Decision Trees

Exercise 1 : Decision Trees

Construct by hand decision trees corresponding to each of the following Boolean formulas. The examples $(\mathbf{x}, c) \in D$ consist of a feature vector $\mathbf{x}$ where each component corresponds to one of the Boolean variables $(A, B, \dots)$ used in the formula, and each example corresponds to one interpretation (i.e. assignment of 0/1 to the Boolean variables). The target concept c is the truth value of the formula given that interpretation. Assume the set D contains examples with all possible combinations of attribute values.

Hint: It may be helpful to write out the set D for each formula as a truth table.

(a) $A \wedge \neg B$

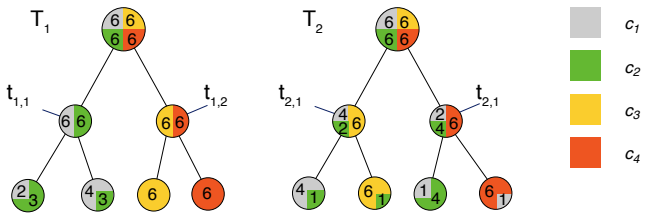
(b) A *XOR* B

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

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

Exercise 2 : Impurity Functions

Let D be a set of examples over a feature space $\mathbf{X}$ 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 $D(t)$ associated with a node t of a tree; 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 $\Delta\iota(D, \{D(t_{1,1}), D(t_{1,2})\})$ and $\Delta\iota(D, \{D(t_{2,1}), D(t_{2,2})\})$ with (1) the misclassification rate $\iota_{misclass}$ and (2) the entropy criterion $\iota_{entropy}$ as splitting criterion.

Interpret the results: which of $\{D(t_{1,1}), D(t_{1,2})\}$ or $\{D(t_{2,1}), D(t_{2,2})\}$ is the better first split?

(b) If we compare T_1 and T_2 in terms of their misclassification rate on D, which one is the better decision tree?

(c) Assuming the splits shown are the only possibilities, which of T_1 or T_2 would the ID3 algorithm construct, and why?

Exercise 3 : Decision Trees

Given is the following dataset to classifiy whether a dog is dangerous or well-behaved in character:

Color	Fur	Size	Character (C)
brown	ragged	small	well-behaved
black	ragged	big	dangerous
black	smooth	big	dangerous
black	curly	small	well-behaved
white	curly	small	well-behaved
white	smooth	small	dangerous
red	ragged	big	well-behaved

(a) Use the ID3 algorithm to determine a decision tree, where the attributes are to be chosen with the maximum average information gain *iGain*:

$$\textit{iGain}(D, A) \equiv H(D) \; - \sum_{a \in A} \frac{|D_a|}{|D|} \cdot H(D_a) \quad \text{with} \quad H(D) = -p_{\oplus} \log_2(p_{\oplus}) - p_{\ominus} \log_2(p_{\ominus})$$

(b) Classify the new example (Color=black, Fur=ragged, Size=small) using your decision tree.