Lab

The following exercises must be submitted: 18.11.2021 , 05:00PM

Write a program in python for the scenario in the follwing scenarios

Upload Link:

https://forms.office.com/Pages/ResponsePage.aspx?id=p1CEoozbM0yRKkQzEbIt26j5uR98u7pHk4YAJdVRMgxUM1JFWlhJRUFESlBOSzVSRVFLSkhQRUwxNy4u

For those questions where justification is asked , write in hand answers for that

Exercise 1 : Decision Trees

Construct a decision tree corresponding to each of the following Boolean functions. The examples $\mathbf{x} \in D$ have one attribute for each Boolean variable A, B, . . . in the formula; the target concept $c(\mathbf{x})$ is the truth value of the formula itself. Assume the set D contains examples with all possible combinations of attribute values.

Hint: It may be helpful to write out the set D, i.e., the truth table for the variables and formula, first.

(a) $A \wedge \neg B$

(b) $A \vee (B \wedge C)$

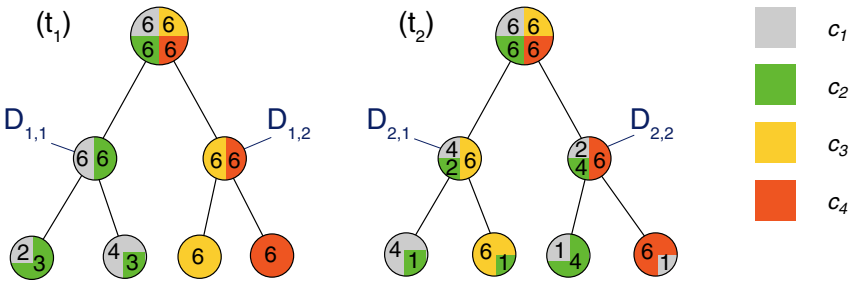
(c) A *XOR* B

(d) $(A \wedge B) \vee (C \wedge D)$

(e) $(A \vee B) \wedge (C$ *XOR* $D)$

Exercise 2 : Impurity Functions

Let D be a set of examples over a feature space X and a set of classes $C = \{c_1, c_2, c_3, c_4\}$, with $|D| = 24$. Consider the following illustration of two possible decision trees, t_1 and t_2 – the colors represent the classes present in each document set associated with the nodes of the trees; the numbers denote how many examples of each class are present.



(a) First, consider only the first split that each of the two trees makes: compute $\Delta\iota(\{D_{1,1}, D_{1,2}\})$ and $\Delta\iota(\{D_{2,1}, D_{2,2}\})$ with the misclassification rate $\iota_{misclass}$ and the entropy criterion $\iota_{entropy}$ as splitting criterion. Interpret the results: which of $\{D_{1,1}, D_{1,2}\}$ or $\{D_{2,1}, D_{2,2}\}$ is the better first split?

(b) Which of t_1, t_2 is the better decision tree, and why?

(c) Assuming the splits shown are the only possibilities, which of t_1 or t_2 would the ID3 algorithm construct, and why?