

Exercises and Labs 2 for Lecture “Authentication „ (M.Sc.)

Exercise 2.1 Suppose that the probability of five events are $P(1) = 0.5$, and $P(2) = P(3) = P(4) = P(5) = 0.125$. Calculate the entropy. Explain in words what this means.

Exercise 2.2 Three binary nodes, N_1 , N_2 , and N_3 , split examples into (0,6), (1,5), and (3,3), respectively. For each node, calculate its entropy, Gini impurity, and classification error.

Exercise 2.3 Build a decision tree that computes the logical AND function.

Exercise 2.4 Imagine that there are four things that you like to do in the evening: going to a pub, watching TV, going to a party, or studying (!). The choice is sometimes made for you — if you have an assignment due the next day, you need to study, if you’re feeling lazy then the pub isn’t for you, and if there isn’t a party then you can’t go to it. You are looking for a decision tree which will help you decide what to do each evening. Here is a list of everything you’ve done in the past 10 days.

Deadline?	Is there a party?	Lazy?	Activity
Urgent	Yes	Yes	Party
Urgent	No	Yes	Study
Near	Yes	Yes	Party
None	Yes	No	Party
None	No	Yes	Pub
None	Yes	No	Party
Near	No	No	Study
Near	No	Yes	TV
Near	Yes	Yes	Party
Urgent	No	No	Study

Remark: The first thing to do is to work out which feature to use as the starting (root) node. For this you need to compute the entropy, and then find out which feature has the maximal information gain.

Exercise 2.5 Write out the steps in the ID3 algorithm in pseudo code.

Exercise 2.6 Consider the training data from the alphabet $A = \{a, b, c\}$:

$\omega 1$	$\omega 2$	$\omega 3$
aabbc	Bccba	caaaa
ababcc	Bbbca	cbcaab
babbcc	cbbaaaa	baaca

Use the edit distance to classify each of the following strings [if there are ambiguities in the classification, state which two (or all three) categories are candidates]: “abacc,” “abca,” “ccbba,” “bbaaac”.