Machine Learning – Decision Trees

1. Consider the following data relative to the classification of the quality of apartments, using 14 examples. The target-concept is "Acceptable", boolean, "Yes" or "No".

The three attributes are in the table.

Apartment	Furniture	Nº of rooms	New kitchen	Acceptable
1	No	1	Yes	No
2	Yes	1	No	No
3	Yes	1	Yes	Yes
4	No	2	Yes	Yes
5	Yes	2	No	No
6	Yes	2	Yes	Yes
7	No	2	No	No
8	Yes	3	No	No
9	No	4	Yes	No
10	Yes	3	Yes	Yes
11	Yes	4	No	No
12	No	3	Yes	No
13	No	4	No	No
14	Yes	4	Yes	Yes

i) Calculate the entropy of the target concept.

ii) Built the decision tree of the problem that you will train with the ID3 algorithm.

Show the information gain values for each candidate of each attribute in each step of ID3.

2. A company wants to launch a new gadget and for that it made market research, questioning several people about their decision relative to buying the gadget. The following results have been obtained.

	Age	Education	Income	Marital status	Will buy?
1	36-55	Master	high	single	Yes
2	18-35	Secondary	low	single	No
3	36-55	Master	low	single	No
4	18-35	Bachelor	high	single	No
5	< 18	Secondary	low	single	Yes
6	18-35	Bachelor	high	married	Yes
7	36-55	Bachelor	low	married	No
8	> 55	Bachelor	high	single	Yes
9	36-55	Master	low	married	No
10	> 55	Master	low	married	Yes
11	18-35	Master	high	married	Yes
12	> 55	Master	high	single	Yes
13	< 18	Secondary	high	single	No
14	> 55	Master	high	married	Yes
15	36-55	Secondary	low	single	Yes
16	< 18	Secondary	low	married	Yes
17	18-35	Master	low	married	No
18	> 55	Secondary	high	married	Yes
19	> 55	Bachelor	low	single	Yes
20	36-55	Secondary	high	married	No

a) Construct the decision tree for the problem using the ID3 trained with the first 14 examples.
b) Test the obtained tree with the last 6 examples, critically analyzing the results.

Note that this form of definition of numerical intervals allows to extend ID3 to the case of numerical attributes. Instead, verbal attributes could be used: young, young adult, adult, aged.

Coimbra, September 12th, 2023.