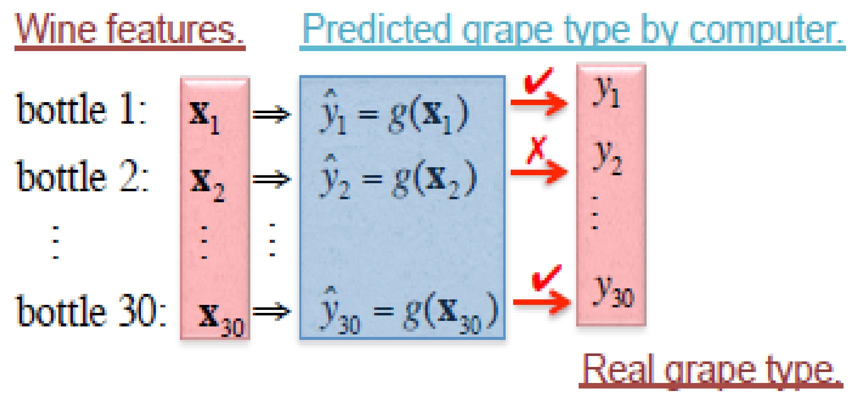
$$\mathbf{x}_{30} = [x_{30,1}, x_{30,2}, x_{30,3}, \ldots, x_{30,12}, x_{30,13}], \quad y_{30} = \text{grape type } 1$$

30 bottles in total, 10 bottles for each tree type, each bottle is charaterised by 13 features.

Example: Wine Classification

- Design a mathematical model to predict the grape type. The model below is controlled by 14 parameters: $[w_1, w_2, \ldots, w_{13}, b]$

$$\hat{y} = g(\mathbf{x}) = \begin{cases} \text{type 1}, & \text{if } \sum_{i=1}^{13} w_i x_i + b \geq 0 \\ \text{type 2}, & \text{if } \sum_{i=1}^{13} w_i x_i + b < 0 \end{cases}$$



- System training is the process of finding the best model parameters by minimising a loss function.

$$[w_1^*, w_2^*, \ldots, w_{13}^*, b^*] = \underset{w_1, w_2, \ldots, w_{13}, b}{\arg\min} \; O_{loss}(w_1, w_2, \ldots, w_{13}, b)$$

Loss: predictive error

Example: Wine Classification

- Now, given an unseen bottle of wine: