Artificial Intelligence 2 – SS 2020 Assignment 10: Decision Trees – Given June 19., Due June 28. –

Hint: Exercises need to be handed in via StudOn at 23:59 on the day they are due or earlier. Please use only the exercise group of your tutor to hand in your work.

If any concepts here seem unfamiliar to you or you have no idea how to proceed, consult the lecture materials, ask a fellow student, your tutor, or on the Forum.

If a problem asks for code, comment it or make it otherwise self-explanatory. You do not need to write a lot, but it should be enough to convince your tutor that you understand what the code does. We may deduct up to 30% for uncommented and unclear code, but would prefer not to.

Problems with no points (0pt) will not be graded, but might appear on the exam in a similar form. For these, we will provide a reference solution after the submission deadline. If you find the reference solution unclear, ask about it on the forum or in in a tutorial.

Problem 10.1 (Decision Trees - Errors)

1. Explain why it is possible (even common) that the learning curve (definition 24.4.6., slide 832) never gets to 100% correctness. Give an example of a hypothesis space and function being approximated where this could happen.

0pt

10 pt

40 pt

2. What are the sources of training set errors and validation set errors (for example, in the model selection algorithm on slide 841)?

Problem 10.2 (Home Decisions)

Consider the following decisions on finding a new home. The target is "Acceptable". 60pt

	Attributes			
House Nr.	Furniture	Nr. of rooms	New kitchen	Acceptable
1	No	3	Yes	Yes
2	Yes	3	No	No
3	No	4	No	Yes
4	No	3	No	No
5	Yes	4	No	Yes

- 1. Compute the information entropy for the priors in the table.
- Apply the decision tree learning algorithm on slide 826 on this table. Show the value of the information gain for each candidate attribute at each step in the construction of the tree. Do not use the house number as an attribute.
- 3. Draw the resulting decision tree.