BAUHAUS UNIVERSITÄT WEIMAR

INTRODUCTION TO MACHINE LEARNING

ASSIGNMENT 5

SUBMITTED BY

Anjana MuraleedharanNair(125512)
Isabel Maria Binu (125514)
Vishal Sanjay Shivam (125353)
Sharat Anand (125404)

Exercise 1 : Decision Trees

(a).

(a1) $x|_A$ - Projection Operator
(a2) T - Decision Tree
(a3) t - A node in the decision tree
(a4) X(t) - Subset of features X that is represented by node t in the decision tree
(a5) D(t) - Subset of the example set D that is represented by node t in the decision tree
(a6) $\Delta\iota$ - Impurity Reduction

(b). $X = \{x \in X : x|_A \in B\} \cup \{x \in X : x|_A \notin B\}$
Binary splitting induced by a (nominal) feature A

(c). Non-Negativity
Zero Impurity for Pure Nodes
Monotonicity.

(d). The hypothesis space of decision trees refers to the set of all possible decision trees that can be constructed over a given set of features and classes.

(e).The search space of the ID3 algorithm consists of all permutations of the features in the feature set. if the number of features (= dimensionality of a feature vector x) is p, then the search space contains p! Elements.

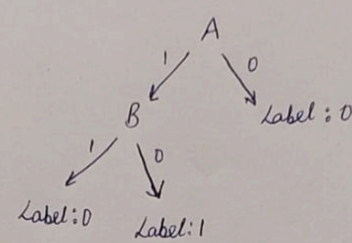
(f). The candidate elimination algorithm has an incomplete hypothesis space, containing only conjunctions of feature-value pairs as hypotheses.This restricted hypothesis space is searched completely by the candidate elimination algorithm, thus exhibiting a restriction bias.

The ID3 algorithm has a complete hypothesis space, since it contains all decision trees that can be constructed over D. But it searches this space incompletely, following a preference bias or search bias.

Exercise 2 : Decision Trees

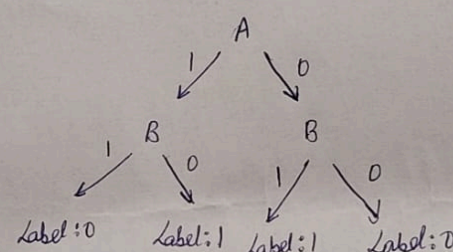
a) $A \wedge \neg B$

A	B	$C(x)$
0	0	0
0	1	0
1	0	1
1	1	0



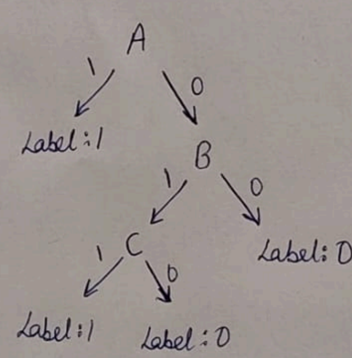
b) A XOR B

A	B	$C(x)$
0	0	0
0	1	1
1	0	1
1	1	0



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

A	B	C	$B \wedge C$	$C(x)$
0	0	0	0	0
0	0	1	0	0
0	1	0	0	0
0	1	1	1	1
1	0	0	0	1
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1



Exercise 3 : Impurity Functions

(a)

3(a). Using the misclassification rate

$$T_1 : \Delta_{misclass}(D, \{D(t_1,1), D(t_1,2)\}) = \text{Impurity}(D) - [P_1 * \text{impurity}(D(t_1,1)) + P_2 * \text{impurity}(D(t_1,2))]$$

$$T_2 : \Delta_{misclass}(D, \{D(t_2,1), D(t_2,2)\}) = \text{Impurity}(D) - [P_1 * \text{impurity}(D(t_2,1)) + P_2 * \text{impurity}(D(t_2,2))]$$

For $T_1 \rightarrow$

$$\Delta_{misclass} = i(1/4, 1/4, 1/4, 1/4) - [1/2\, i(1/2, 1/2, 0, 0) + 1/2\, i(0, 0, 1/2, 1/2)]$$

$$= 3/4 - (1/2 \times 1/2 + 1/2 \times 1/2)$$

$$= \underline{\underline{0.25}}$$

For $T_2 \rightarrow$

$$\Delta_{misclass} = i(1/4, 1/4, 1/4, 1/4) - [1/2\, i(1/3, 1/6, 1/2, 0) + 1/2\, i(1/6, 1/3, 0, 1/2)]$$

$$= 3/4 - (1/2 \times 1/2 + 1/2 \times 1/2)$$

$$= \underline{\underline{0.25}}$$

Using Entropy criteria

For $T_1 \rightarrow$

$$\Delta_{entropy}(D, \{D(t_1,1), D(t_1,2)\}) = i_{entropy}(1/4, 1/4, 1/4, 1/4) - 1/2 \times i_{entropy}(1/2, 1/2, 0, 0) - 1/2 \times i_{entropy}(0, 0, 1/2, 1/2)$$

$$= -(1/4 \log_2(1/4) + 1/4 \log_2(1/4) + 1/4 \log_2(1/4) + 1/4 \log_2(1/4)) + 1/2\,(1/2 \log_2(1/2) + 1/2 \log_2(1/2)) + 1/2\,(1/2 \log_2(1/2) + 1/2 \log_2(1/2))$$

$$= \underline{\underline{1}}$$

For $T_2 \rightarrow$

$$\Delta_{entropy}(D, \{D(t_2,1), D(t_2,2)\}) = i_{entropy}(1/4, 1/4, 1/4, 1/4) - 1/2\, i_{entropy}(1/3, 1/6, 1/2, 0) - 1/2\, i_{entropy}(1/6, 1/3, 0, 1/2)$$

$$= -4(1/4 \log_2(1/4)) + 1/2\,(1/3 \log_2(1/3) + 1/6 \log_2(1/6) + 1/2 \log_2(1/2)) + 1/2\,(1/6 \log_2(1/6) + 1/3 \log_2(1/3) + 1/2 \log_2(1/2))$$

$$= \underline{\underline{0.54}}$$

The first split, $\{D(t_1,1), D(t_1,2)\}$ is considered better because of its higher impurity reduction based on entropy criteria.

Exercise 4 : Decision Trees
(a)

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

Attribute : color

values $\rightarrow$ Black, White, Red, Brown

$$H(A|B_i) = -\sum_{i=1}^{k} P(A_i|B_i)\cdot\log_2(P(A_i|B_i))$$

$H:[4+,3-] = -4/7\log_2(4/7) - 3/7\log_2(3/7) = \cancel{0.985}\ 0.985$

$H_{Black} = [1+,2-] \rightarrow$ Entropy $(H_{Black}) = -1/3\log_2(1/3) - 2/3\log_2(2/3) = 0.9182$

$H_{White} = [1+,1-] \rightarrow$ Entropy $(H_{White}) = 1$

$H_{Red} = [1+,0] \rightarrow$ Entropy $(H_{Red}) = 0$

$H_{Brown} = [1+,0] \rightarrow$ Entropy $(H_{Brown}) = 0$

$\Delta\text{entropy} = \text{entropy}(D) - \sum_{i=1}^{m}\frac{|D_i|}{|D|}\,\text{entropy}(D_i) = 0.985 - (3/7\times 0.9182 + 2/7\times 1 + 0)$

$= \cancel{0.305} = \underline{\underline{0.305}}$

Attribute : Fur

values $\rightarrow$ Ragged, Smooth, Curly.

$H = [4+,3-] = \cancel{0.985}.\ 0.985$

$H_{Ragged} = [2+,1-] \rightarrow$ Entropy $(H_{ragged}) = -2/3\log_2(2/3) - 1/3\log_2(1/3) = 0.9182$

$H_{smooth} = [0+,2-] \rightarrow$ Entropy $(H_{smooth}) = 0$

$H_{curly} = [2+,0-] \rightarrow$ Entropy $(H_{curly}) = 0$

$\Delta\text{entropy} = 0.985 - (3/7\times 0.9182) = \cancel{0.837}\ \underline{\underline{0.591}}$

Attribute : Size

values $\rightarrow$ Small, Big.

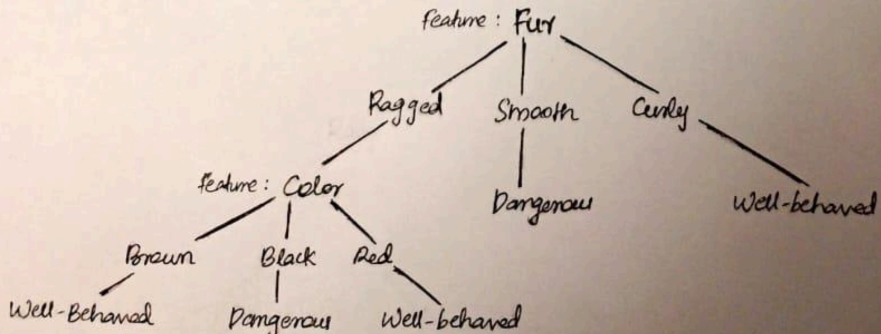
$H = [4+,3-] = 1.231.$

$H_{small} = [3+,1-] \rightarrow$ Entropy $(H_{small}) = -3/4\log_2(3/4) - 1/4\log_2(1/4) = 0.8112$

$H_{big} = [1+,2-] \rightarrow$ Entropy $(H_{big}) = 0.9182$

$\Delta\text{entropy} = 0.985 - (4/7\times 0.8112 + 3/7\times 0.9182) = \cancel{0.3739}\ \underline{0.128}$

Maximum information gain (Δentropy) $\cancel{0.591}$ 0.591 $\rightarrow$ Fur (select as root node). = 0.591



Color	Size	Character
Brown	Small	Well-behaved
Black	Big	dangerous
Red	Big	Well-behaved.

→ Attribute : color

values → Brown, Black, Red.

$H_{furi} = [2+,1-] = -2/3 \log_2(2/3) - 1/3 \log_2(1/3) = 0.9182$

$H_{brown} = [1+,0] \rightarrow$ Entropy $(H_{brown}) = 0$

$H_{black} = [0,1-] \rightarrow$ Entropy $(H_{black}) = 0$

$H_{red} = [1+,0] \rightarrow$ Entropy $(H_{red}) = 0$.

Δ entropy $= \underline{\underline{0.9182}}$

→ Attribute : size

values → small, big.

$H_{furi} = [2+,1-] = 0.9182$

$H_{small} = [1+,0-] = 0$

$H_{big} = [1+,1-] = 1$.

Δ entropy $= 0.9182 - (2/3 \times 1) \approx \underline{0.251}$

Information gain is maximum for the attribute color. So it is selected as the next node.

Exercise 5 : Cost functions

(a) For node ‘small’: toxic has 2 misclassifications edible has 1 misclassification
For the node ‘large’: toxic has 2 misclassifications edible has 0 misclassification Based on minimum cost, the labels for the node small is edible and large is edible.

(b) Cost(c’,c) = { **0** ***if c ' = edible and c = poisonous***
1 ***otherwise***