

Prac Sheet - Decision Trees

Last modified on Wednesday, 24 April 2019 by f.maire@qut.edu.au

Decision Trees (DTs) are a supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from a dataset.

The following exercises are typical exam questions.

Exercise 1

- Consider the data table D below where C is the class label column (one example per row).

D =

X_1	X_2	X_3	X_4	C
F	F	F	F	P
F	F	T	T	P
F	T	F	T	P
T	T	T	F	P
T	F	F	F	N
T	T	T	T	N
T	T	T	F	N

$$\mathbf{X} = \{X_1, X_2, X_3, X_4\}$$

- What is the entropy of D?
- What is the information gain of X_1 ?
- What is the information gain of X_2 ?
- Build a DT to a depth of 3.

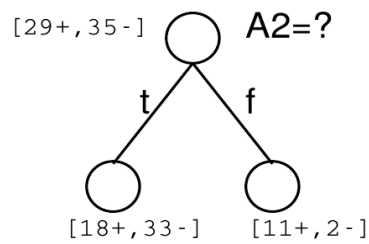
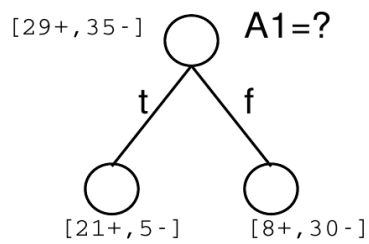
Exercise 2

- Recall that

$Gain(S, A) =$ expected reduction in entropy due to sorting on A

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

- Compute the information gain of the attribute A1 and A2



Exercise 3

- Which attribute should be used for a decision stump, *Humidity* or *Wind*?

