Techniques of Artificial Intelligence

Exercises – Decision Trees

Andrei Covaci & Arthur Moraux
andrei.covaci@vub.be

May 2, 2022

Representation and Interpretation of Boolean Functions

Symbol	**Name**
0	FALSE
1	TRUE
A	A
B	B
$!A/\neg A$	NOT A
$!B/\neg B$	NOT B
$A \wedge B$	A AND B
$A \vee B$	A OR B
$\neg A \wedge B$	NOT A AND B
$A \wedge \neg B$	A AND NOT B
$\neg A \vee B$	NOT A OR B
$A \vee \neg B$	A OR NOT B
$A \oplus B$	A XOR B
$A \bar{\vee} B$	A NOR B
A XNOR B	A XNOR B
$A \bar{\wedge} B$	A NAND B

1. Give decision trees to represent the following Boolean functions:
   (a) $A \wedge \neg B$
   (b) $A \vee (B \wedge C)$
   (c) $A \oplus B$
   (d) $(A \wedge B) \vee (C \wedge D)$
   (e) A XNOR B (hint: XNOR $= (A \wedge B) \vee (\neg A \wedge \neg B)$)

2. **True or false:** Any Boolean function can be expressed as a decision tree. Give a proof.

3. **More general than** (Exercise 3.3 from Tom Mitchel's book)
   True or False: If a decision tree D2 is an elaboration of tree D1, then D1 is more general than D2. Assume D1 and D2 are decision trees representing arbitrary Boolean functions, and that D2 is an elaboration of D1 if ID3 could extend D1 into D2. If true, give a proof; if false, give a counterexample.

   More-general-than is defined as: Let h_j and h_k be Boolean-valued functions defined over X. Then h_j is more general than or equal to h_k if and only if $(\forall x \in X)[(h_k(x) = 1 \rightarrow (h_j(x) = 1)]$

4. **ID3**
   In order to evaluate the quality of the tree which is grown by ID3, one could compare its performance to a baseline performance. The baseline performance is often the performance of a very simple machine learning algorithm. Consider the following approach:

   You have a dataset consisting of 25 examples of two classes. You plan to use leave-one-out cross validation. As a baseline, you use a simple majority classifier (a majority classifier is given a set of training data and then always outputs the class that is in the majority in the training set, regardless the input). Such a majority classifier is expected to score about 50%, but with this example of leave-one-out cross-validation, it does not. What will be its performance and why?

5. **Decision Trees with ID3**
   Consider the following data on the hair color, body weight, body height and the usage of lotion of eight different people. The table shows whether the people got sunburned after an afternoon in the sun.

Hair	**Height**	**Weight**	**Lotion**	**Sunburned?**
Blonde	Average	Light	No	Yes
Blonde	Tall	Average	Yes	No
Brown	Short	Average	Yes	No
Blonde	Short	Average	No	Yes
Red	Average	Heavy	No	Yes
Brown	Tall	Heavy	No	No
Brown	Average	Heavy	No	No
Blonde	Short	Light	Yes	No

   (a) Perform average entropy calculations on the following complete dataset for each of the four attributes. Select the attribute which minimizes the entropy; draw the first level of the decision tree.

   (b) Grow the tree until you reach the proper identification of all the samples.

   (c) Establish the rules from the tree found in the previous question.

Hair	Height	Weight	Lotion	Sunburned?
Red	Tall	Average	Yes	No

(d) The factual value of the training instance, in the dataset is: **Sunburned? – Yes**. Consider the following additional instance in **your** dataset:

Will the inclusion of this new instance in your training dataset impact the structure of the decision tree? What are such instances considered to be? Does their presence or absence in the training dataset impact the outcome of the built model?

(e) Consider the herewith provided alternative dataset:

Hair	Height	Weight	Lotion	Sunburned?
?	Average	Light	No	Yes
Blonde	Tall	Average	Yes	No
Brown	Short	Average	Yes	No
Blonde	Short	Average	No	Yes
Red	Average	Heavy	No	Yes
Brown	Tall	Heavy	No	No
Brown	Average	?	No	No
Blonde	Short	Light	?	No

Is this dataset suitable for training your model and will you be able to build an optimal decision tree? Explain briefly how you can deal with this situation.