Cse352 AI
Homework 2 Solutions

PART ONE
Classification: Characteristic and Discriminant Rules

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 = $\{a_1, a_2, \ldots, a_n\}$ of attributes and a **class attribute C** with values $\{c_1, c_2, \ldots, c_k\}$ (k classes),
any expression;
$a_1 = v_1 \wedge \ldots \wedge a_k = v_k$, where a_i `in A`, c_k `in C 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
$C = c_k \Rightarrow a_1 = v_1 \wedge \ldots \wedge a_k = v_k$,
We write it shortly as

CLASS => DESCRIPTION

Definition 3

A **DISCRIMINANT formula** is any expression

$a_1 = v_1 \Lambda \ldots \Lambda a_k = v_k \Rightarrow C = c_k$

written shortly as

DESCIPTION => CLASS

Definition 4

A characteristic formula **CLASS => DESCRIPTION** is called a **CHARACTERISITIC RULE** of the classification dataset DB iff it is **TRUE** in DB, i.e. when the following holds

{o: DESCRIPTION} ∩ {o: CLASS} not= empty set

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

Definition 5

A discriminant formula DESCRIPTION => CLASS is called a **DISCRIMINANT RULE** of DB iff it is **TRUE in DB,** i.e. the following two conditions hold

1. **{o: DESCRIPTION} not= empty set**
2. **{o: DESCRIPTION} included in {o: CLASS}**

PROBLEM 1: Given a dataset:

Record	a_1	a_2	a_3	a_4	**C**
o_1	1	1	1	0	1
o_2	2	1	2	0	2
o_3	0	0	0	0	0
o_4	0	0	2	1	0
o_5	2	1	1	0	1

C – class attribute

Find {o :DESCRIPTION} for the following descriptions
Write solution in space provided

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

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

$a_3 = 1$

Record	a_1	a_2	a_4	C
o_1	1	1	0	1
o_5	2	1	0	1

$a_3 = 1 \Lambda a_4 = 0$

Record	a_1	a_2	C
o_1	1	1	1
o_5	2	1	1

$\{o : a_3 = 1 \Lambda a_4 = 0\} = \{ o_1 , o_5 \}$

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

$a_2 = 0$

Record	a_1	a_3	a_4	C
o_3	0	0	0	0
o_4	0	2	1	0

$a_2 = 1 \wedge a_3 = 2$

Record	a_1	a_4	C
o_4	0	1	0

$\{o : a_2 = 1 \wedge a_3 = 2\} = \{ o_4 \}$

3) c=1

C=1

Record	a_1	a_2	a_3	a_4
o_1	1	1	1	0
o_5	2	1	1	0

$\{o : C = 1 \} = \{ o_1, o_5 \}$

4) c=0

C=0

Record	a_1	a_3	a_3	a_4
o_3	0	0	0	0
o_4	0	2	2	1

$\{o : C = 0 \} = \{ o_3, o_4 \}$

2: Given a dataset:

Record	***a_1***	***a_2***	***a_3***	***a_4***	**C**

o_1	1	1	1	0	1
o_2	2	1	2	0	2
o_3	0	0	0	0	0
o_4	0	0	2	1	0
o_5	2	1	1	0	1

C – class attribute

For the following formulas use proper definitions stated above to determine (it means prove) whether they are or they are not **DISCRIMINANT / CHARACTERISTIC RULES** in our dataset

5) $a_1 = 1 \wedge a_2 = 1 \Rightarrow C = 1$

DESCIPTION ⇒ CLASS
DISCRIMINANT FORMULA

$a_1=1$

Record	a_2	a_3	a_4	C
o_1	1	1	0	1

$a_1 = 1 \wedge a_2 = 1$

Record	a_3	a_4	C
o_1	1	0	1

C=1

Record	a_1	a_2	a_3	a_4
o_1	1	1	1	0
o_5	2	1	1	0

$\{o : a_1 = 1 \wedge a_2 = 1\} = \{ o_1 \}$

$\{o: \text{DESCRIPTION}\}$ not= $\emptyset$

$\{o : C = 1\} = \{ o_1, o_5 \}$

$\{o : a_1 = 1 \wedge a_2 = 1\} = \{ o_1 \} \subseteq \{o : C = 1\} = \{ o_1, o_5 \}$

$\{o: \text{DESCRIPTION}\} \subseteq \{o: \text{CLASS}\}$

Rule in the dataset = true

6) $C = 1 \Rightarrow a_1 = 0 \wedge a_2 = 1 \wedge a_3 = 1$

CLASS ⇒ DESCRIPTION
CHARACTERISTIC FORMULA

C=1

Record	a_1	a_2	a_3	a_4
o_1	1	1	1	0
o_5	2	1	1	0

a_1=0

Record	a_2	a_3	a_4	C
o_3	0	0	0	0
o_4	0	2	1	0

$a_1 = 0 \wedge a_2 = 1$
None
$a_1 = 0 \wedge a_2 = 1 \wedge a_3 = 1$
None

$\{o : a_1 = 0 \wedge a_2 = 1 \wedge a_3 = 1\} = \emptyset$
$\{o : C = 1\} = \{ o_1, o_5 \}$
$\{o : a_1 = 0 \wedge a_2 = 1 \wedge a_3 = 1\} = \emptyset \cap \{o : C = 1\} = \{ o_1, o_5 \} = \emptyset$
$\{o: \text{DESCRIPTION}\} \cap \{o: \text{CLASS}\} = \emptyset$

Rule in the Dataset = false

7) $C = 2 \Rightarrow a_1 = 1$

CLASS ⇒ DESCRIPTION

CHARACTERISTIC FORMULA

C=2

Record	a_1	a_3	a_3	a_4
o_2	2	1	2	1

a_1=1

Record	a_2	a_3	a_4	C
o_1	1	1	0	1

$\{o : a_1 = 1\} = \{ o_1 \}$
$\{o : C = 2\} = \{ o_2 \}$
$\{o : a_1 = 1\} = \{ o_1 \} \cap \{o : C = 2\} = \{ o_2 \} = \emptyset$
$\{o\text{: DESCRIPTION}\} \cap \{o\text{: CLASS}\} = \emptyset$

Rule in the Dataset = false

8) $C = 0 \Rightarrow a_1 = 1 \wedge a_4 = 0$

CLASS ⇒ DESCRIPTION
CHARACTERISTIC FORMULA

C=0

Record	a_1	a_3	a_3	a_4
o_3	0	0	0	0
o_4	0	2	2	1

a_1=1

Record	a_2	a_3	a_4	C
o_1	1	1	0	1

$a_1 = 1 \wedge a_4 = 0$

Record	a_2	a_3	C
o_1	1	1	1

$\{o : C = 0\} = \{o_3, o_4\}$
$\{o : a_1 = 1 \wedge a_4 = 0\} = \{o_1\}$
$\{o : C = 0\} = \{o_3, o_4\} \cap \{o : a_1 = 1 \wedge a_4 = 0\} = \{o_1\} = \emptyset$
$\{o\text{: DESCRIPTION}\} \cap \{o\text{: CLASS}\} = \emptyset$

Rule in the Dataset = false

9) $a_1 = 2 \wedge a_2 = 1 \wedge a_3 = 1 \Rightarrow C = 0$

DESCIPTION $\Rightarrow$ CLASS
DISCRIMINANT FORMULA

a_1=2

Record	a_2	a_3	a_4	C
o_2	2	1	2	2
o_5	2	1	1	1

$a_1 = 2 \wedge a_2 = 1$
None

$a_1 = 2 \wedge a_2 = 1 \wedge a_3 = 1$
None

C=0

Record	a_1	a_3	a_3	a_4
o_3	0	0	0	0
o_4	0	2	2	1

$\{o : a_1 = 0 \wedge a_2 = 1 \wedge a_3 = 1\} = \emptyset$
$\{o\text{: DESCRIPTION}\} = \emptyset$

Rule of Dataset = false

10) $a_1 = 0 \ \Lambda \ a_3 = 2 \Rightarrow C = 1$

DESCIPTION ⇒ CLASS
DISCRIMINANT FORMULA

a_1=0

Record	a_2	a_3	a_4	C
o_3	0	0	0	0
o_4	0	2	1	0

a1 = 0 Λ a3 = 2

Record	a_2	a_4	C
o_3	0	0	0

C=1

Record	a_1	a_2	a_3	a_4
o_1	1	1	1	0
o_5	2	1	1	0

$\{o : a_1 = 0 \ \Lambda \ a_3 = 2\} = \{ o_3 \}$
{o: DESCRIPTION} not= Ø
$\{o : C = 1 \} = \{ o_1, o_5 \}$
$\{o : a_1 = 0 \ a_3 = 2\} = \{ \emptyset \} \subseteq \{o : C = 1\} = \{ o_1, o_5 \}$
{o: DESCRIPTION} not⊆ {o: CLASS}

Rule in the Dataset = false

PART TWO:
Decision Tree Learning 1

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

Age	Income	Student	Credit Rating	Buys Computer
<=30	High	No	Fair	No
<=30	High	No	Excellent	No
31…40	High	No	Fair	Yes
>40	medium	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	No
31…4	Low	Yes	Excellent	Yes

0				
<=30	mediu m	No	Fair	No
<=30	Low	Yes	Fair	Yes
>40	mediu m	Yes	Fair	Yes
<=30	mediu m	Yes	Excellent	Yes
31…4 0	mediu m	No	Excellent	Yes
31…4 0	High	Yes	Fair	Yes
>40	mediu m	No	Excellent	No

Problem 1

Use the Training Data to create **two decision trees**:

Tree 1 : Build the decision tree using **general majority voting heuristic**, defined as follows:
You CAN use MAJORITY Vote for the majority class at any table at any level of the tree – when you choose so.
Use CREDIT RATING as the **root attribute**, and nodes attributes of your own choice;
Write down all the rules determined by your tree in the **predicate form**

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

SHOW WORK.

Tree 2 Use **Basic ID3 algorithm**
Use INCOME as **root attribute**, and nodes attributes of your choice;

2. Write down all the rules determined by your tree in the **predicate form**
3. EVALUATE predictive accuracy for the set of your rules with respect to the TEST Dataset below
SHOW WORK.

TEST DATA SET

Obj	Age	Income	Student	Credit_Rating	Class
1	<=30	High	Yes	Fair	Yes
2	31…40	Low	No	Fair	Yes
3	31…40	High	Yes	Excellent	No
4	>40	Low	Yes	Fair	Yes
5	>40	Low	Yes	Excellent	No
6	<=30	Low	No	Fair	No

Problem 2

Create test data sets of at least 6 records for your sets of rules rules corresponding to **Tree 1** and **Tree 2** that **guarantees 100% predictive accuracy.**

PROVE that your Example is correct.

Extra Credit

EVALUATE **Information Gain** **for 2 attributes** on one NODE of your tree.

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.

SOLUTIONS
ATTENTION:
TREES PUBLISHED here MAY have MISTAKES!!
YOu have to find what and where they are is wrong
- I explained it in class many times
It is an exercise for proper application Of the **Decision Tree Algorithm TERMINATING Conditions.**

Also In the case of the **general voting heuristic**
Evebody can have different anwrrs of the one published here

Tree 1 and Rules from Tree 1
Use general voting heuristic, defined as follows: You CAN use MAJORITY Vote for the majority class at any table at any level of the tree – when you choose so

Training Data:

Age	Income	Student	Credit Rating	Buys Computer
<=30	High	No	Fair	No
<=30	High	No	Excellent	No
31…40	High	No	Fair	Yes
>40	Medium	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	No

31…40	Low	Yes	Excellent	Yes
<=30	Medium	No	Fair	No
<=30	Low	Yes	Fair	Yes
>40	Medium	Yes	Fair	Yes
<=30	Medium	Yes	Excellent	Yes
31…40	Medium	No	Excellent	Yes
31…40	High	Yes	Fair	Yes
>40	Medium	No	Excellent	No

We choose **Credit Rating** as the root attribute

Credit Rating

Fair

Age	Income	Student	Buys Computer
<=30	High	No	No
31…40	High	No	Yes
>40	Medium	No	Yes
>40	Low	Yes	Yes
<=30	Medium	No	No
<=30	Low	Yes	Yes
>40	Medium	Yes	Yes
31…40	High	Yes	Yes

General Majority Voting = Yes

Excellent

Age	Income	Student	Buys Computer
<=30	High	No	No
>40	Low	Yes	No
31…40	Low	Yes	Yes
<=30	Medium	Yes	Yes
31…40	Medium	No	Yes
>40	Medium	No	No

Student

No

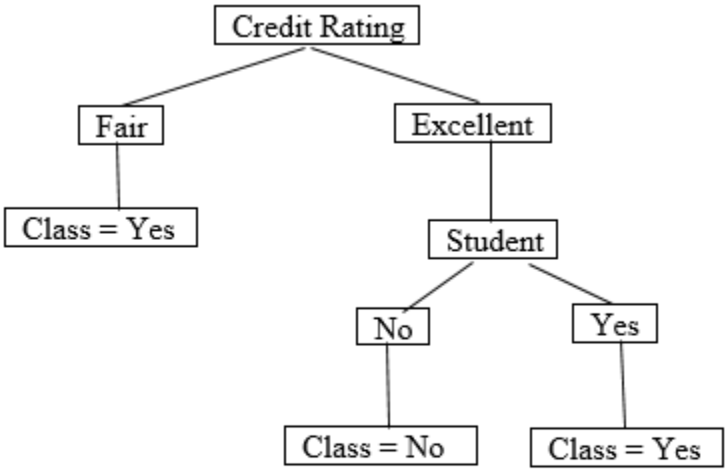
Age	Income	Buys Computer
<=30	High	No
31…40	Medium	Yes
>40	Medium	No

General Majority Voting = No

Yes

Age	Income	Buys Computer
>40	Low	No
31…40	Low	Yes
<=30	Medium	Yes

General Majority Voting = Yes

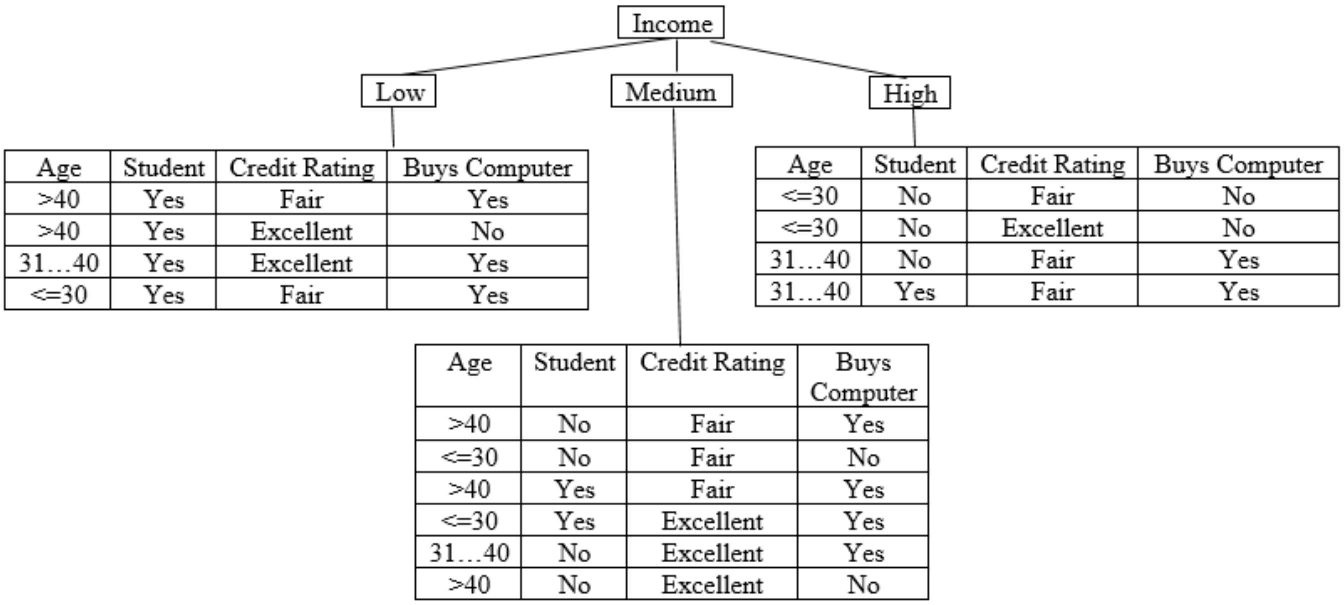


2. Rules

R1. Credit Rating(x, Fair) ⇒ buysComputer(x, Yes)
R2. Credit Rating(x, Excellent) Λ Student(x, No) ⇒ buysComputer(x, No)
R3. Credit Rating(x, Excellent) Λ Student(x, Yes) ⇒ buysComputer(x, Yes)

Tree 2 and Rules from Tree 2

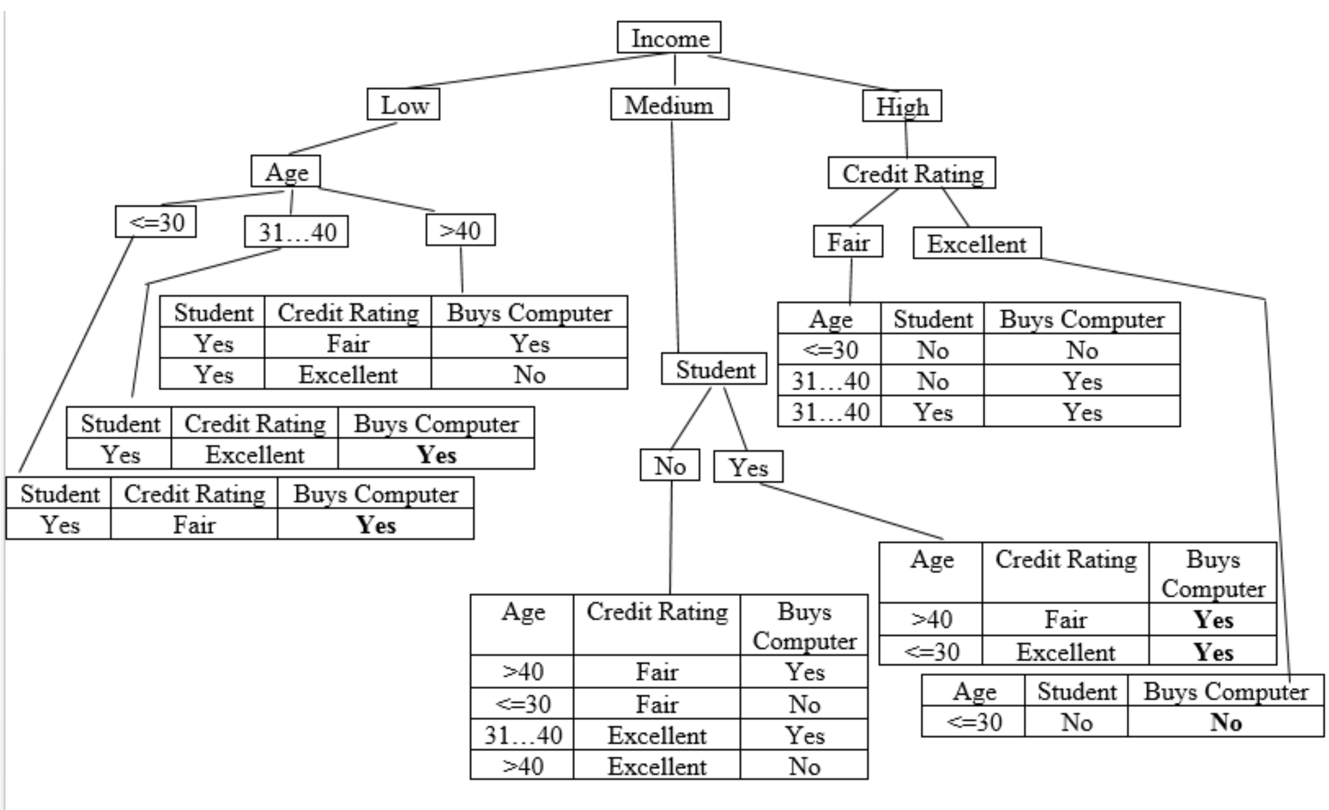
Age	Income	Student	Credit Rating	Buys Computer
<=30	High	No	Fair	No
<=30	High	No	Excellent	No
31…40	High	No	Fair	Yes
>40	Medium	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	No
31…40	Low	Yes	Excellent	Yes
<=30	Medium	No	Fair	No
<=30	Low	Yes	Fair	Yes
>40	Medium	Yes	Fair	Yes
<=30	Medium	Yes	Excellent	Yes
31…40	Medium	No	Excellent	Yes
31…40	High	Yes	Fair	Yes
>40	Medium	No	Excellent	No



Age	Student	Credit Rating	Buys Computer
>40	Yes	Fair	Yes
>40	Yes	Excellent	No
31…40	Yes	Excellent	Yes
<=30	Yes	Fair	Yes

Age	Student	Credit Rating	Buys Computer
>40	No	Fair	Yes
<=30	No	Fair	No
>40	Yes	Fair	Yes
<=30	Yes	Excellent	Yes
31…40	No	Excellent	Yes
>40	No	Excellent	No

Age	Student	Credit Rating	Buys Computer
<=30	No	Fair	No
<=30	No	Excellent	No
31…40	No	Fair	Yes
31…40	Yes	Fair	Yes



Student	Credit Rating	Buys Computer
Yes	Fair	**Yes**

Student	Credit Rating	Buys Computer
Yes	Excellent	**Yes**

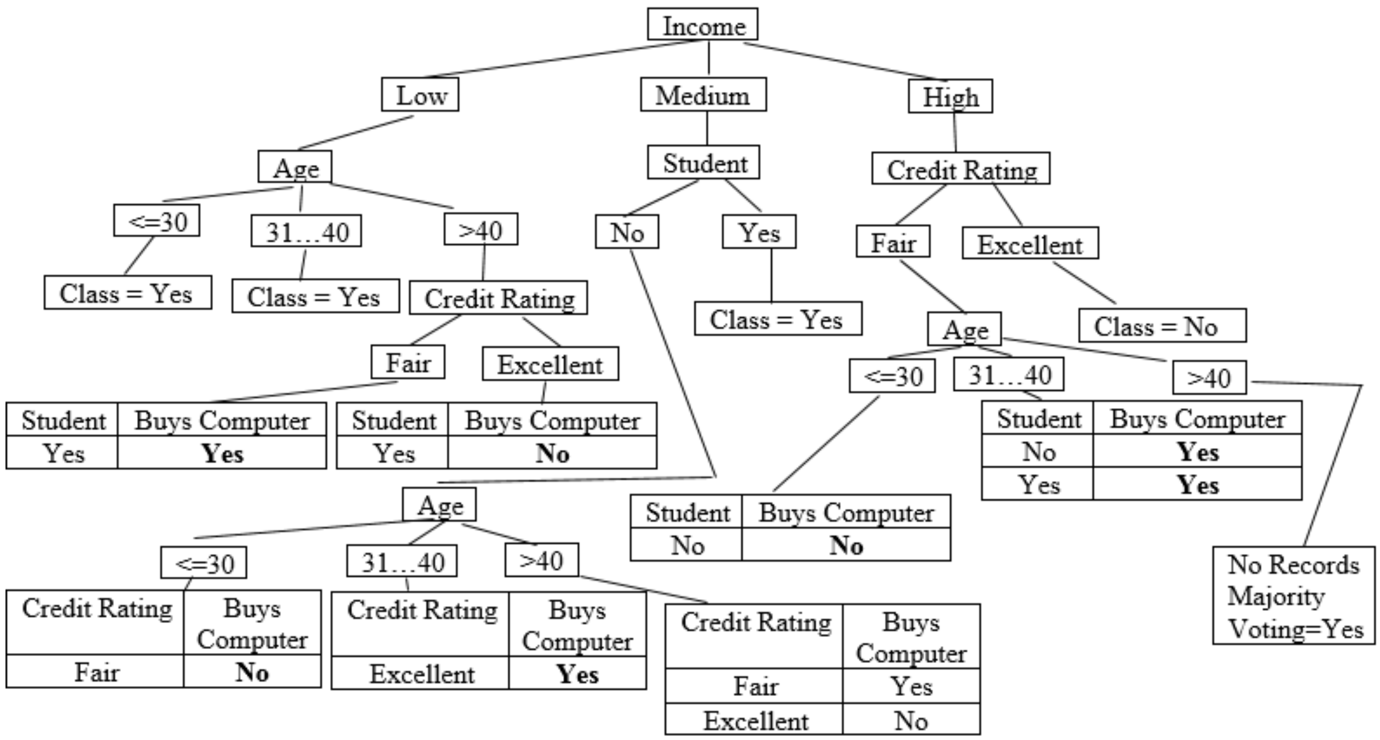
Student	Credit Rating	Buys Computer
Yes	Fair	Yes
Yes	Excellent	No

Age	Credit Rating	Buys Computer
>40	Fair	Yes
<=30	Fair	No
31…40	Excellent	Yes
>40	Excellent	No

Age	Credit Rating	Buys Computer
>40	Fair	**Yes**
<=30	Excellent	**Yes**

Age	Student	Buys Computer
<=30	No	No
31…40	No	Yes
31…40	Yes	Yes

Age	Student	Buys Computer
<=30	No	**No**



R1. Income(x, Low) Λ Age(x, <=30) ⇒ buysComputer(x, Yes)
R2. Income(x, Low) Λ Age(x, 31…40) ⇒ buysComputer(x, Yes)
R3. Income(x, Low) Λ Age(x, >40) Λ CreditRating(x, Fair) ⇒ buysComputer(x, Yes)
R4. Income(x, Low) Λ Age(x, >40) Λ CreditRating(x, Excellent) ⇒ buysComputer(x, No)

R5. Income(x, Medium) Λ Student(x, No) Λ Age(x, <=30) ⇒ buysComputer(x, No)
R6. Income(x, Medium) Λ Student(x, No) Λ Age(x, 31…40) ⇒ buysComputer(x, Yes)
R7. Income(x, Medium) Λ Student(x, No) Λ Age(x, >40) Λ ⇒ buysComputer(x, Yes)
R8. Income(x, Medium) Λ Student(x, No) Λ Age(x, >40) Λ CreditRating(x, Excellent) ⇒ buysComputer(x, No)
R9. Income(x, Medium) Λ Student(x, Yes) ⇒ buysComputer(x, Yes)
R10. Income(x, High) Λ CreditRating(x, Fair) Λ Age(x, <=30) ⇒ buysComputer(x, No)
R11. Income(x, High) Λ CreditRating(x, Fair) Λ Age(x, 31…40) ⇒ buysComputer(x, Yes)
R12. Income(x, High) Λ CreditRating(x, Fair) Λ Age(x, >40) ⇒ buysComputer(x, No)
R13. Income(x, High) Λ CreditRating(x, Excellent) ⇒ buysComputer(x, No)

Predictive Accuracy for Tree 1

Obj	Age	Income	Student	Credit_Rating	Class
1	<=30	High	Yes	Fair	Yes
2	31…40	Low	No	Fair	Yes
3	31…40	High	Yes	Excellent	No
4	>40	Low	Yes	Fair	Yes
5	>40	Low	Yes	Excellent	No
6	<=30	Low	No	Fair	No

Obj 1: Fits - must say which rule
Obj 2: Fits Fits - must say which rulke
Obj 3: Fails R3

Obj 4: Fits Fits - must say which rule
Obj 5: Fails R3
Obj 6: Fails R1

3 Passed/out of 6=50% Predictive Accuracy

Predictive Accuracy for Tree 2

Obj	Age	Income	Student	Credit_Rating	Class
1	<=30	High	Yes	Fair	Yes
2	31…40	Low	No	Fair	Yes
3	31…40	High	Yes	Excellent	No
4	>40	Low	Yes	Fair	Yes
5	>40	Low	Yes	Excellent	No
6	<=30	Low	No	Fair	No

Obj 1 Fails R10
Obj Fits - must say which rule
Obj 3 Fits - must say which rule
Obj 4 Fits - must say which rule
Obj 5 Fits - must say which rule
Obj 6 Fails R1

4 Passed/out of 6=66.7% Predictive Accuracy

Problem 2 Solution for Tree 1

Obj	Age	Income	Student	Credit_Rating	Class
1	<=30	Low	Yes	Fair	Yes

2	31…40	Low	Yes	Fair	Yes
3	31…40	High	Yes	Excellent	Yes
4	31…40	Medium	No	Fair	Yes
5	>40	Medium	No	Excellent	No
6	<=30	Low	No	Fair	Yes

Obj 1 Fits - must say which rule
Obj 2 Fits I - must say which rule
Obj 3 Fits] - must say which rule
Obj 4 Fits - must say which rule
Obj 5 Fits - must say which rule
Obj 6 Fits - must say which rule

6 Passed/out of 6=100% Predictive Accuracy

Problem 2 Solution for Tree 2

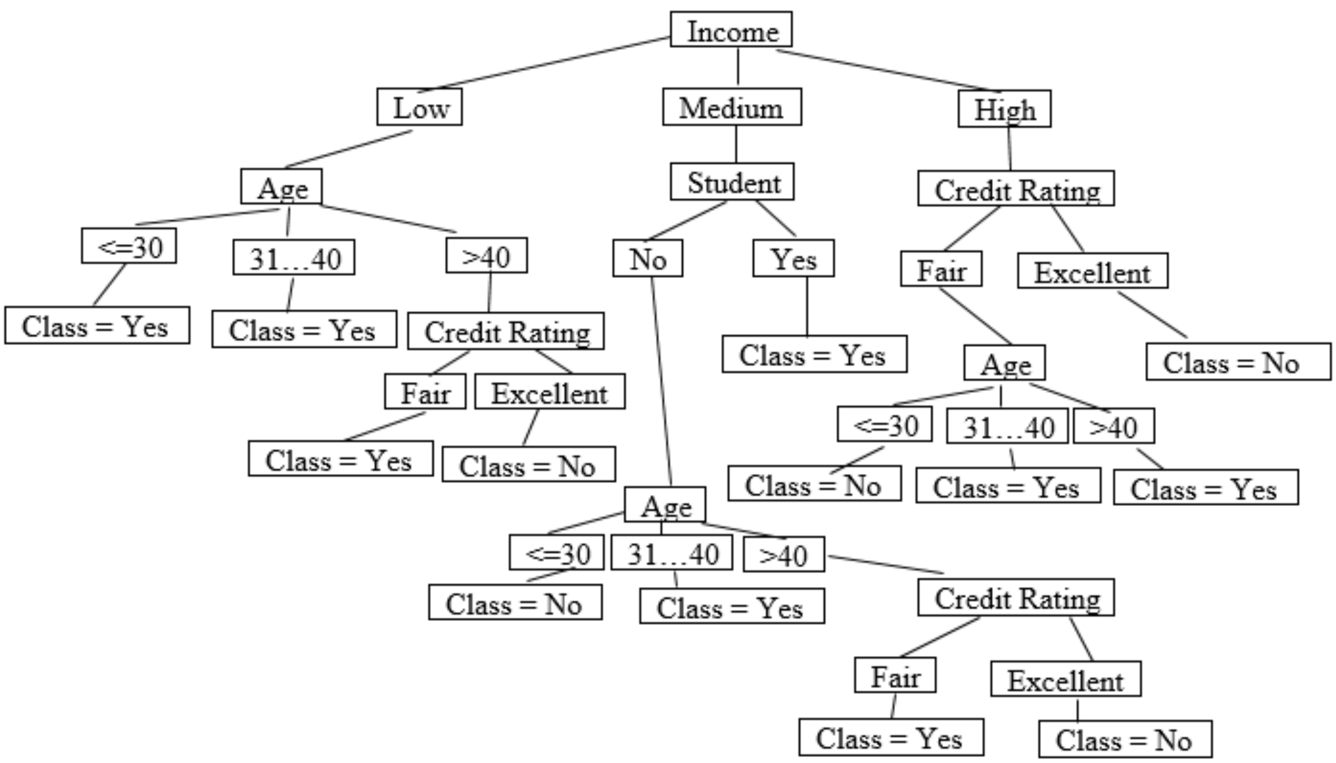
Obj	Age	Income	Student	Credit_Rating	Class
1	<=30	Low	Yes	Fair	Yes
2	>40	High	Yes	Excellent	No
3	31…40	High	Yes	Excellent	No
4	>40	Low	No	Fair	Yes
5	>40	Medium	Yes	Excellent	Yes
6	31…40	Low	No	Excellent	Yes

Obj 1 Fits - must say which rule
Obj 2 Fits I - must say which rule
Obj 3 Fits] - must say which rule
Obj 4 Fits - must say which rule
Obj 5 Fits - must say which rule

Obj 6 Fits - must say which rule

6 Passed/out of 6=100% Predictive Accuracy

EXTRA CREDIT Solution



Node Selected: Income-Medium

P Class: buys_Computer=Yes
N Class: buys_Computer=No

Records at Node
Attributes examined
Student
Credit_Rating

Age	Student	Credit Rating	Buys Computer
>40	No	Fair	Yes
<=30	No	Fair	No
>40	Yes	Fair	Yes
<=30	Yes	Excellent	Yes
31…40	No	Excellent	Yes
>40	No	Excellent	No

$Info(D)=I(4,2)=(-4/6)\log_2(4/6)-(2/6)\log_2(2/6)=-0.667\log_2(0.667)-0.333\log_2(0.333)=-0.667(-0.584)-0.333(-1.586)=+0.390+0.528=0.918$
$Info_{Student}(D)=(2/6)I(2,0)+(4/6)I(2,2)$

$I(2,0)=(-2/2)\log_2(2/2)-(0/2)\log_2(0/2)=-1\log_2(1)-0\log_2(0)=-1(0)-0(0)=0-0=0$

$I(2,2)=(-2/4)\log_2(2/4)-(2/4)\log_2(2/4)=-0.5\log_2(0.5)-0.5\log_2(0.5)=-1(-1)-0(-1)=1-0=1$

$Info_{Student}(D)=(2/6)I(2,0)+(4/6)I(2,2)=0.333(0)+0.667(1)=0+0.667=0.667$

$Info(D)-Info_{Student}(D)=0.918-0.667=0.251$=Information Gain for Student Attribute at Node

$Info_{Credit\_Rating}(D)=(3/6)I(2,1)+(3/6)I(2,1)$

$I(2,1)=(-2/3)\log_2(2/3)-(1/3)\log_2(1/3)=-0.667\log_2(0.667)-0.333\log_2(0.333)=-0.667(-0.584)-0.333(-1.586)=+0.390+0.528=0.918$

$Info_{Credit\_Rating}(D)=(3/6)I(2,1)+(3/6)I(2,1)=0.5(0.918)+.5(0.918)=0.459+0.459=0.918$

$Info(D)-Info_{Credit\_Rating}(D)=0.918-0.918=0$=Information Gain for Credit_ Rating Attribute at Node

Information Gain on Student is Greater than on Credit Rating