COS30018 – Week 4:
Introduction to Machine Learning (ML)

SWINBURNE UNIVERSITY OF TECHNOLOGY

1

Machine Learning (ML)

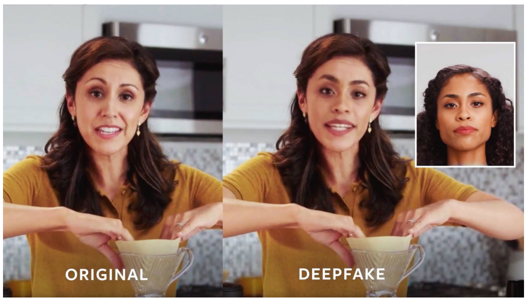
SWINBURNE UNIVERSITY OF TECHNOLOGY

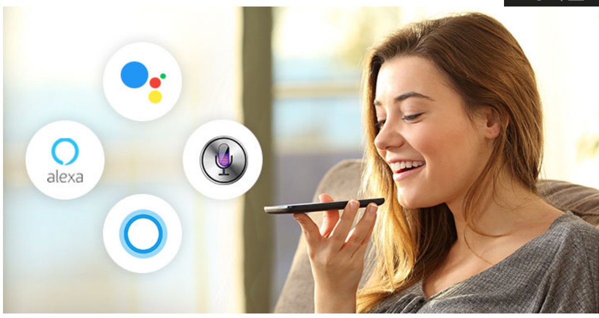
Machine Learning is one of today's most exciting technologies:

- Powering a majority of AI systems
- Democratizing technologies & AI

2

What is Machine Learning? Examples



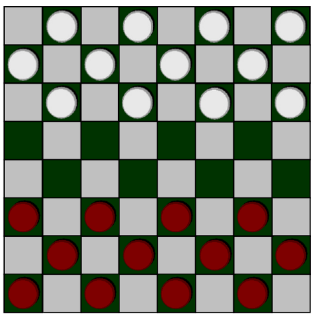


3

What is Machine Learning? Definition

Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed.

(Arthur Samuel, 1959)



Samuel's Checkers program (approx. 1952 – 1956) uses a combination of:

- Rote learning ~ memorizing the game board positions and storing their values
- Generalization learning ~ updating the reward functions as more board positions are discovered and assigned a value (via self-playing)
- Self-playing ~ generating vast amount of training data

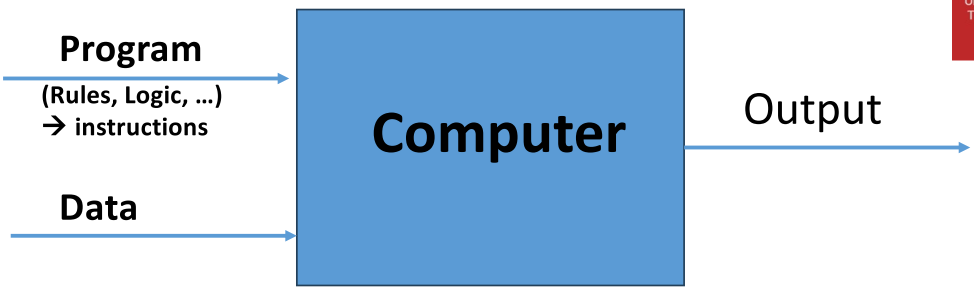
4

Questions for AI researchers, developers & practitioners:

5

Traditional programming

SWINBURNE UNIVERSITY OF TECHNOLOGY



6

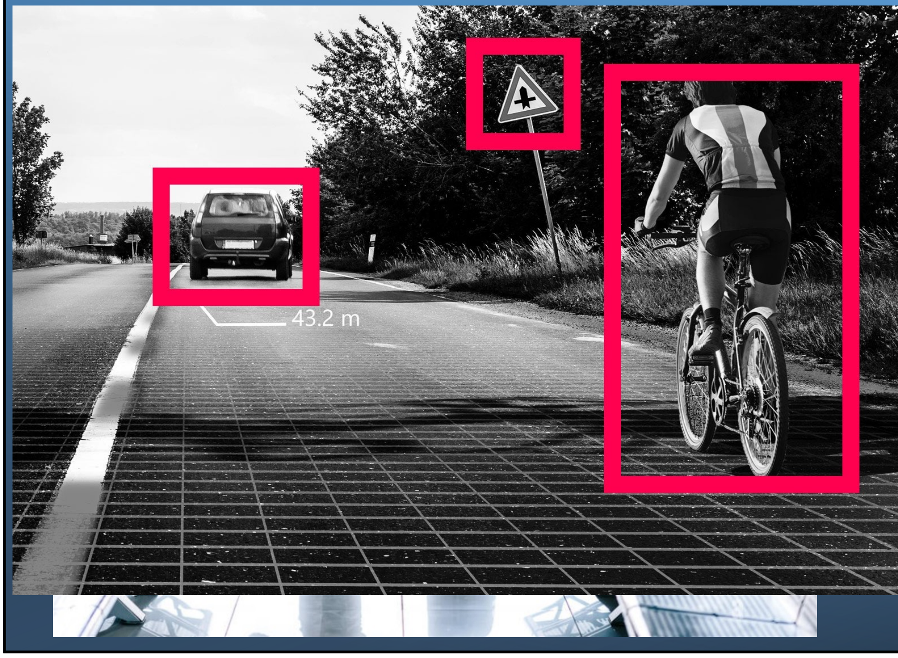
Issues with traditional approach

7

Issues with traditional approach

- What are the rules & logic you use to recognise your friend??

8



SWiNBURNE
SWINBURNE UNIVERSITY OF TECHNOLOGY

Issues with traditional approach

- What are the rules & logic you use to recognise your friend??

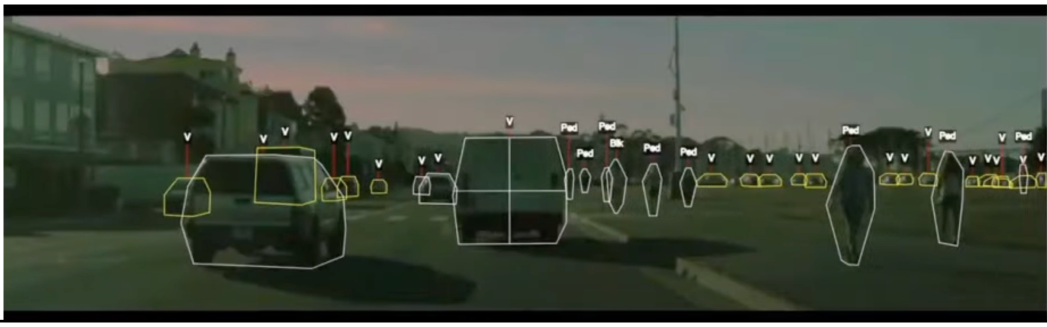
9

Machine Learning: Our hero??

SWiNBURNE
SWINBURNE UNIVERSITY OF TECHNOLOGY

Data → **Computer** → "Program"

Output → **Computer**



10

Machine Learning (ML):

Our hero??

Data

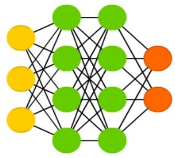
Output

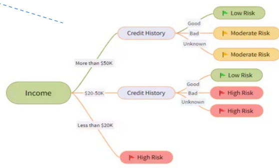
Computer

(ML Algorithm)

"Program"

(ML Model)





$$y = f(x_1, x_2, x_3, x_4) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

11

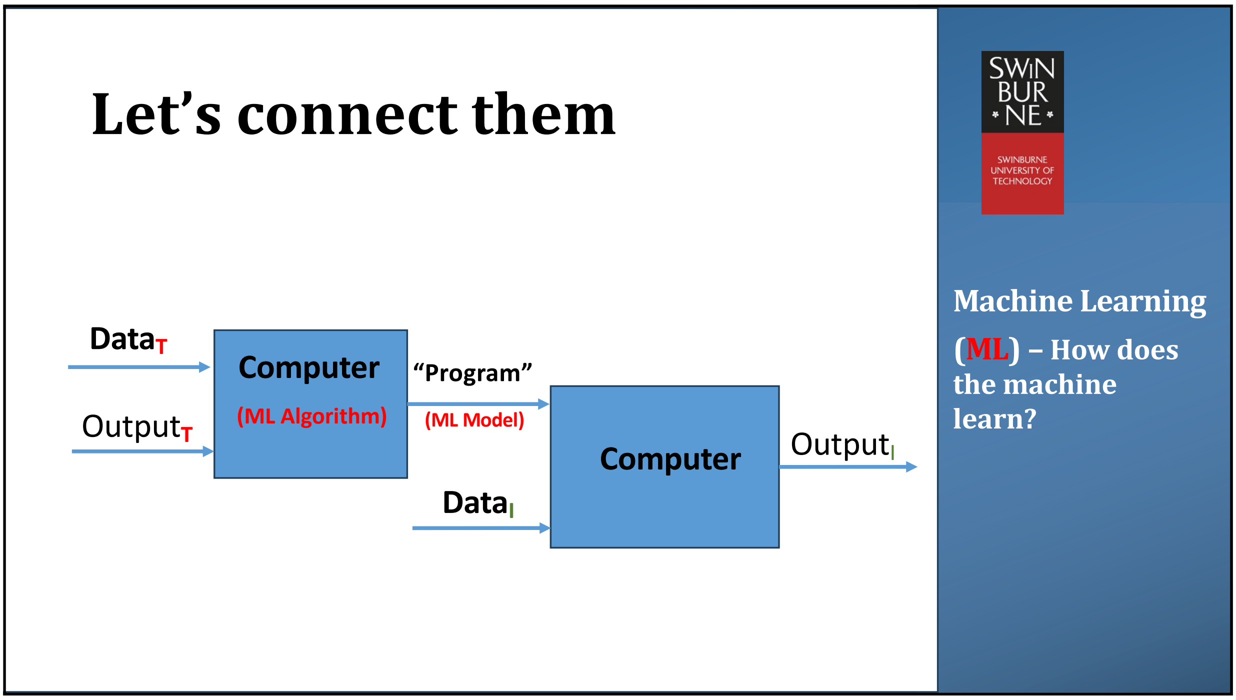
Machine Learning (ML) – How does the machine learn?

- A typical problem

12

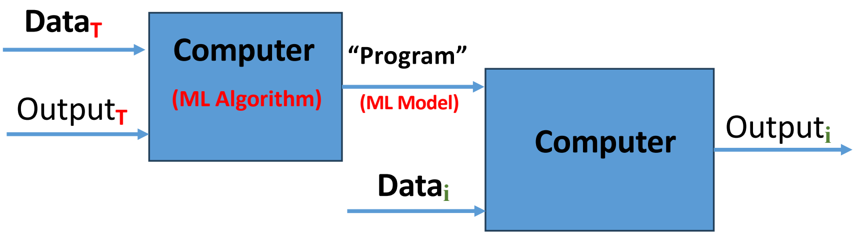

Leonardo DiCaprio
Ben Affleck
Taylor Swift
SWINBURNE UNIVERSITY OF TECHNOLOGY
Machine Learning (ML) – How does the machine learn?
• A typical problem

13


Let's connect them
$Data_T$
$Output_T$
Computer
(ML Algorithm)
"Program"
(ML Model)
Computer
$Data_I$
$Output_I$
SWINBURNE UNIVERSITY OF TECHNOLOGY
Machine Learning (ML) – How does the machine learn?

14

Let’s connect them



The most well-known models currently?
They are known as **Large Foundation Models**/**Large Language Models**, such as models from the **GPT**’s family, **Gemini**’s family, **Grok**’s family, **Llama**’s family

SWINBURNE UNIVERSITY OF TECHNOLOGY

Machine Learning (ML) – How does the machine learn?

15

SWINBURNE UNIVERSITY OF TECHNOLOGY

Machine Learning (ML) – How does the machine learn?

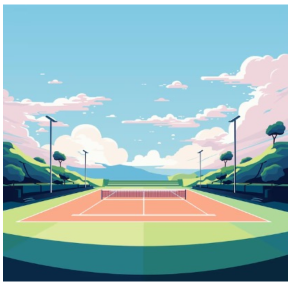
- From a dataset consisting of many examples
- From the experiences it gains from the direct interaction with environment

16

Machine Learning (ML) – How does the machine learn?

- From a dataset consisting of many examples
- From the experiences it gains from the direct interaction with environment

17



Machine Learning (ML) – Basic concepts

- A simple problem – When to play tennis??

18

Source:
Tom Mitchell, 1997

Example	outlook	temp	humidity	windy	play
0	sunny	hot	high	False	no
1	sunny	hot	high	True	no
2	overcast	hot	high	False	yes
3	rainy	mild	high	False	yes
4	rainy	cool	normal	False	yes
5	rainy	cool	normal	True	no
6	overcast	cool	normal	True	yes
7	sunny	mild	high	False	no
8	sunny	cool	normal	False	yes
9	rainy	mild	normal	False	yes
10	sunny	mild	normal	True	yes
11	overcast	mild	high	True	yes
12	overcast	hot	normal	False	yes
13	rainy	mild	high	True	no

SWINBURNE UNIVERSITY OF TECHNOLOGY

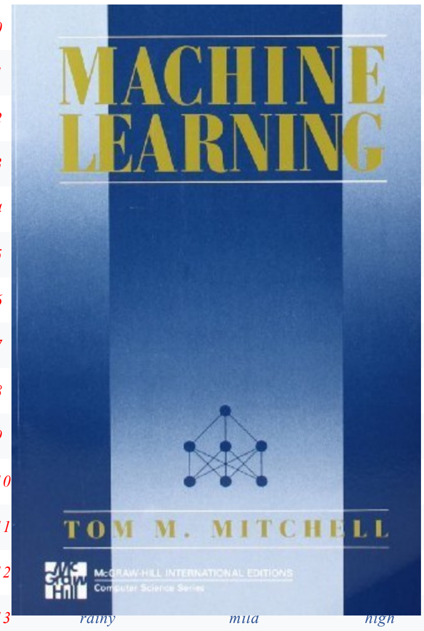
Machine Learning (ML) – Basic concepts

- A simple problem – When to play tennis??

19

Source:
Tom Mitchell, 1997

Example	outlook	temp	humidity	windy	play
0				False	no
1				True	no
2				False	yes
3				False	yes
4				False	yes
5				True	no
6				True	yes
7				False	no
8				False	yes
9				False	yes
10				True	yes
11				True	yes
12				False	yes
13	rainy	mild	high	True	no



SWINBURNE UNIVERSITY OF TECHNOLOGY

Machine Learning (ML) – Basic concepts

- A simple problem – When to play tennis??

20

Features
(Attributes)

Example	outlook	temp	humidity	windy	play
0	sunny	hot	high	False	no
1	sunny	hot	high	True	no
2	overcast	hot	high	False	yes
3	rainy	mild	high	False	yes
4	rainy	cool	normal	False	yes
5	rainy	cool	normal	True	no
6	overcast	cool	normal	True	yes
7	sunny	mild	high	False	no
8	sunny	cool	normal	False	yes
9	rainy	mild	normal	False	yes
10	sunny	mild	normal	True	yes
11	overcast	mild	high	True	yes
12	overcast	hot	normal	False	yes
13	rainy	mild	high	True	no

Source:
Tom Mitchell, 1997

SWINBURNE UNIVERSITY OF TECHNOLOGY

Machine Learning (ML) – Basic concepts

- **A simple problem – When to play tennis??**

21

Features
(Attributes)

Target
(Labels)

Example	outlook	temp	humidity	windy	play
0	sunny	hot	high	False	no
1	sunny	hot	high	True	no
2	overcast	hot	high	False	yes
3	rainy	mild	high	False	yes
4	rainy	cool	normal	False	yes
5	rainy	cool	normal	True	no
6	overcast	cool	normal	True	yes
7	sunny	mild	high	False	no
8	sunny	cool	normal	False	yes
9	rainy	mild	normal	False	yes
10	sunny	mild	normal	True	yes
11	overcast	mild	high	True	yes
12	overcast	hot	normal	False	yes
13	rainy	mild	high	True	no

Source:
Tom Mitchell, 1997

SWINBURNE UNIVERSITY OF TECHNOLOGY

Machine Learning (ML) – Basic concepts

- **A simple problem – When to play tennis??**

22

Example	outlook	temp	humidity	windy	play
0	sunny	hot	high	False	no
1	sunny	hot	high	True	no
2	overcast	hot	high	False	yes
3	rainy	mild	high	False	yes
12	overcast	hot	normal	False	yes
13	rainy	mild	high	True	no

Features (Attributes)

Target (Labels)

In ML lingua:

- Each example *i* is expressed as a pair (X^i, y^i), where $X^i = (X^i_1, ..., X^i_k)$ is also called ***Input*** and the Target y^i is also called ***Output.*** (Here k is the number of features)
- For example, in the "Play tennis" problem: Example 0 can be expressed as (X^0, y^0), where $X^0 = (X^0_{outlook}$ = sunny, X^0_{temp} = hot, $X^0_{humidity}$ = high, X^0_{windy} = False) and y^0 = no.

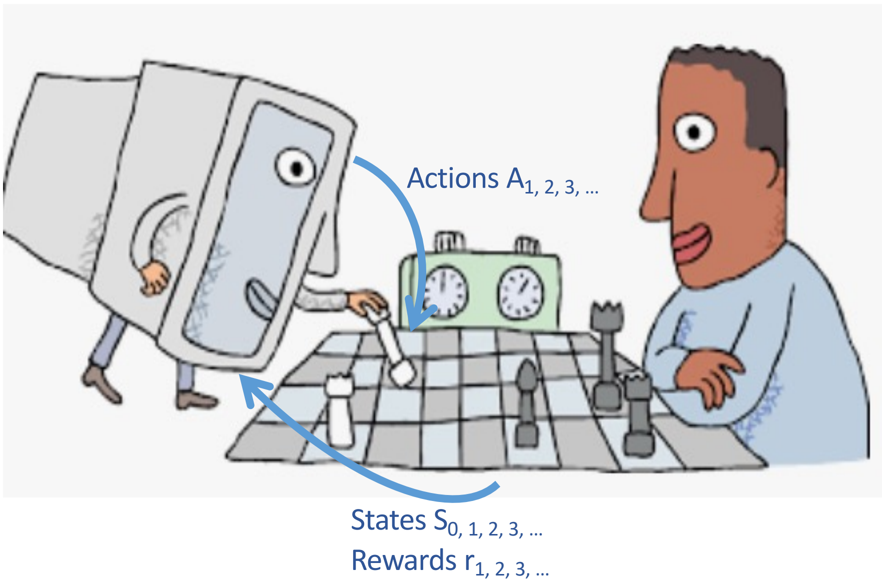
Source:
Tom Mitchell, 1997

SWINBURNE UNIVERSITY OF TECHNOLOGY

Machine Learning (ML) – Basic concepts

- **A simple problem – When to play tennis??**

23

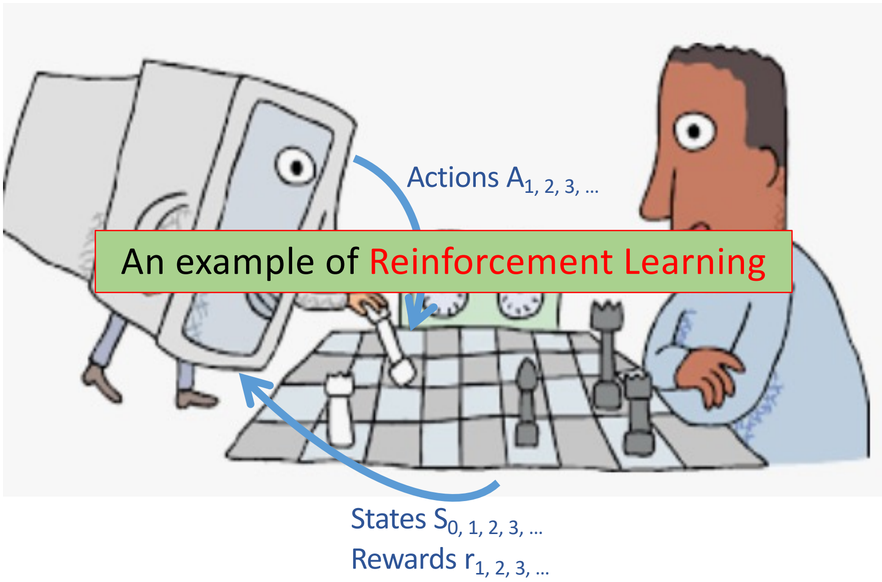


SWINBURNE UNIVERSITY OF TECHNOLOGY

Machine Learning (ML) – Basic concepts

- **A machine that plays chess**

24



Machine Learning (ML) – Basic concepts

- **A machine that plays chess**

25

Types of Learning

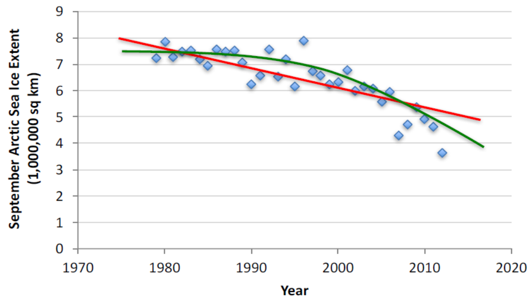
- Supervised (inductive) learning
  - Given: training data + desired outputs (labels)
- Unsupervised learning
  - Given: training data (without desired outputs)
- Semi-supervised learning
  - Given: training data + a few desired outputs
- Reinforcement learning
  - Rewards from sequence of actions

26

Supervised Learning: Regression and Classification

Given the dataset of observations $(x_1, t_1), (x_2, t_2), \ldots, (x_n, t_n)$

- Learn a function $f(x)$ to predict the target variable t given x



Data from G. Witt. Journal of Statistics Education, Volume 21, Number 1 (2013)

t is real-valued == regression

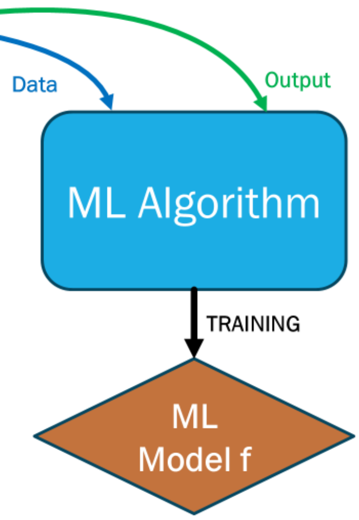
Example	outlook	temp	humidity	windy	play
0	sunny	hot	high	False	no
1	sunny	hot	high	True	no
2	overcast	hot	high	False	yes
3	rainy	mild	high	False	yes
4	rainy	cool	normal	False	yes
5	rainy	cool	normal	True	no
6	overcast	cool	normal	True	yes
7	sunny	mild	high	False	no
8	sunny	cool	normal	False	yes
9	rainy	mild	normal	False	yes
10	sunny	mild	normal	True	yes
11	overcast	mild	high	True	yes
12	overcast	hot	normal	False	yes
13	rainy	mild	high	True	no

t is categorical == classification

27

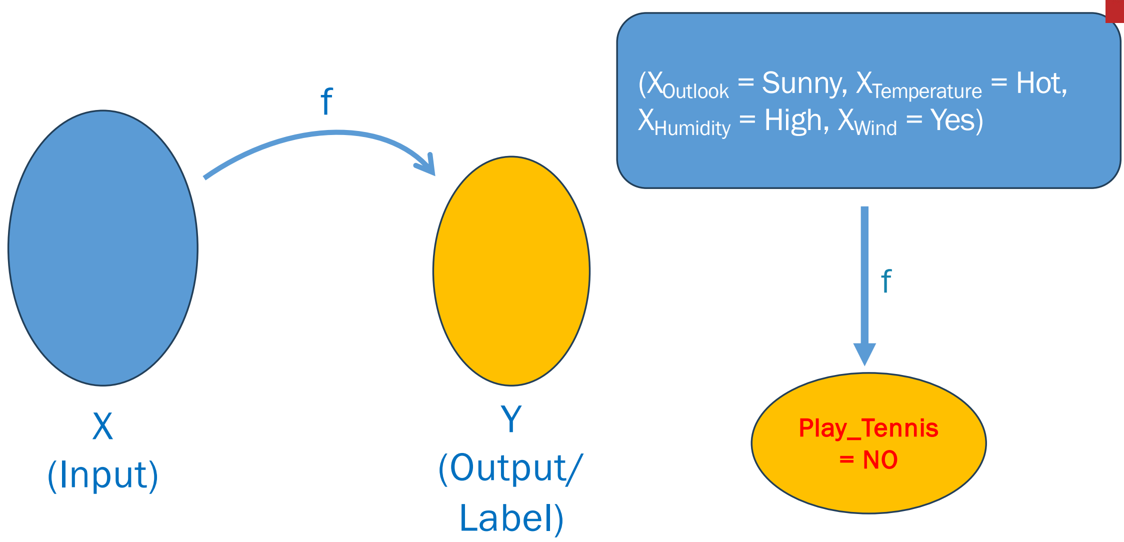
SUPERVISED LEARNING

	F_1	F_2	.....	F_k	T
Obs^1	Obs^1_1	Obs^1_2		Obs^1_k	t^1
Obs^2	Obs^2_1	Obs^2_2		Obs^2_k	t^2
...					...
...					...
Obs^i	Obs^i_1	Obs^i_2		Obs^i_k	t^i
...					
Obs^n	Obs^n_1	Obs^n_2		Obs^n_k	t^n



28

SUPERVISED LEARNING



29

SUPERVISED LEARNING

Input (X)	Output (Y)	Application
email →	spam? (0/1)	spam filtering
audio →	text transcripts	speech recognition
English →	Spanish	machine translation
ad, user info →	click? (0/1)	online advertising
image, radar info →	position of other cars	self-driving car
image of phone →	defect? (0/1)	visual inspection

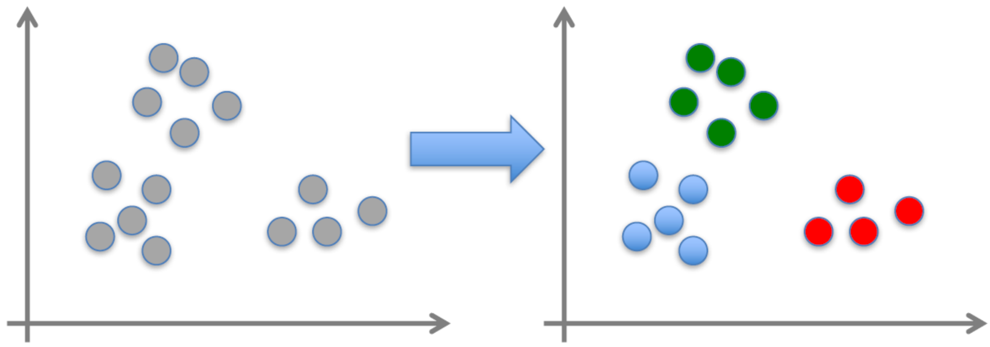
Source: Andrew Ng/Stanford Online

30

Unsupervised Learning

Given $x^1, x^2, \ldots, x^n$ (without labels)

- Output hidden structure behind the x's
  - E.g., clustering

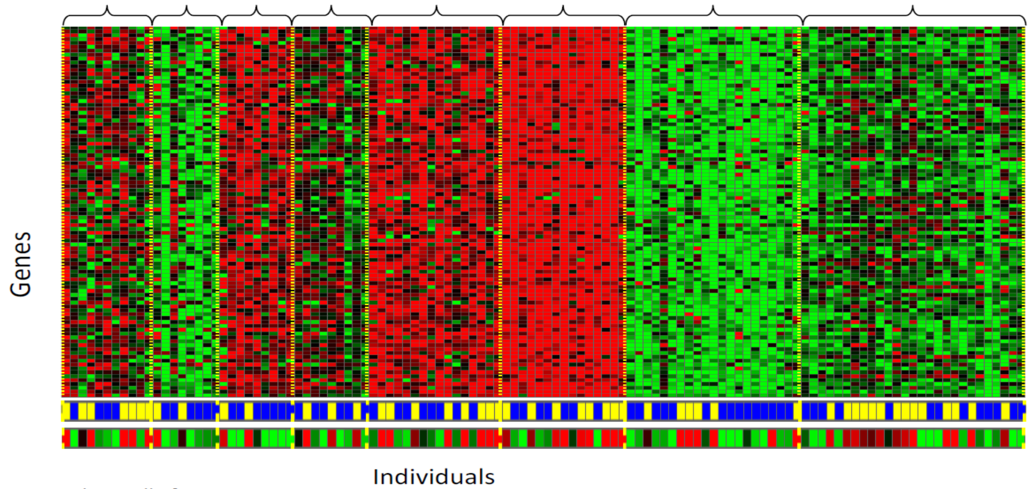


E.g., https://news.google.com uses clustering

31

Unsupervised Learning

Genomics application: group individuals by genetic similarity



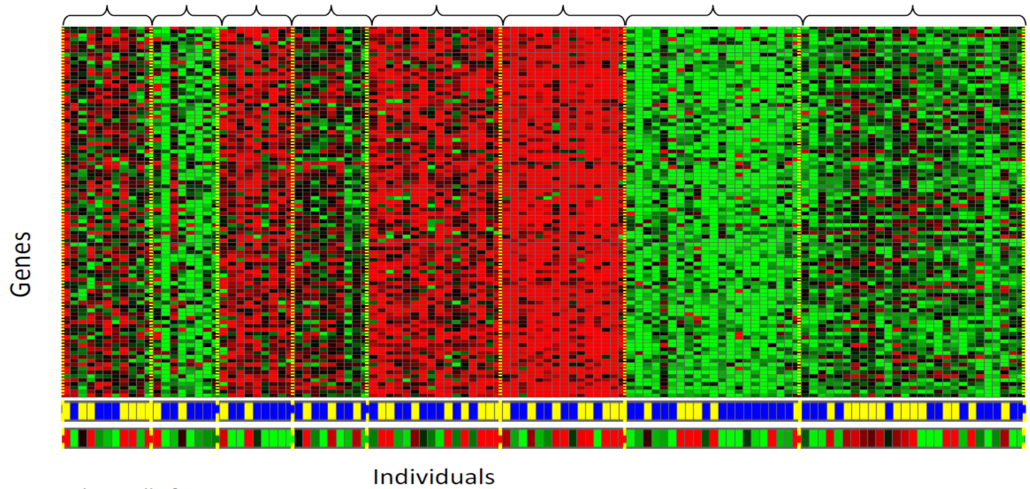
[Source: Daphne Koller]

32

Unsup

Genomics

If members of a genomics group have a higher chance of getting a disease X then an individual from the same group should be monitored for X to ensure that the condition can be detected early.



[Source: Daphne Koller]

33

Semi-supervised Learning

Can you build a spam filter for emails?

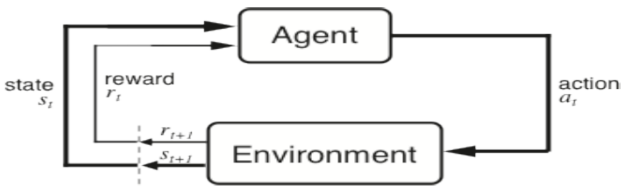
34

Semi-supervised Learning

The given dataset does not have a label for every example

35

Reinforcement Learning: The Agent-Environment Interface



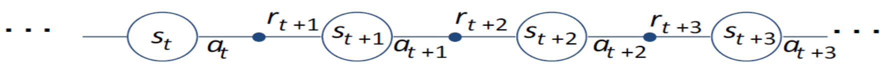
Agent and environment interact at discrete time steps : $t = 0, 1, 2, K$

Agent observes state at step t: $s_t \in S$

produces action at step t: $a_t \in A(s_t)$

gets resulting reward : $r_{t+1} \in \Re$

and resulting next state : s_{t+1}



Slide credit: Sutton & Barto

36

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

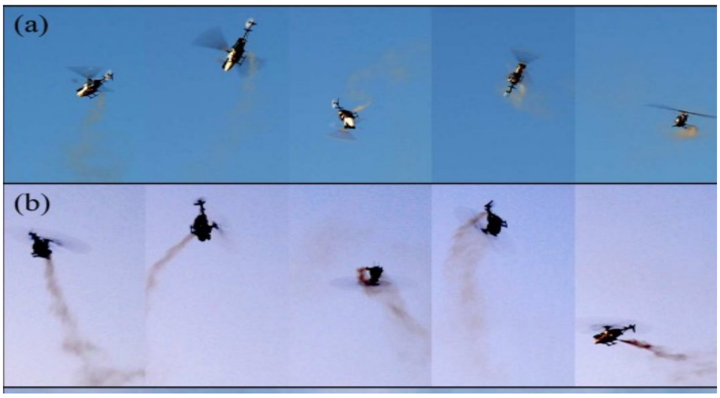
Examples:

– Credit assignment problem
– Game playing
– Robot in a maze
– Balance a pole on your hand

37

Reinforcement Learning

Learn policy from user demonstrations

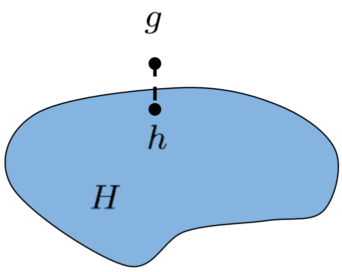


Stanford Autonomous Helicopter
http://heli.stanford.edu/

38

Learning from a Dataset of Examples: Inductive Learning

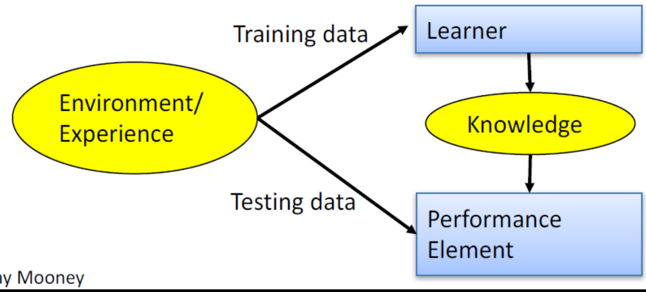
- Simplest form: learn a function from examples
  - A target function: g
  - Examples: input-output pairs $(x, g(x))$
  - E.g. x is an email and $g(x)$ is spam / ham
  - E.g. x is a house and $g(x)$ is its selling price
- Problem:
  - Given a hypothesis space H
  - Given a training set of examples x_i
  - Find a hypothesis $h(x)$ such that $h \sim g$
- Includes:
  - Classification (outputs = class labels)
  - Regression (outputs = real numbers)



39

Designing a Learning System

- Choose the training experience
- Choose exactly what is to be learned – i.e. the ***target function***
- Choose how to represent the target function
- Choose a learning algorithm to infer the target function from the experience



Based on slide by Ray Mooney

40

Training vs. Test Distribution

- We generally assume that the training and test examples are independently drawn from the same overall distribution of data
  - We call this "i.i.d" which stands for "independent and identically distributed"
- If examples are not independent, requires ***collective classification***
- If test distribution is different, requires ***transfer learning***

41

ML in a Nutshell

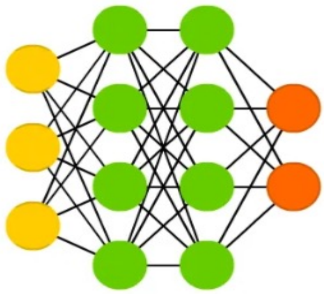
- Tens of thousands of machine learning algorithms
  - Hundreds new every year
- Every ML algorithm has three components:
  - **Representation**
  - **Optimization**
  - **Evaluation**

42

Various Function Representations

SWINBURNE UNIVERSITY OF TECHNOLOGY

- Numerical functions
  - Linear regression
  - Neural networks
  - Support vector machines



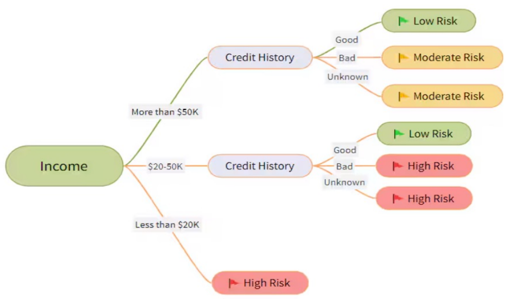
$$y = f(x_1, x_2, x_3, x_4) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

43

Various Function Representations

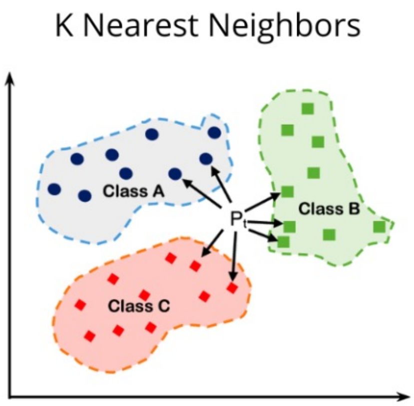
SWINBURNE UNIVERSITY OF TECHNOLOGY

- Numerical functions
  - Linear regression
  - Neural networks
  - Support vector machines
- Symbolic functions
  - Decision trees
  - Rules in propositional logic
  - Rules in first-order predicate logic



44

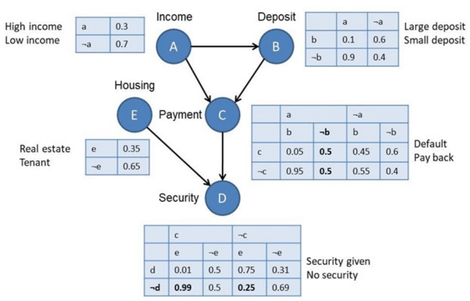
Various Function Representations

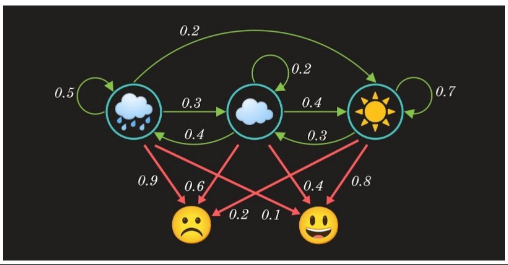


- Instance-based functions
  - Nearest-neighbor
  - Case-based
- Probabilistic Graphical Models
  - Naïve Bayes
  - Bayesian networks
  - Hidden-Markov Models (HMMs)
  - Probabilistic Context Free Grammars (PCFGs)
  - Markov networks

45

Various Function Representations





- Instance-based functions
  - Nearest-neighbor
  - Case-based
- Probabilistic Graphical Models
  - Naïve Bayes
  - Bayesian networks
  - Hidden-Markov Models (HMMs)
  - Probabilistic Context Free Grammars (PCFGs)
  - Markov networks

46

Various Search/Optimization Algorithms

- Gradient descent
  - Perceptron
  - Backpropagation
- Dynamic Programming
  - HMM Learning
  - PCFG Learning
- Divide and Conquer
  - Decision tree induction
  - Rule learning
- Evolutionary Computation
  - Genetic Algorithms (GAs)
  - Genetic Programming (GP)
  - Neuro-evolution

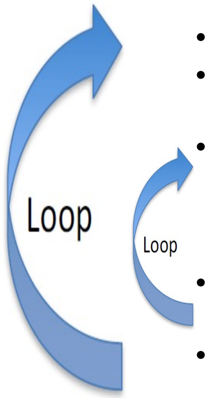
47

Evaluation

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

48

ML in Practice



- Understand domain, prior knowledge, and goals
- Data gathering, data integration, selection, cleaning, pre-processing, etc.
- Learn models
  - Model Selection
  - Model Training
  - Model Evaluation
- Interpret results
  - Hyperparameter tuning
- Consolidate and deploy discovered knowledge

49

Summary

- Learning can be viewed as using direct or indirect experience to approximate a chosen target function.
- Function approximation can be viewed as a search through a space of hypotheses (representations of functions) for one that best fits a set of training data.
- Different learning methods assume different hypothesis spaces (representation languages) and/or employ different search techniques.

50