Funded by
the European Union

ASSISTANT

Course on Artificial Intelligence

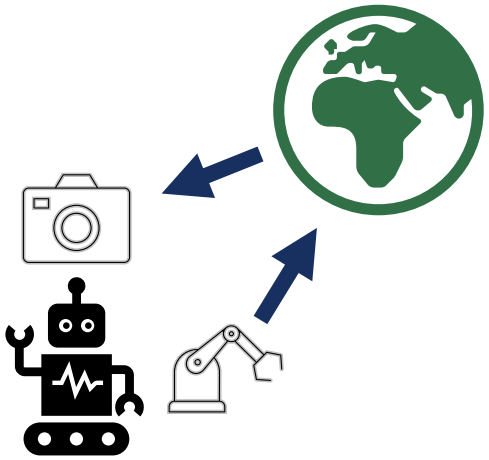
José Coelho, 2024

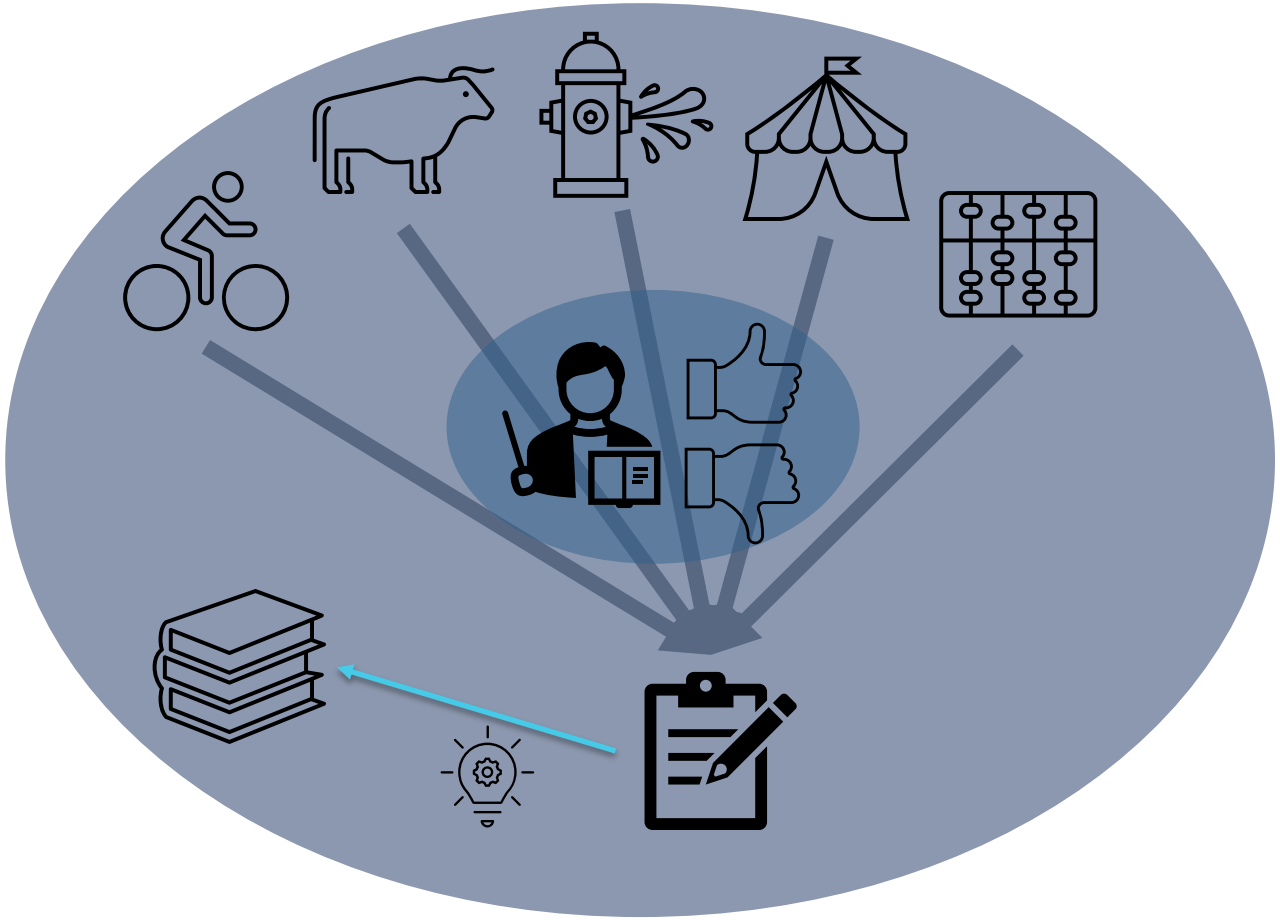
Lesson 3: Introduction to Machine Learning: Learning from Examples – Part A

Table of contents

Introduction 1/3

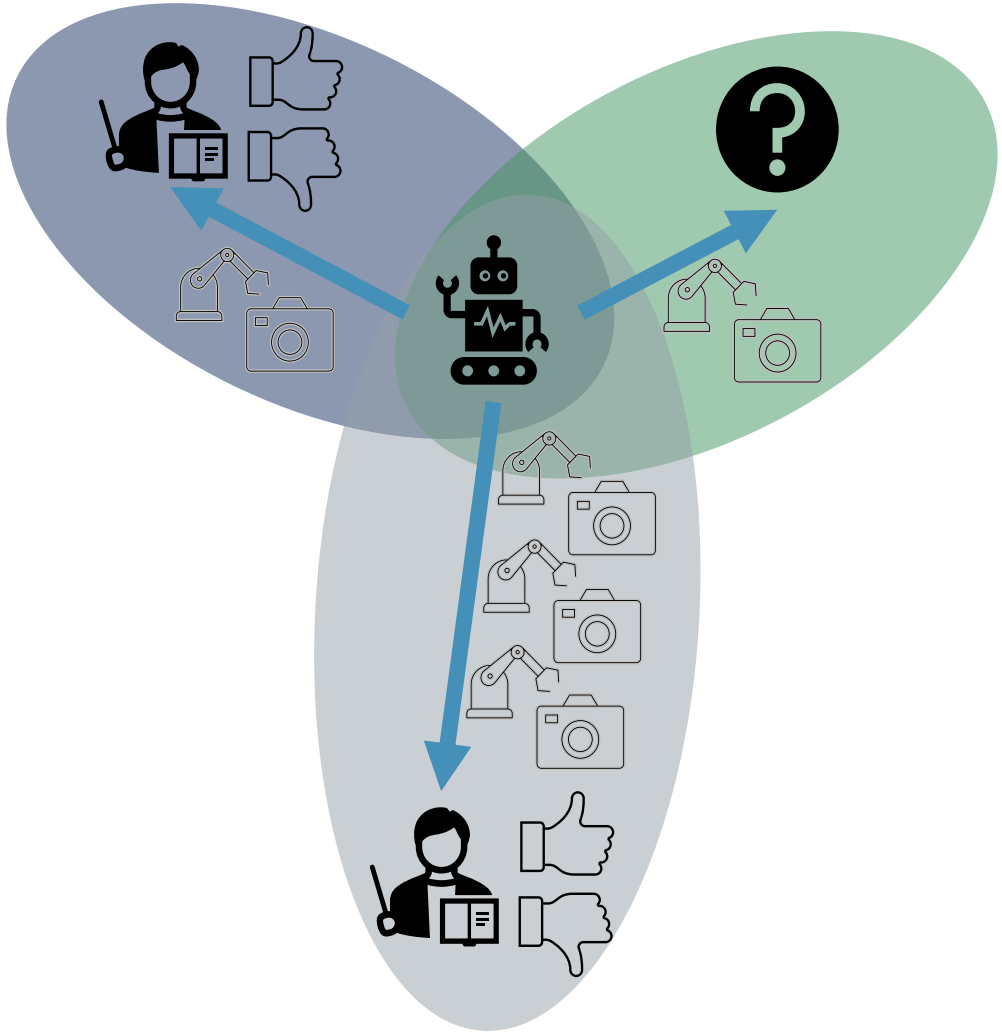
- Learning
- Induction/deduction
- Classification/Regression





Introduction 2/3

- Types of learning
  - Supervisioned
  - Unsupervised
  - Reinforced



Introduction 3/3

- Training set: $\{(x_1, y_1), \ldots, (x_n, y_n)\}$
- Real function: $y = f(x)$
- Hypothesis: $h(x) = \hat{y}, h \in \mathbb{H}$
- Example
  - Reality:
    - Students in a course
  - Observable variables:
    - Reading of materials;
    - Carrying out training activities;
    - Interventions in the forum;
    - Carrying out assessment activities.
  - Learn:
    - passing or failing the course

Case	Materials	Formative Activities	Interventions	Evaluations	Grade
1	1	2	2	1	1
2	3	2	1	3	2
3	2	2	1	2	1
4	2	2	1	3	2
5	2	2	2	1	1
6	2	3	1	2	2
7	1	2	3	2	1
8	2	2	3	2	2
9	2	1	1	2	1
10	3	2	3	3	2
11	1	2	2	1	1
12	3	1	2	3	2
13	2	2	1	1	1
14	2	2	2	3	2
15	1	1	3	1	1
16	3	3	3	2	2

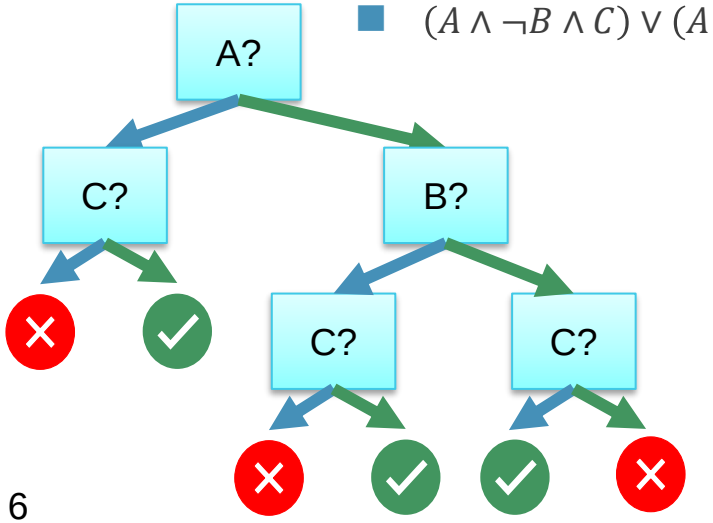
1	not done
2	partial done
3	complete

1	reproval
2	approval

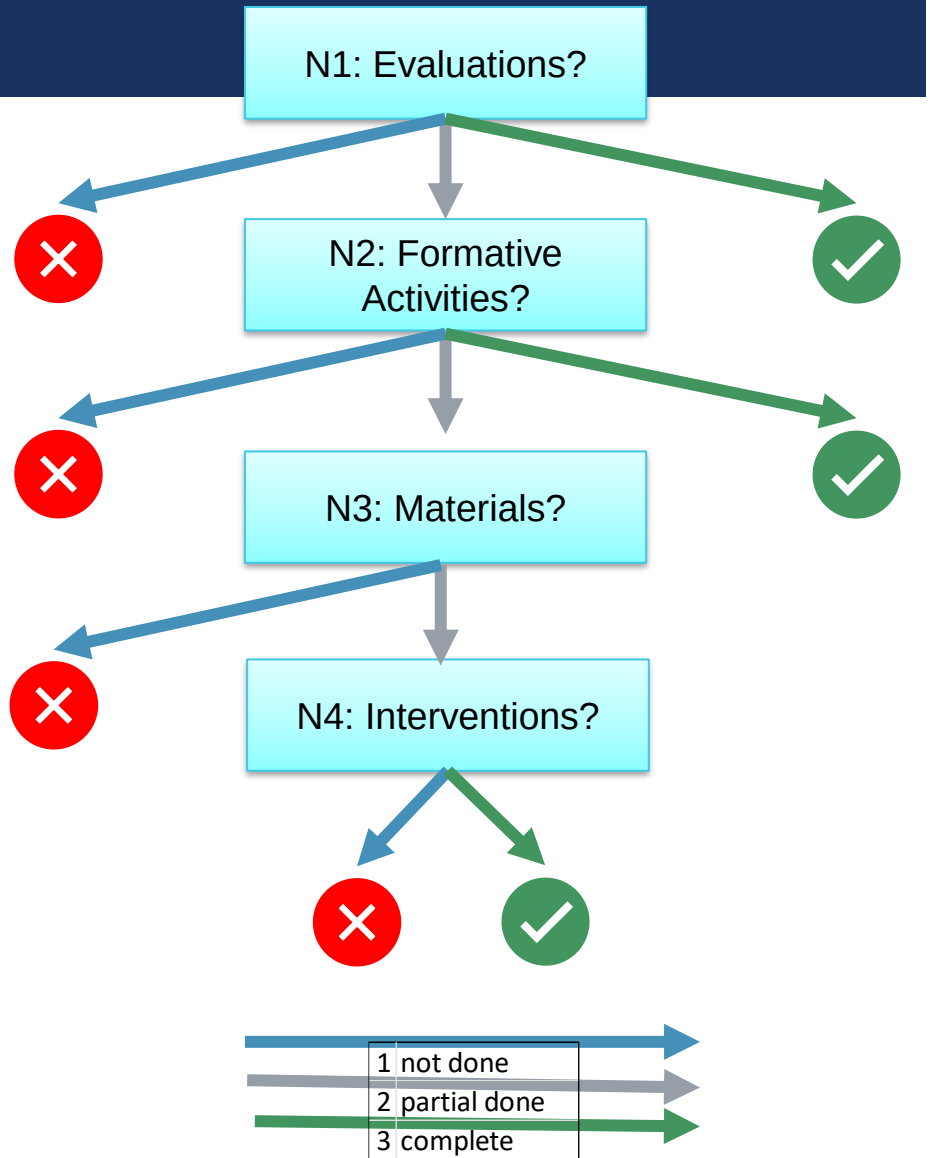
Decision Trees 1/3

- Definition
  - Example
- Expressiveness
  - propositional logic (DNF):
    - $(A \wedge \neg B \wedge C) \vee (A \wedge B \wedge \neg C) \vee (\neg A \wedge C)$

N1: Evaluations?					
1		2			3
Grade 1	N2: Formative Activities?				Grade 2
	1		2	3	
	Grade 1	N3: Materials?		Grade 2	
		1	2		
		Grade 1	N4: Interventions?		
			1	3	
			Grade 1	Grade 2	



A?			
0		1	
C?		B?	
0	1	0	1
0	1	C?	C?
		0 1	0 1
		0 1	1 0



Decision Trees 2/3

Case	Materials	Interventions	Grade
7	1	3	1
3	2	1	1
8	2	3	2

Case	Materials	Formative Activities	Interventions	Evaluations	Grade
1	1	2	2	1	1
5	2	2	2	1	1
11	1	2	2	1	1
13	2	2	1	1	1
15	1	1	3	1	1
3	2	2	1	2	1
7	1	2	3	2	1
9	2	1	1	2	1
6	2	3	1	2	2
8	2	2	3	2	2
16	3	3	3	2	2
2	3	2	1	3	2
4	2	2	1	3	2
10	3	2	3	3	2
12	3	1	2	3	2
14	2	2	2	3	2

- Build from cases:
  - Choose the most relevant attribute
  - Split cases by attribute value

Case	Materials	Formative Activities	Interventions	Grade
9	2	1	1	1
3	2	2	1	1
7	1	2	3	1
8	2	2	3	2
6	2	3	1	2
16	3	3	3	2

Case	Interventions	Grade
3	1	1
8	3	2

N1: Evaluations?					
1			2		3
Grade 1	N2: Formative Activities?				Grade 2
	1		2	3	
	Grade 1	N3: Materials?		Grade 2	
		1	2		
		Grade 1	N4: Interventions?		
			1	3	
			Grade 1	Grade 2	

Decision Trees 3/3

- Choosing the most relevant attribute
  - More solved cases
  - Entropy $H(V) = -\sum_k P(v_k) \log_2 P(v_k)$
    - Boolean variable, probability q
      - $B(q) = -(qlog_2 q + (1-q)\log_2(1-q))$
    - Information with p positive cases, n negative:
      - $B\left(\frac{p}{p+n}\right)$
  - Information gain (A attribute with d values)
    - $Gain(A) = B\left(\frac{p}{p+n}\right) - \sum_{k=1}^{d} \frac{p_k+n_k}{p+n} B\left(\frac{p_k}{p_k+n_k}\right)$
- Prune the tree
  - Remove the nodes without statistic relevance
- Continuous variables
  - Example: study hours
  - Cutoff
    - Choose the point with the highest gain

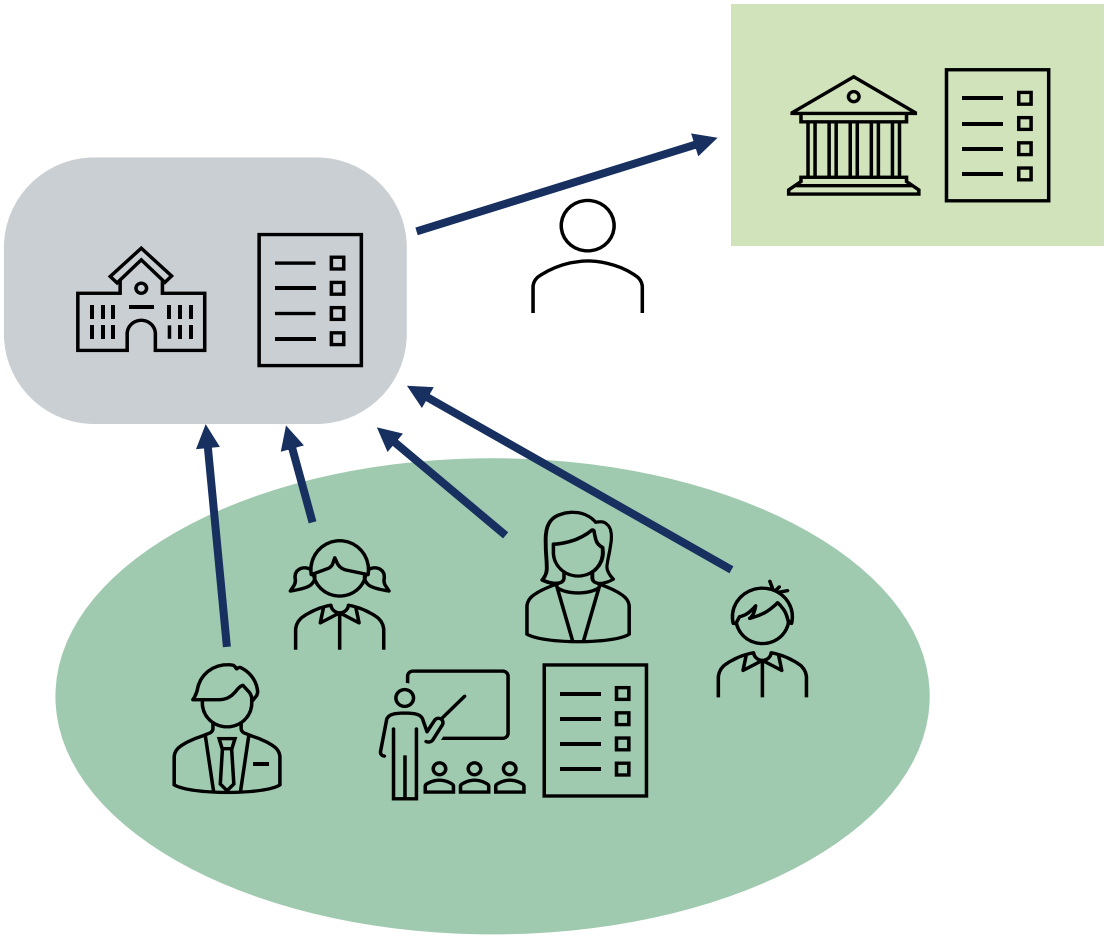
Values	Grade	Materials	Formative Activities	Interventions	Evaluations
1	1	4	2	3	5
1	2	0	1	3	0
2	1	4	6	3	3
2	2	4	5	2	3
3	1	0	0	2	0
3	2	4	2	3	5
	p+n	16	16	16	16
	p	8	8	8	8
	Solve	8	2	0	10
	Entropy	1,0000	1,0000	1,0000	1,0000
	Entropy 1	0,0000	0,9183	1,0000	0,0000
		0,2500	0,1875	0,3750	0,3125
	Entropy 2	1,0000	0,9940	0,9710	1,0000
		0,5000	0,6875	0,3125	0,3750
	Entropy 3	0,0000	0,0000	0,9710	0,0000
		0,2500	0,1250	0,3125	0,3125
	Gain	0,5000	0,1444	0,0182	0,6250

Values	Grade	Materials	Formative Activities	Interventions	Evaluations
1	1	1	1	2	0
1	2	0	0	1	0
2	1	2	2	0	3
2	2	2	1	0	3
3	1	0	0	1	0
3	2	1	2	2	0
	p+n	6	6	6	6
	p	3	3	3	3
	Solve	2	3	0	0
	Entropy	1,0000	1,0000	1,0000	1,0000
	Entropy 1	0,0000	0,0000	0,9183	0,0000
		0,1667	0,1667	0,5000	0,0000
	Entropy 2	1,0000	0,9183	0,0000	1,0000
		0,6667	0,5000	0,0000	1,0000
	Entropy 3	0,0000	0,0000	0,9183	0,0000
		0,1667	0,3333	0,5000	0,0000
	Gain	0,3333	0,5409	0,0817	0,0000

Values	Grade	Materials	Formative Activities	Interventions	Evaluations
1	1	1	0	1	0
1	2	0	0	0	0
2	1	1	2	0	2
2	2	1	1	0	1
3	1	0	0	1	0
3	2	0	0	1	0
	p+n	3	3	3	3
	p	1	1	1	1
	Solve	1	0	1	0
	Entropy	0,9183	0,9183	0,9183	0,9183
	Entropy 1	0,0000	0,0000	0,0000	0,0000
		0,3333	0,0000	0,3333	0,0000
	Entropy 2	1,0000	0,9183	0,0000	0,9183
		0,6667	1,0000	0,0000	1,0000
	Entropy 3	0,0000	0,0000	1,0000	0,0000
		0,0000	0,0000	0,6667	0,0000
	Gain	0,2516	0,0000	0,2516	0,0000

Assessment

- We consider that cases:
  - They have equal probability
  - They are independent
- Error: $\mathrm{P}(h(x) \neq y)$
- Sets:
  - Train
  - Validation
  - Test
- Model parameters
- Cross-validation



Thank you for attention

Resources:

- Microsoft Power Point / Clipchamp / DeepL

- Russell, S. J. & Norvig, P. (2010). Artificial intelligence: A modern approach (3rd ed). Prentice Hall.

Funded by the European Union

ASSISTANT - CHALLENGE BASED LEARNING IN AI ENHANCED DIGITAL TRANSFORMATION CURRICULAR NO. 2022-1-LT01-KA220-HED-000086555