CSE352 AI

HOMEWORK 2  - 10pts

Homework 2 has 3 Problems in 2 PARTS and an extra credit problem.

SOLVE  ONLY TWO problems of your choice

I will post Solutions so you could compare your solutions with mine. PROBLEMS similar to Hmks will appear on your TEST

You must TYPE the statement of each problem you are solving; otherwise problem will not be considered for correction. You can DRAW your TREES by hand.

Submit HMK as a TEAM by e-mail to TA

You must include TEAM Number and Names and IDs of all team members on a FRONT page.
Here are some **DEFINITIONS** from the **Lecture Notes** that **YOU NEED** for your Homework

**Definition 1**

Given a classification dataset DB with a set $A = \{a_1, a_2, ..., a_n\}$ of attributes and a **class attribute** $C$ with values $\{c_1, c_2, ..., c_k\}$ ($k$ classes), any expression $a_i = v_i \wedge \ldots \wedge a_k = v_k$, where $a_i \in A$ and $v_i$ are values of attributes, is called a **DESCRIPTION**.

In particular, $C = c_k$ is called a **CLASS DESCRIPTION**.

**Definition 2**

A **CHARACTERISTIC FORMULA** is any expression $C = c_k \Rightarrow a_i = v_i \wedge \ldots \wedge a_k = v_k$.

We write it shortly as

**CLASS ⇒ DESCRIPTION**

**Definition 3**

A **DISCRIMINANT formula** is any expression $a_i = v_i \wedge \ldots \wedge a_k = v_k \Rightarrow C = c_k$

written shortly as

**DESCRIPTION ⇒ CLASS**
**Definition 4**

A characteristic formula $\text{CLASS} \Rightarrow \text{DESCRIPTION}$ is called a **CHARACTERISTIC RULE** of the classification dataset $\text{DB}$ iff it is **TRUE** in $\text{DB}$, i.e. when the following holds

$$\{o : \text{DESCRIPTION}\} \cap \{o : \text{CLASS}\} \neq \emptyset$$

where $\{o : \text{DESCRIPTION}\}$ is the set of all records of $\text{DB}$ corresponding to the description $\text{DESCRIPTION}$ and $\{o : \text{CLASS}\}$ is the set of all records of $\text{DB}$ corresponding to the description $\text{CLASS}$

**Definition 5**

A discriminant formula $\text{DESCRIPTION} \Rightarrow \text{CLASS}$ is called a **DISCRIMINANT RULE** of $\text{DB}$ iff it is **TRUE in DB**, i.e. the following two conditions hold

1. $\{o : \text{DESCRIPTION}\} \neq \emptyset$

2. $\{o : \text{DESCRIPTION}\} \subseteq \{o : \text{CLASS}\}$
PART ONE:
Classification: Characteristic and Discriminant Rules

Given a dataset:

<table>
<thead>
<tr>
<th>Record</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$o_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$o_2$</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>$o_3$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$o_4$</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$o_5$</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

C – class attribute

Problem 1

1. Find sets \{o : DESCRIPTION\} for the following descriptions

Follow the Example below when writing your solutions

Example: for description $a_1 = 2 \land a_2 = 1$ you have evaluate the set:
\{o : a_1 = 2 \land a_2 = 1\} = \{ o_2, o_5 \}

1) $a_3 = 1 \land a_4 = 0$

2) $a_2 = 0 \land a_3 = 2$

3) $C=1$

4) $C=0$
2. For the following formulas use proper definitions stated above to determine, it means use proper definitions to prove whether they are or they are not DISCRIMINANT / CHARACTERISTIC RULES of our dataset.

Example:
\[ a_1 = 2 \Lambda a_2 = 1 \Rightarrow C = 1 \]
is a DISCRIMANT Formula that is NOT DISCRIMINANT RULE because
\[ \{o : a_1 = 2 \Lambda a_2 = 1 \} = \{ o_2, o_5 \}, \quad \{o : C=1\} = \{ o_1, o_5 \} \]
and \[ \{ o_2, o_5 \} \text{ is NOT a subset of } \{ o_1, o_5 \} \]

5) \[ a_1 = 1 \Lambda a_2 = 1 \Rightarrow C = 1 \]
6) \[ C = 1 \Rightarrow a_1 = 0 \Lambda a_2 = 1 \Lambda a_3 = 1 \]
7) \[ C = 2 \Rightarrow a_1 = 1 \]
8) \[ C = 0 \Rightarrow a_1 = 1 \Lambda a_4 = 0 \]
9) \[ a_1 = 2 \Lambda a_2 = 1 \Lambda a_3 = 1 \Rightarrow C = 0 \]
10) \[ a_1 = 0 \Lambda a_3 = 2 \Rightarrow C = 1 \]
PART TWO:

Decision Tree Learning 1

Here is the TRAINING DATA SET FOR THE HOMEWORK:
Class Attribute: Buys Computer

<table>
<thead>
<tr>
<th>Age</th>
<th>Income</th>
<th>Student</th>
<th>Credit Rating</th>
<th>Buys Computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;=30</td>
<td>high</td>
<td>No</td>
<td>Fair</td>
<td>No</td>
</tr>
<tr>
<td>&lt;=30</td>
<td>high</td>
<td>No</td>
<td>Excellent</td>
<td>No</td>
</tr>
<tr>
<td>31…40</td>
<td>high</td>
<td>No</td>
<td>Fair</td>
<td>Yes</td>
</tr>
<tr>
<td>&gt;40</td>
<td>medium</td>
<td>No</td>
<td>Fair</td>
<td>Yes</td>
</tr>
<tr>
<td>&gt;40</td>
<td>Low</td>
<td>Yes</td>
<td>Fair</td>
<td>Yes</td>
</tr>
<tr>
<td>&gt;40</td>
<td>low</td>
<td>Yes</td>
<td>Excellent</td>
<td>No</td>
</tr>
<tr>
<td>31…40</td>
<td>low</td>
<td>Yes</td>
<td>Excellent</td>
<td>Yes</td>
</tr>
<tr>
<td>&lt;=30</td>
<td>medium</td>
<td>No</td>
<td>Fair</td>
<td>No</td>
</tr>
<tr>
<td>&lt;=30</td>
<td>low</td>
<td>Yes</td>
<td>Fair</td>
<td>Yes</td>
</tr>
<tr>
<td>&gt;40</td>
<td>medium</td>
<td>Yes</td>
<td>Fair</td>
<td>Yes</td>
</tr>
<tr>
<td>&lt;=30</td>
<td>medium</td>
<td>Yes</td>
<td>Excellent</td>
<td>Yes</td>
</tr>
<tr>
<td>31…40</td>
<td>medium</td>
<td>No</td>
<td>Excellent</td>
<td>Yes</td>
</tr>
<tr>
<td>31…40</td>
<td>high</td>
<td>Yes</td>
<td>Fair</td>
<td>Yes</td>
</tr>
<tr>
<td>&gt;40</td>
<td>medium</td>
<td>No</td>
<td>Excellent</td>
<td>No</td>
</tr>
</tbody>
</table>

When building your DECISION TREES follow Examples in the Lectures; you must include ALL steps of their constructions not only the final results.
Problem 2

Use the Training Data to create two decision trees described as follows:

Tree 1:

1. Build the decision tree using **SIMPLE heuristic** as defined in **Lecture 8**

   Use **CREDIT RATING** as the root attribute, and nodes attributes of your own choice;
   YOU MUST use at least 3 attributes as nodes.

2. Write down all the rules determined by your tree in the description and in the predicate forms

3. EVALUATE predictive accuracy for the set of your rules with respect to the **TEST Dataset** below

Tree 2:

1. Use **Basic ID3 algorithm** from **Lecture 8** (without Information Gain).
   Use **INCOME** as root attribute, and nodes attributes of your choice;

2. Write down all the rules determined by your tree in the description and predicate forms

3. Evaluate correctness of your rules, i.e. the predictive accuracy with respect to the TRAINING data

4. Evaluate predictive accuracy for the set of your rules with respect to the **TEST Dataset** below.
   Must show work.
TEST DATA SET

<table>
<thead>
<tr>
<th>Obj</th>
<th>Age</th>
<th>Income</th>
<th>Student</th>
<th>Credit_Rating</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;=30</td>
<td>High</td>
<td>Yes</td>
<td>Fair</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>31…40</td>
<td>Low</td>
<td>No</td>
<td>Fair</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>31…40</td>
<td>High</td>
<td>Yes</td>
<td>Excellent</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>&gt;40</td>
<td>Low</td>
<td>Yes</td>
<td>Fair</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>&gt;40</td>
<td>Low</td>
<td>Yes</td>
<td>Excellent</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>&lt;=30</td>
<td>Low</td>
<td>No</td>
<td>Fair</td>
<td>No</td>
</tr>
</tbody>
</table>

Problem 3

Create test data set of at least 6 records for your sets of rules corresponding to Tree 1 or Tree 2 that guarantees 100% predictive accuracy. Prove that your example is correct.

Extra Credit  - 5pts

EVALUATE Information Gain for 2 attributes on one NODE of your choice of your tree Tree 1 or 2

You must show work, not a final number; in fact you can write proper formulas for its computation without evaluating (calculator) the numbers. I want to SEE if you understand the formulas