

COMP5318/COMP4318 Machine Learning and Data Mining

Week 5 Tutorial exercises Decision Trees

Exercise 1. Decision trees and information gain (parts a) and b) – done in class; the rest in your own time)

Consider the following set of training examples:

shape	color	class
circle	blue	+
circle	blue	+
square	blue	-
triangle	blue	-
square	red	+
square	blue	-
square	red	+
circle	red	+

Adapted from M. Kubat, Introduction to Machine Learning, Springer, 2021

- What is the entropy of this collection of training examples with respect to the class?
- What is the information gain of the attribute *shape*?
- Which attribute will be selected as root of the tree based on information gain?
- Build the whole decision tree. Draw the tree after each selected attribute.

You may use this table to calculate information gain:

x	y	$-(x/y) * \log_2(x/y)$	x	y	$-(x/y) * \log_2(x/y)$	x	y	$-(x/y) * \log_2(x/y)$	x	y	$-(x/y) * \log_2(x/y)$
1	2	0.50	4	5	0.26	6	7	0.19	5	9	0.47
1	3	0.53	1	6	0.43	1	8	0.38	7	9	0.28
2	3	0.39	5	6	0.22	3	8	0.53	8	9	0.15
1	4	0.50	1	7	0.40	5	8	0.42	1	10	0.33
3	4	0.31	2	7	0.52	7	8	0.17	3	10	0.52
1	5	0.46	3	7	0.52	1	9	0.35	7	10	0.36
2	5	0.53	4	7	0.46	2	9	0.48	9	10	0.14
3	5	0.44	5	7	0.35	4	9	0.52			