MachineLearnAthon - Microlecture Introduction to Machine Learning

Recorded by Lara Kuhlmann

MachineLearnAthon
A project Co-funded by the Erasmus+ programme of the European Union

Learning outcomes of today

After successfully completing this micro-lecture, you are able to….

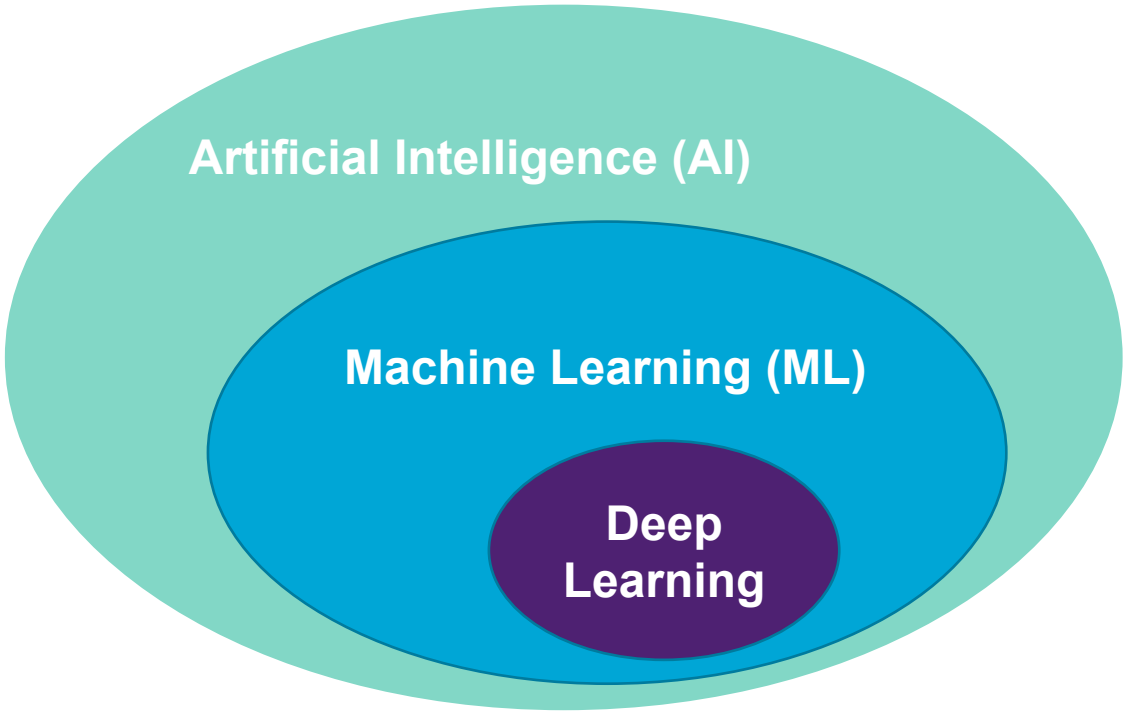
- Distinguish the terms artificial intelligence, machine learning and deep learning
- Explain the difference between supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning
- Identify the matching machine learning paradigm for a use case

Agenda for today

- Definition of Machine Learning (ML)
- Overview of ML paradigms
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Definition of Machine Learning

“Machine learning is the technique that improves system performance by learning from experience via computational methods. In computer systems, experience exists in the form of data, and the main task of machine learning is to develop learning algorithms that build models from data.”[1]



[1] Zhou, Z. H. (2021). *Machine learning*. Springer nature., p. 2.
Mirtaheri, S. L., & Shahbazian, R. (2022). *Machine learning: theory to applications*. CRC Press.

Overview

ML paradigms	Common tasks
Supervised Learning	Classification
Semi-Supervised Learning	Outlier detection
Unsupervised Learning	Clustering
Reinforcement Learning	Game playing

Question of a producing company:
Is the pallet that my supplier sent me damaged?

Question of an IT service provider:
Which emails should be marked as spam?

Question of a marketing agency:
What customer groups should I target separately?

Question of a chess player:
What should my next move be?

Mirtaheri, S. L., & Shahbazian, R. (2022). *Machine learning: theory to applications*. CRC Press.

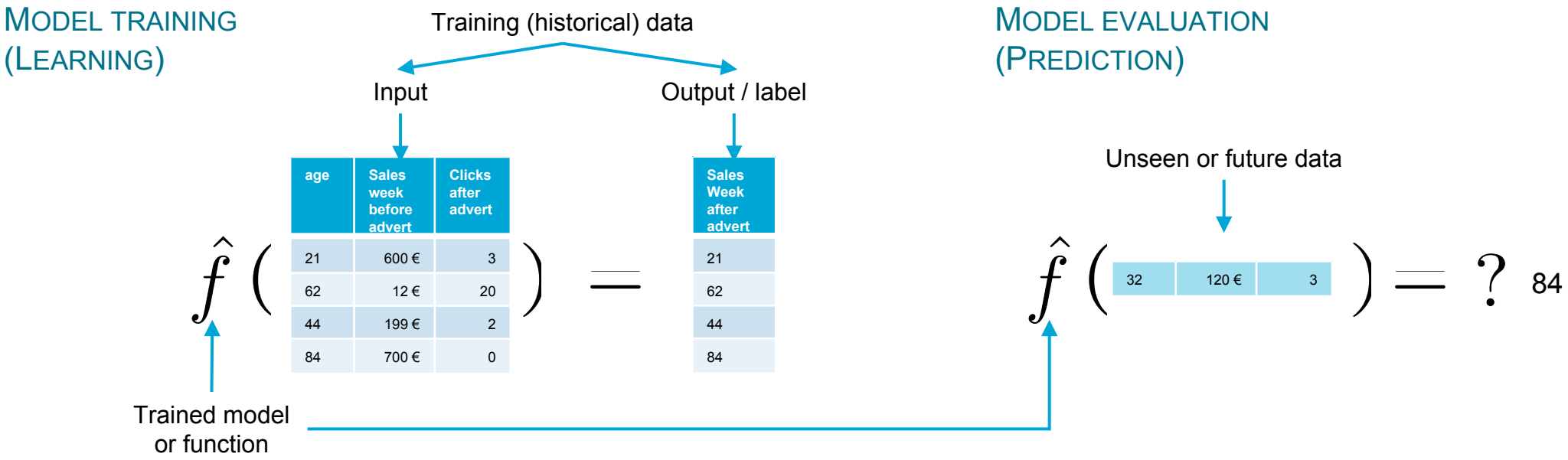
technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Supervised Learning

Input: Historical input data with a known result or output (label). Output (label) can have the form of a numerical value or category.

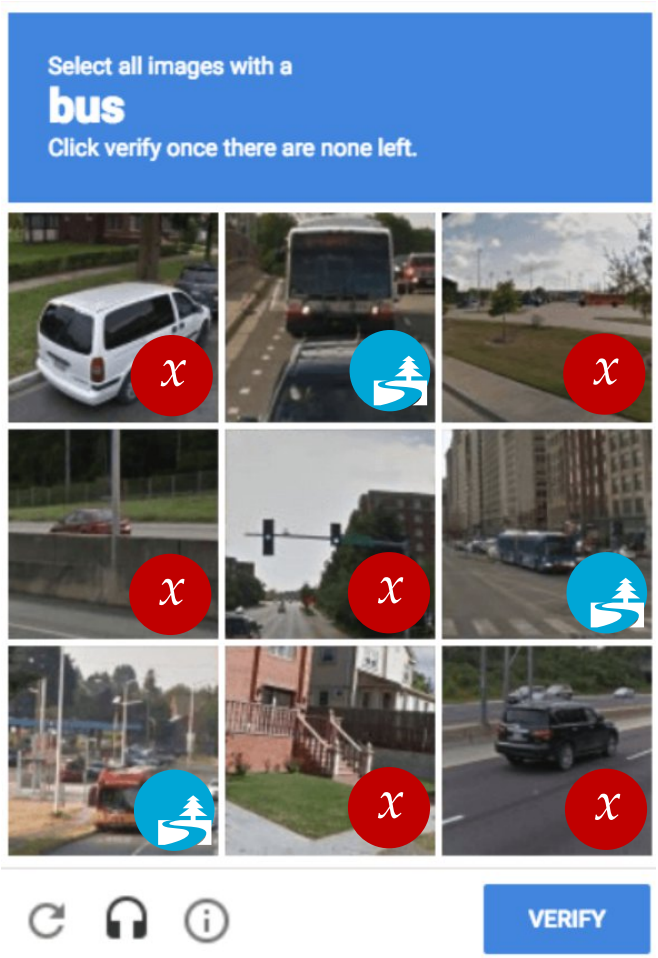
Goal: Predict the output for unseen or future data.

Approach: Learn a function (model) that describes the relationship (or mapping) between the historic input and the output variables and apply this function to the unseen or future data.



Igual, L., & Seguí, S. (2024). Introduction to data science. In *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications* (pp. 1-4). Cham: Springer International Publishing.

Supervised Learning: Classification



Known (historical) data

Input X → Label y

pxl_1	pxl_2	...	$pxl_{n \times m}$	Bus
0	90	...	77	Yes
234	48	...	159	No
...	...	...	...	...
180	50	...	42	No
84	254	...	41	No

$\hat{f}$ ( ) = ...

Unseen (future) data

pxl_1	pxl_2	...	$pxl_{n \times m}$	Bus
0	90	...	77	?

$\hat{f}$ ( ) =

Given:

- Samples $X := (x_1, ... x_n)$
- Categorial Labels $y := (y_1, ..., y_n)$

Goal:

- Given *new* sample X', predict y'

Training/Learning:

- Find mapping $\hat{f}$ such that
- $\hat{f}(X) = y$

Igual, L., & Seguí, S. (2024). Introduction to data science. In *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications* (pp. 1-4). Cham: Springer International Publishing.

technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Supervised Learning: Regression

Training (historical) data

Input X → Label y

n_{pieces}	$diff.$	...	rating	Price
700	hard	...	5/5	126,99
234	easy	...	5/5	69,99
...	...	...	...	...
180	medium	...	4/5	15,99
84	extreme	...	3/5	13,92

$\hat{f}$ ( ) =

Unseen (future) data

n_{pieces}	$diff.$	...	rating	Price
658	hard	...	4/5	?

$\hat{f}$ ( ) =

Given:

- Samples $X := (x_1, \ldots x_n)$
- **Numeric** Labels $y := (y_1, \ldots, y_n)$

Goal:

- Given *new* sample X, predict y'

Training/Learning:

- Find mapping $\hat{f}$ such that
- $\hat{f}(X) = y$

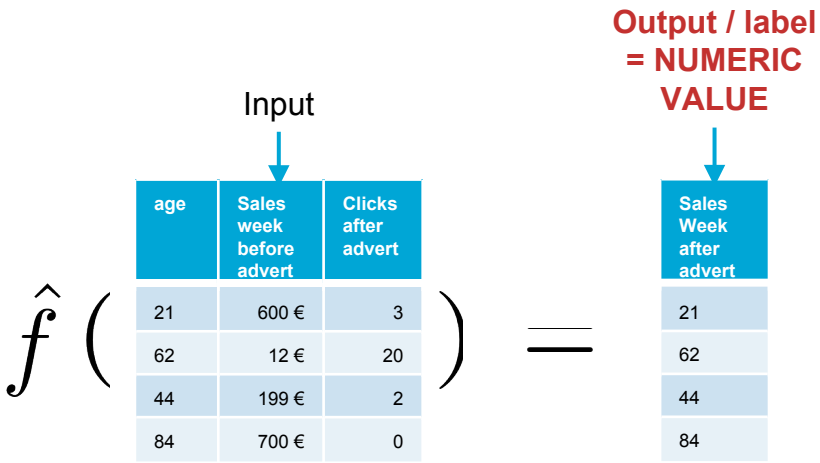
https://www.amazon.de/lego/s?k=lego

Igual, L., & Seguí, S. (2024). Introduction to data science. In *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications* (pp. 1-4). Cham: Springer International Publishing.

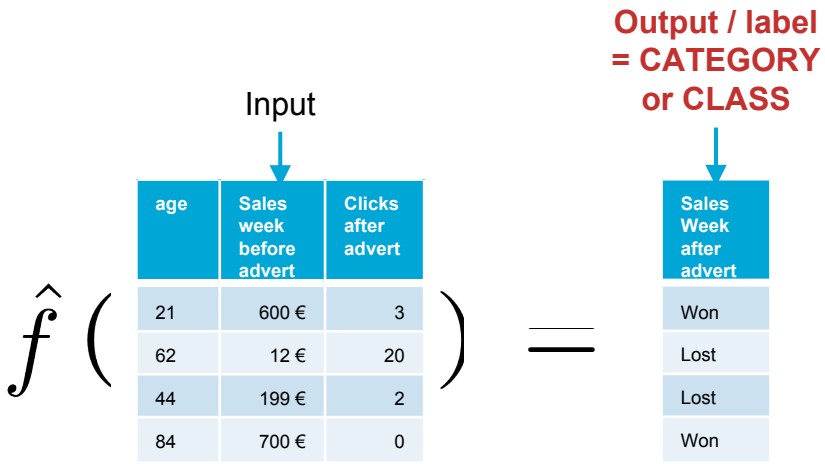
technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Supervised Learning: Regression and Classification

REGRESSION



CLASSIFICATION



Two or more classes!

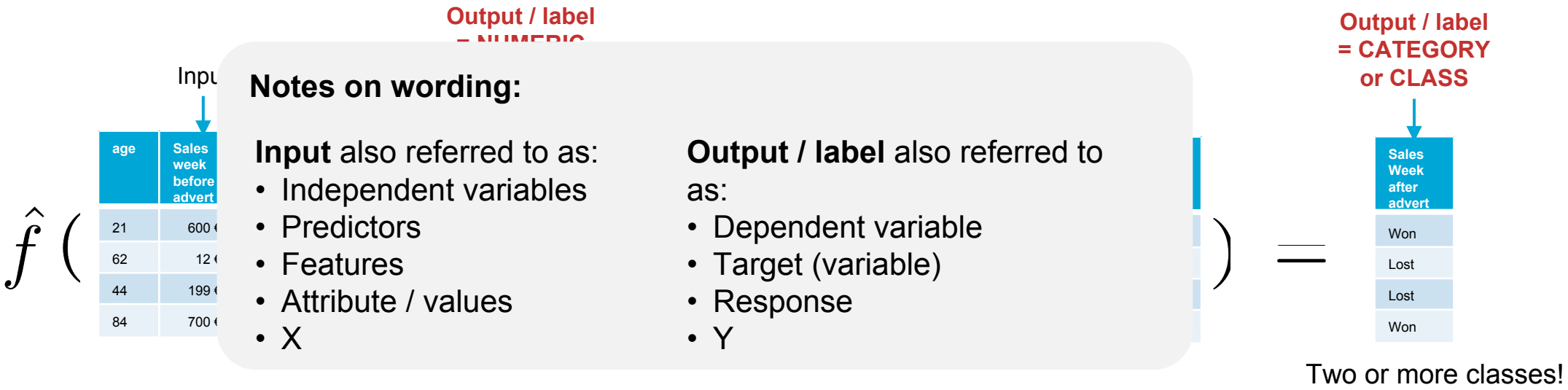
Igual, L., & Seguí, S. (2024). Introduction to data science. In *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications* (pp. 1-4). Cham: Springer International Publishing.

Co-funded by the Erasmus+ programme of the European Union

Supervised Learning: Regression and Classification

REGRESSION

CLASSIFICATION



Notes on wording:

Input also referred to as:
- Independent variables
- Predictors
- Features
- Attribute / values
- X

Output / label also referred to as:
- Dependent variable
- Target (variable)
- Response
- Y

Igual, L., & Seguí, S. (2024). Introduction to data science. In *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications* (pp. 1-4). Cham: Springer International Publishing.

Co-funded by the Erasmus+ programme of the European Union

Unsupervised Learning

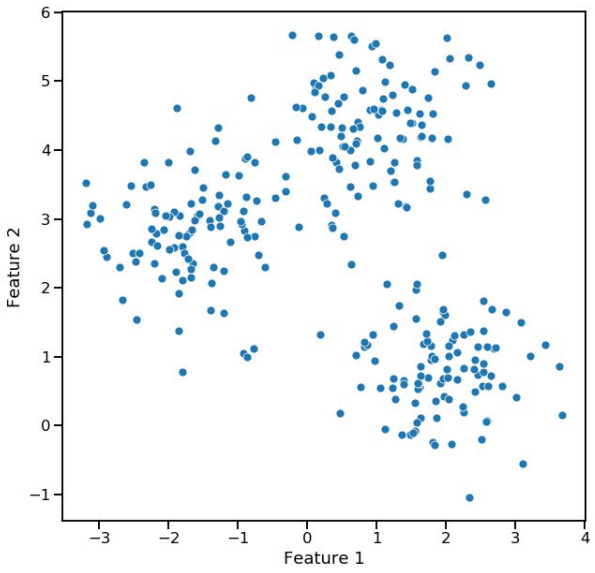
Input: Historical input data without a known result.

Goal: Build meaningful subgroups (clusters), find associations or compress data (reduce dimensionality).

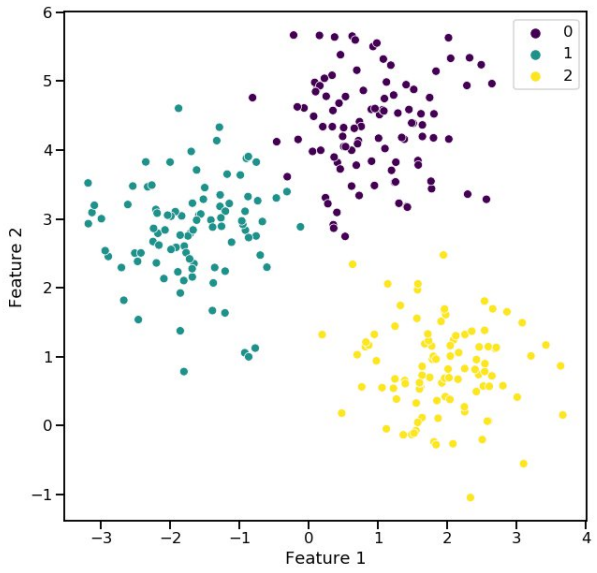
Approach: Find groups of objects sharing similarities, find hidden structures or associations, and remove noise from the data.

CLUSTERING

Input



Output



Find homogeneous subgroups among the observations or stated differently find subgroups of similar observations!

- How many subgroups?
- How to measure similarity?

Igual, L., & Seguí, S. (2024). Introduction to data science. In *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications* (pp. 1-4). Cham: Springer International Publishing.

Co-funded by the Erasmus+ programme of the European Union

Reinforcement Learning

Input: Environment which determines the state and the corresponding reward (e.g. win or lose) after an applied action.

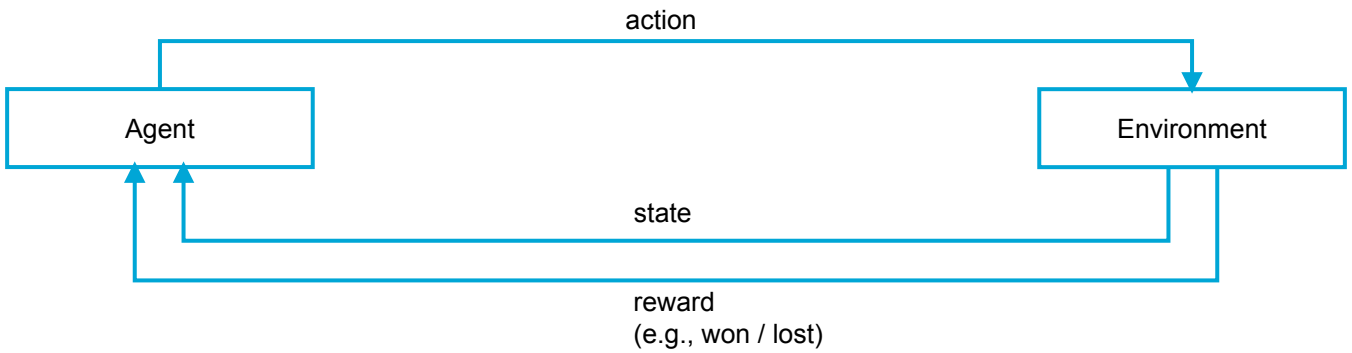
Goal: (Sequentially) select good actions given the current state of the environment.

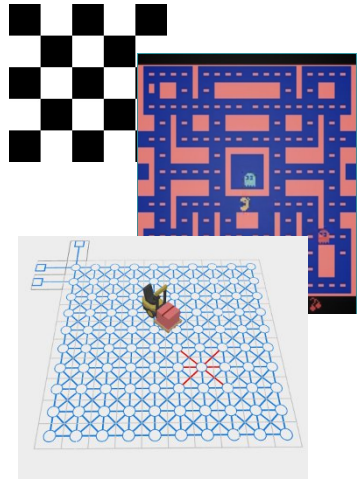
Approach: Agent learns a policy or a value function through interacting with the environment by “trying” actions (play); the policy or value functions maximize the expected rewards of actions depending on the current state.

EXAMPLE Q-LEARNING

Look-up table

State	Action	Q-value
(1,0,0)	A1	0.028
(1,0,0)	A2	0.320
(0,1,0)	A1	0.027
(0,1,0)	A2	0.034
…	…	…





- Agent takes decisions where to go next in the game or which task to schedule next to a fork lift (exploration vs intensification!)
- Environment returns the new state and the corresponding feedback
- Agent tries to learn a look-up table indicating the action with the highest expected reward for a given state

Igual, L., & Seguí, S. (2024). Introduction to data science. In *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications* (pp. 1-4). Cham: Springer International Publishing.

Co-funded by the Erasmus+ programme of the European Union

Overview

Machine Learning

This is not a complete overview of methods!

Supervised Learning

Input: Historical input data with a known result or output (label). Output (label) can have the form of a numerical value or category.
Goal: Predict the output of unseen or future data.
Approach: Learn a function (model) that describes the relationship (or mapping) between the historic input and the output variables and apply this function to the unseen or future data.

Classification

- Logistic regression
- Decision trees and random forests

Regression

- Linear regression
- Decision trees and random forests

Unsupervised Learning

Input: Historical input data without a known result.
Goal: Build meaningful subgroups (clusters), find associations or compress data (reduce dimensionality).
Approach: Find groups of objects sharing similarities, find hidden structures or associations, and remove noise from the data.

Clustering

- k-means
- Hierarchical clustering

Dimensionality reduction

- Principal component analysis (PCA)
- Nonlinear PCA

Association rule learning

- Aprior
- Eclat

Reinforcement Learning

Input: Environment which determines the state and the corresponding reward (e.g. win or lose) after an applied action.
Goal: (Sequentially) select good actions given the current state of the environment.
Approach: Agent learns a policy or a value function through interacting with the environment by "trying" actions (play); the policy or value functions maximize the expected rewards of actions depending on the current state.

Value based

- Deep Q Network

Policy based

- Policy Gradient

technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Recap this lecture

After successfully completing this micro-lecture, you are able to….

- Distinguish the terms artificial intelligence, machine learning and deep learning
- Explain the difference between supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning
- Identify the matching machine learning paradigm for a use case

The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

This material was created with the assistance of an AI language model.

technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon