**INTRODUCTION MACHINE LEARNING**

**EXERCISE 5**

A red sign with white text

Description automatically generated

|  |
| --- |
| **TEACHER:** |
| Johannes Kiesel |

|  |
| --- |
| **GROUP:** |
| Group 16 |
|  |
| **SUBMITTED BY:** |
| Aaron Perez Herrera |
| Cesar Fernando Gamba Tiusaba |
| Chun Ting Lin |
| Olubunmi Emmanuel Ogunleye |

Exercise 1: Decision Trees (1+1+1+1+1+1=6 Points)

1. Name these concepts:

(a1) xlA: This represents a partition of the feature space where condition A is met.

(a2) T: This symbol denotes a decision tree.

(a3) t: This symbol represents a node within the decision tree.

(a4) X(t): This refers to the feature vector associated with a specific node t in the

decision tree.

(a5) D(t): This symbolizes the subset of the example set D that is represented by the node t.

(a6) ∆I: This signifies the reduction in impurity resulting from a split.

1. Name this expression: X = {x E X : x| E B} U {x E X : xl B}

This represents a partitioning of the dataset X into two subsets based on whether the attribute value x∣A​ belongs to the subset B or not.

1. What are the three requirements of an impurity function?
2. Non-negativity: The impurity function must be non-negative (ι(D)≥0\iota(D) \geq 0ι(D)≥0 for any dataset DDD).
3. Maximum at uniform distribution: The impurity function reaches its maximum when all classes in DDD are uniformly distributed.
4. Minimum at pure distribution: The impurity function is zero when all examples in DDD belong to the same class.
5. What is the hypothesis space of decision trees?

The set of possible decision trees over D forms the hypothesis space H.

1. What is the search space of the ID3 algorithm?

The search space of the ID3 algorithm consists of all possible trees that can be constructed by recursively partitioning the dataset using attributes based on information gain.

1. What is the difference between the inductive bias of the candidate elimination algorithm and that of the ID3 algorithm? Hint: search bias and restriction bias.

* Candidate elimination algorithm: Combines search bias (consistent hypotheses with training data) and restriction bias (focus on version space).
* ID3 algorithm: Uses search bias to prefer smaller trees with higher information gain, but its restriction bias assumes all target concepts can be represented as decision trees.

Exercise 2: Decision Trees (1+1+1+0=3 Points)

Construct by hand decision trees corresponding to each of the following Boolean formulas. The examples (x, c) E D consist of a feature vector x where each component corresponds to one of the Boolean variables (A, B, ...) 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.

1. A ^ (NOT B)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | FALSE  1  0  TRUE  FALSE  1  0   |  | | --- | |  | |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| A | B | A and NOT(B) |
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

1. A XOR B

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1  0  FALSE  TRUE  TRUE  FALSE  1  0  0  1   |  | | --- | |  | |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| A | B | A XOR B |
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

1. A V (B ^ C)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1  TRUE  0  1  FALSE  0  1  TRUE  FALSE  0   |  | | --- | |  | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| A | B | C | A or (B and C) |
| 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 1 | 0 | 0 | 1 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 1 |

1. (A ^ B) V (C ^ D)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1  1  1  TRUE  FALSE  0  FALSE  0  TRUE  1  0  0  0  1  FALSE  TRUE  FALSE  1  0   |  | | --- | |  | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A | B | C | D | (A and B) or ( C and D) |
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 1 | 1 |
| 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 | 0 |
| 0 | 1 | 1 | 1 | 1 |
| 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 |
| 1 | 0 | 1 | 0 | 0 |
| 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 | 1 |
| 1 | 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 1 | 1 |

Exercise 5:

(a) Determine the labels of all nodes using the given cost function.

The cost function is defined as:

A black text on a white background

Description automatically generated

1. Training Data:

• Edibility: “toxic” or “edible.”

• Nodes split by “Size” into small and large.

2. Determine labels:

• For each leaf node, assign the most common label from the corresponding subset. If there’s a tie, any label can be assigned.

3. Subsets:

• Size = small: Instances 1, 2, 4 → Toxic: 1, Edible: 2. Majority: Edible.

• Size = large: Instances 3, 5 → Toxic: 1, Edible: 1. Tie, assign any (e.g., Edible).

(b) Devise a new cost function

The new cost function is designed to heavily penalize the classification of toxic mushrooms as edible. It can be expressed as:

A math equation with black text

Description automatically generated with medium confidence

The goal is to design a cost function that penalizes classifying a toxic mushroom as edible more heavily than other errors

This ensures that misclassifying toxic mushrooms has a much higher penalty than other types of errors.

(c) Compute the misclassification costs

For the given decision tree and training data:

Original Cost Function:

A black and white math equation

Description automatically generated with medium confidence

• Misclassification costs:

• Small subset: 1 toxic misclassified → Cost = 1

• Large subset: 1 toxic misclassified → Cost = 1

• Total Cost: 1 + 1 = 2

New Cost Function:

A math equations on a white background

Description automatically generated

• Misclassification costs:

• Small subset: 1 toxic misclassified → Cost = 10

• Large subset: 1 toxic misclassified → Cost = 10

• Total Cost: 10 + 10 =20