

CSE352

HOMEWORK 2 part1, part 2, SOLUTIONS

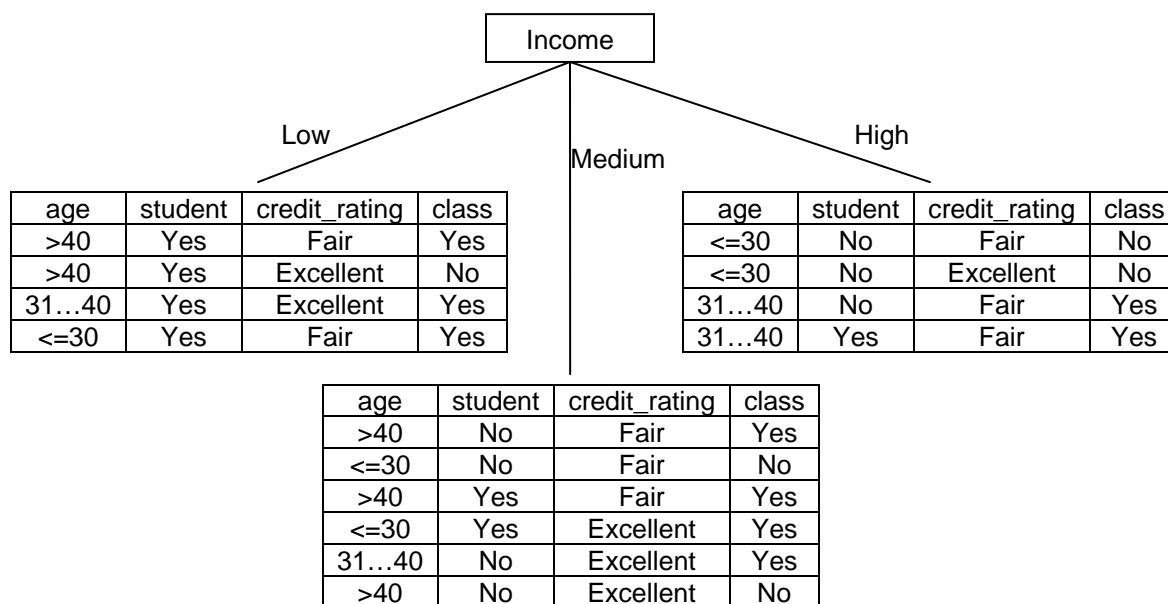
Problem 1

Use the Training Data to create two decision trees:

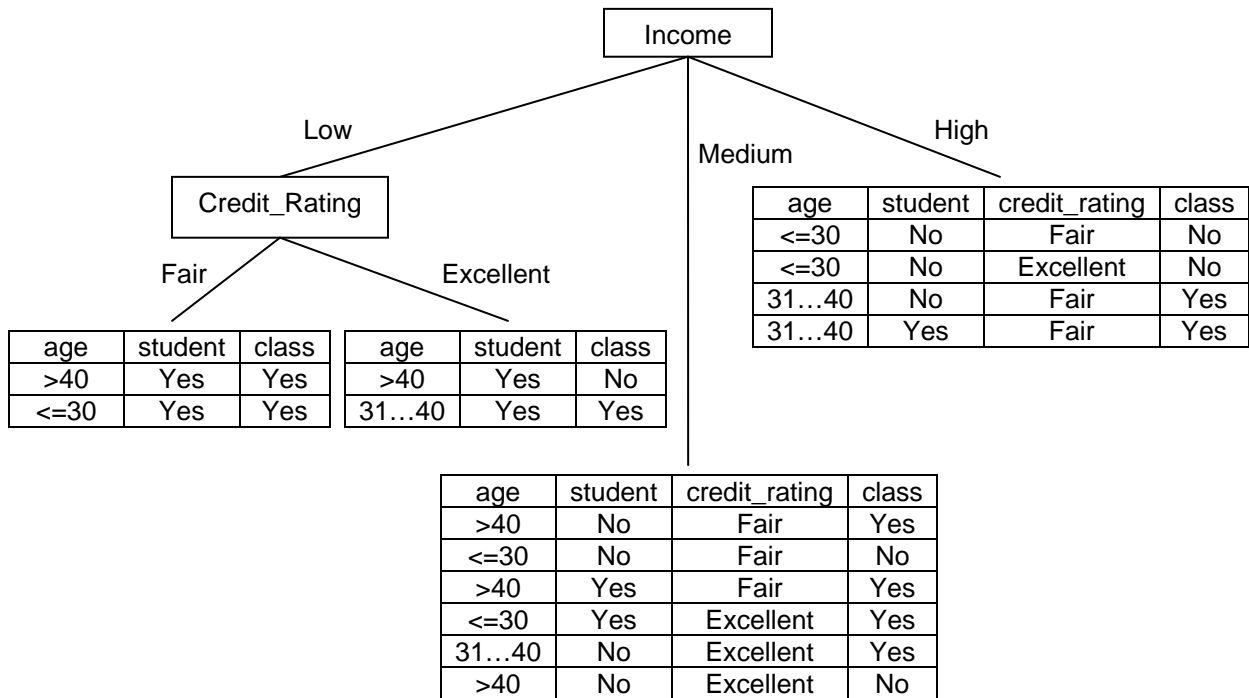
1. one without majority voting;
2. one with general majority voting , i.e. majority voting at any node of your choice.
3. For the tree 1. use INCOME as root attribute, and nodes attributes of your choice;
4. For each of the trees 1. and 2. write down the set of rules determined by the tree.
5. For each set of obtained rules compute PREDICTIVE ACCURACY.

We start with a decision tree without general majority voting.

We build the tree starting with the attribute “Income” as the root.

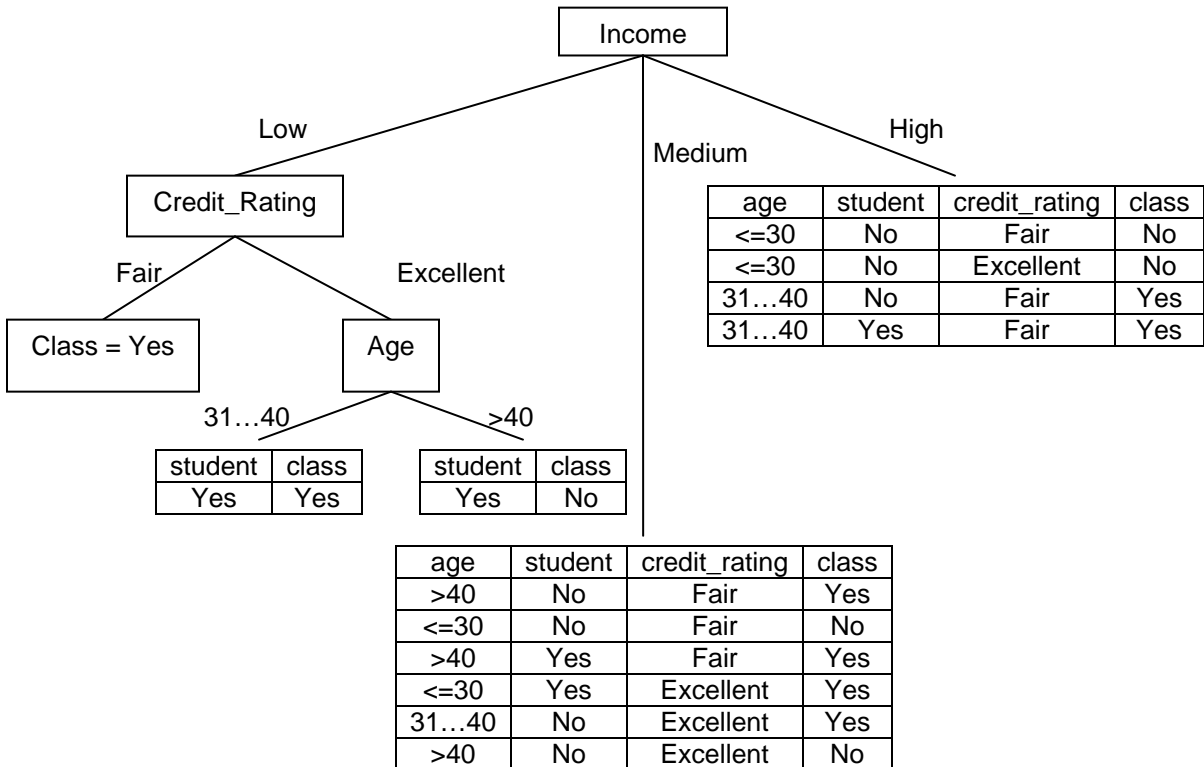


Let's choose the attribute "Credit_Rating" for the "Low" Branch.

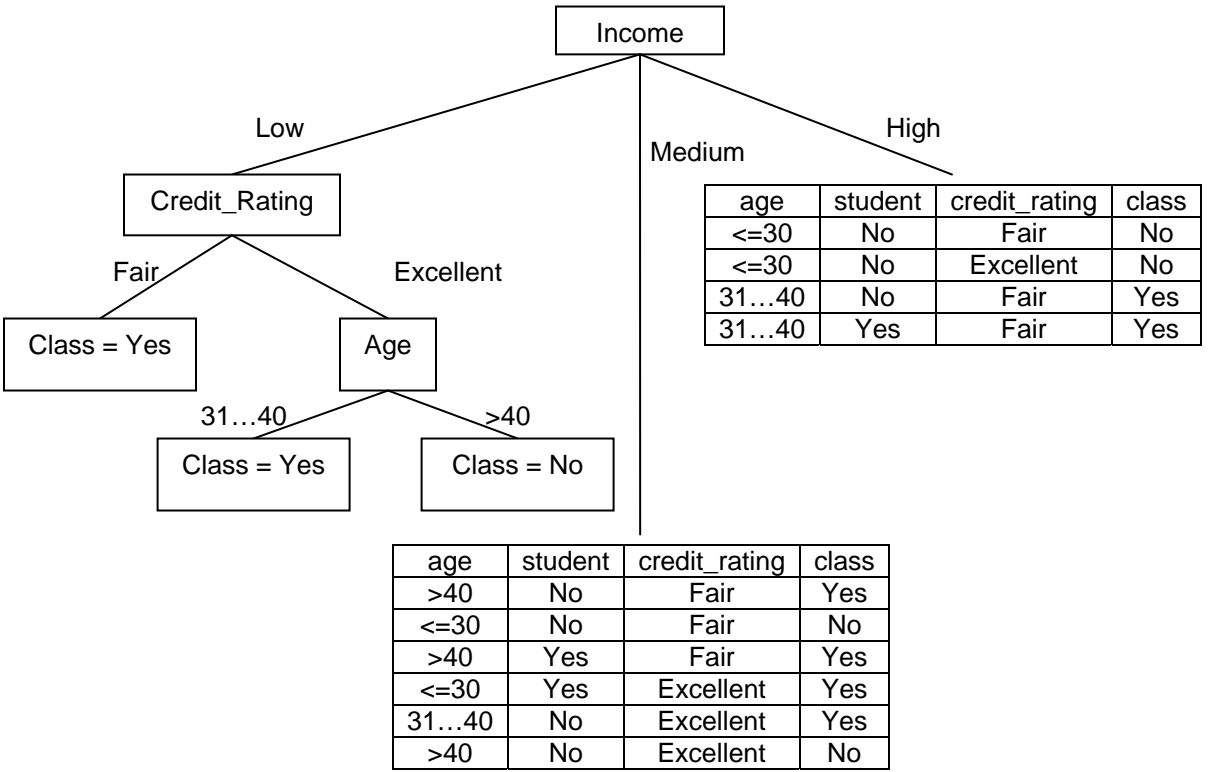


For the "Fair" branch, we set the class to "Yes".

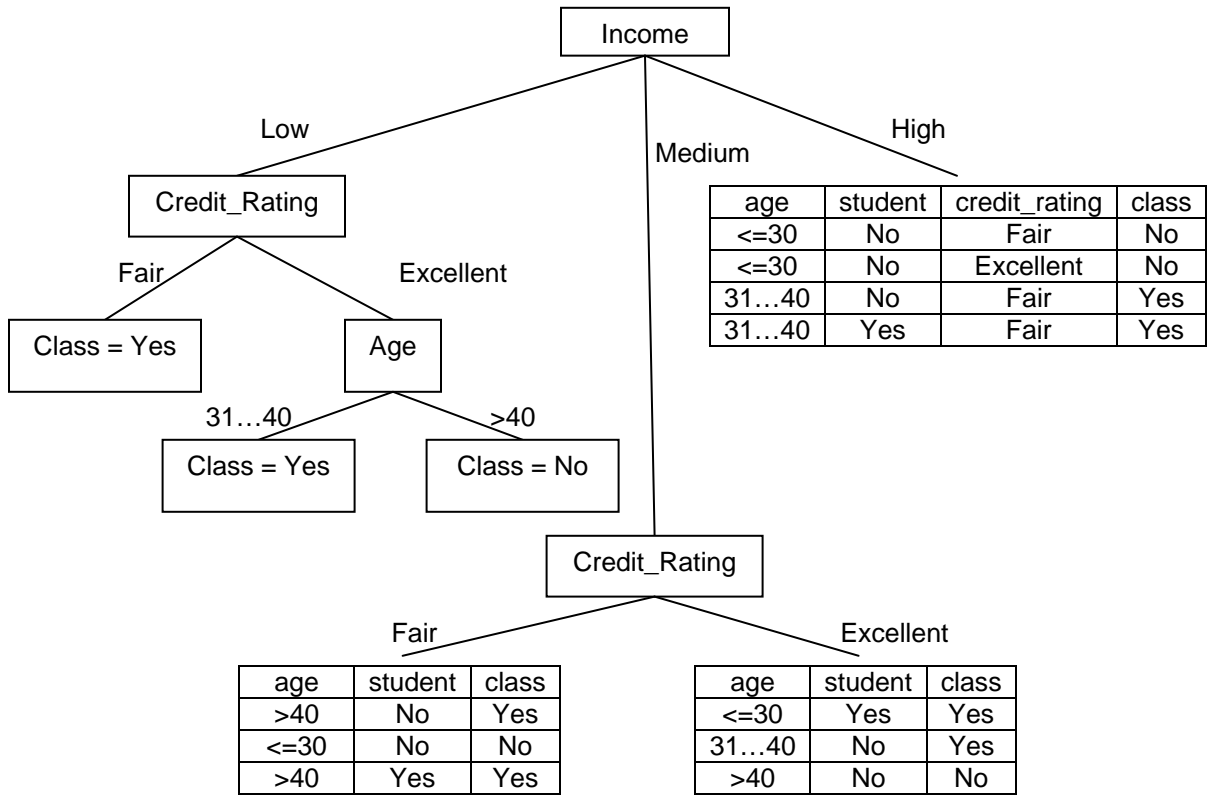
For the "Excellent" branch, we choose the attribute "Age".



For the “31...40” branch, we set the class to “Yes”.
For the “>40” branch, we set the class to “No”.



Let's choose the attribute "Credit_Rating" for the "Medium" branch.

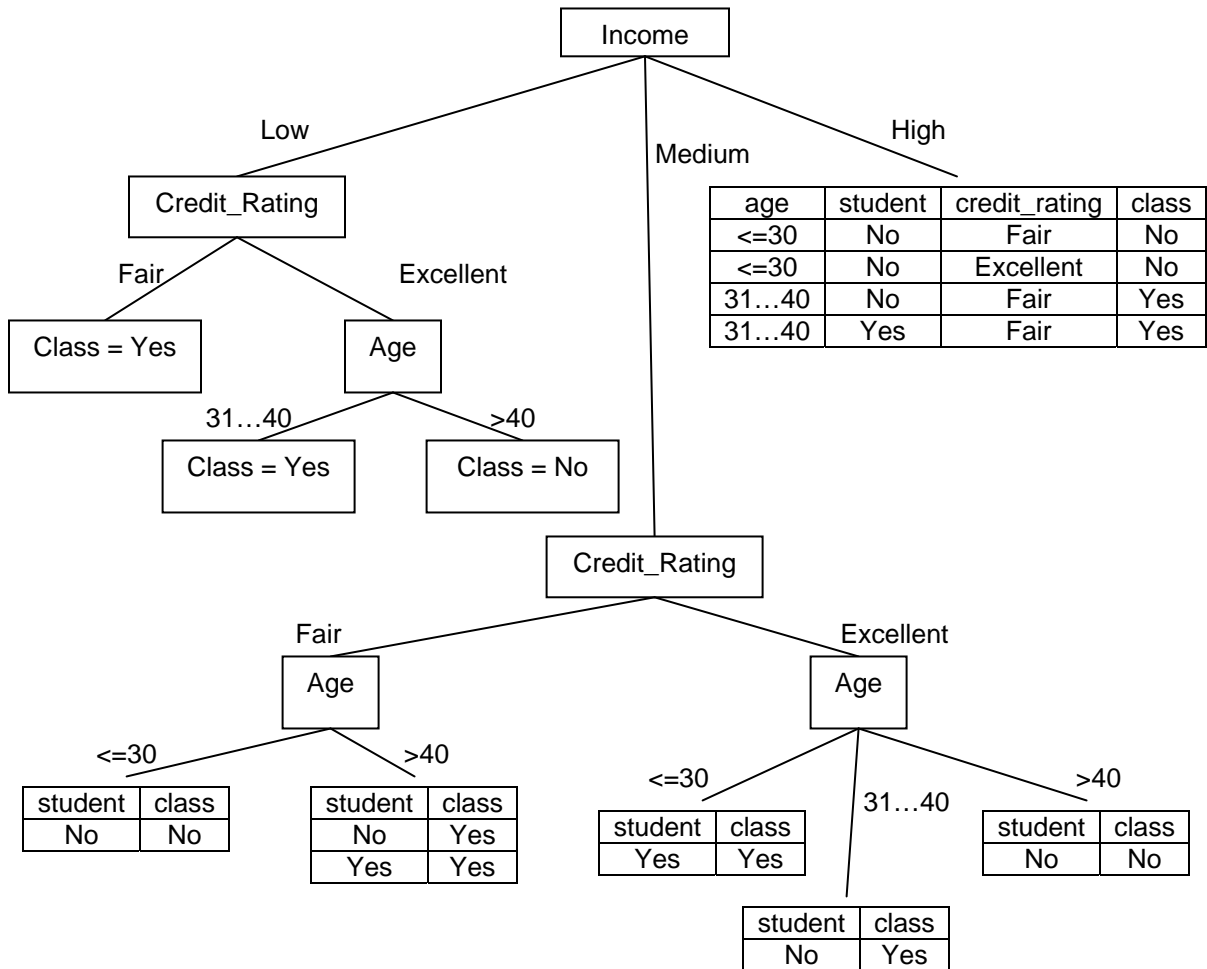


age	student	credit_rating	class
<=30	No	Fair	No
<=30	No	Excellent	No
31...40	No	Fair	Yes
31...40	Yes	Fair	Yes

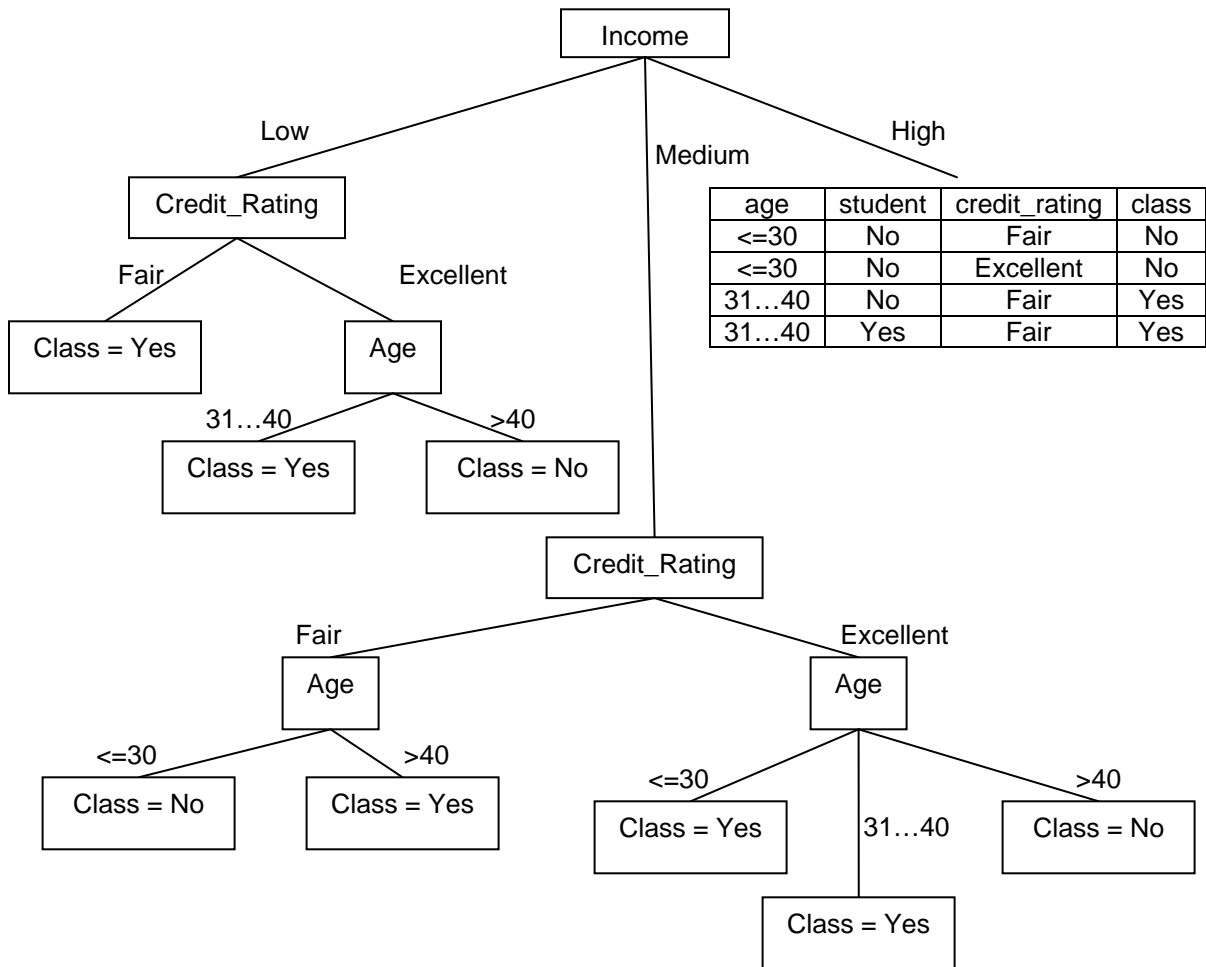
age	student	class
>40	No	Yes
<=30	No	No
>40	Yes	Yes

age	student	class
<=30	Yes	Yes
31...40	No	Yes
>40	No	No

For the “Fair” branch and the “Excellent” branch, we choose the attribute “Age”.

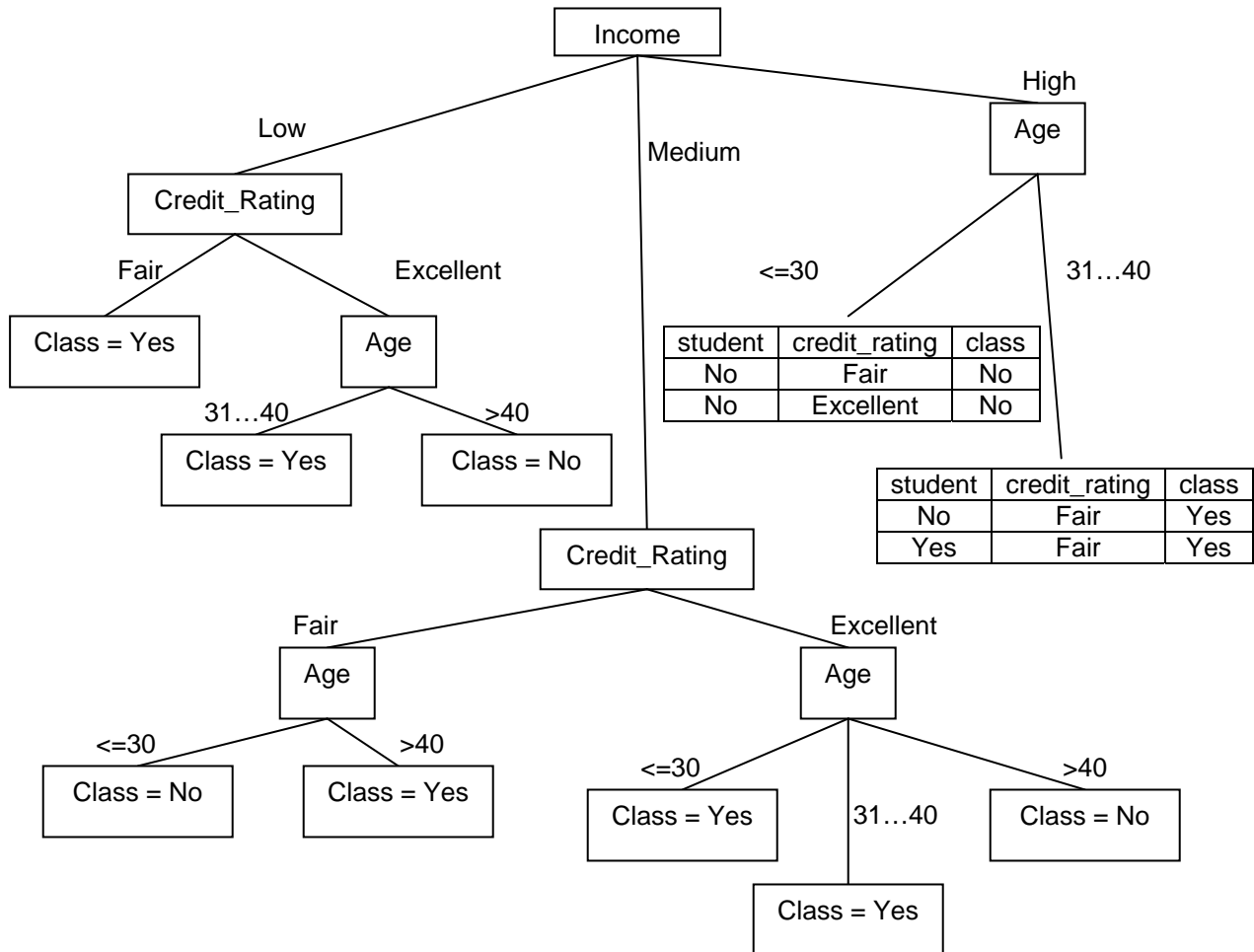


For the left " ≤ 30 " branch, we set the class to "No".
 For the left " > 40 " branch, we set the class to "Yes".
 For the right " ≤ 30 " branch, we set the class to "Yes".
 For the right " $31 \dots 40$ " branch, we set the class to "Yes".
 For the right " > 40 " branch, we set the class to "No".



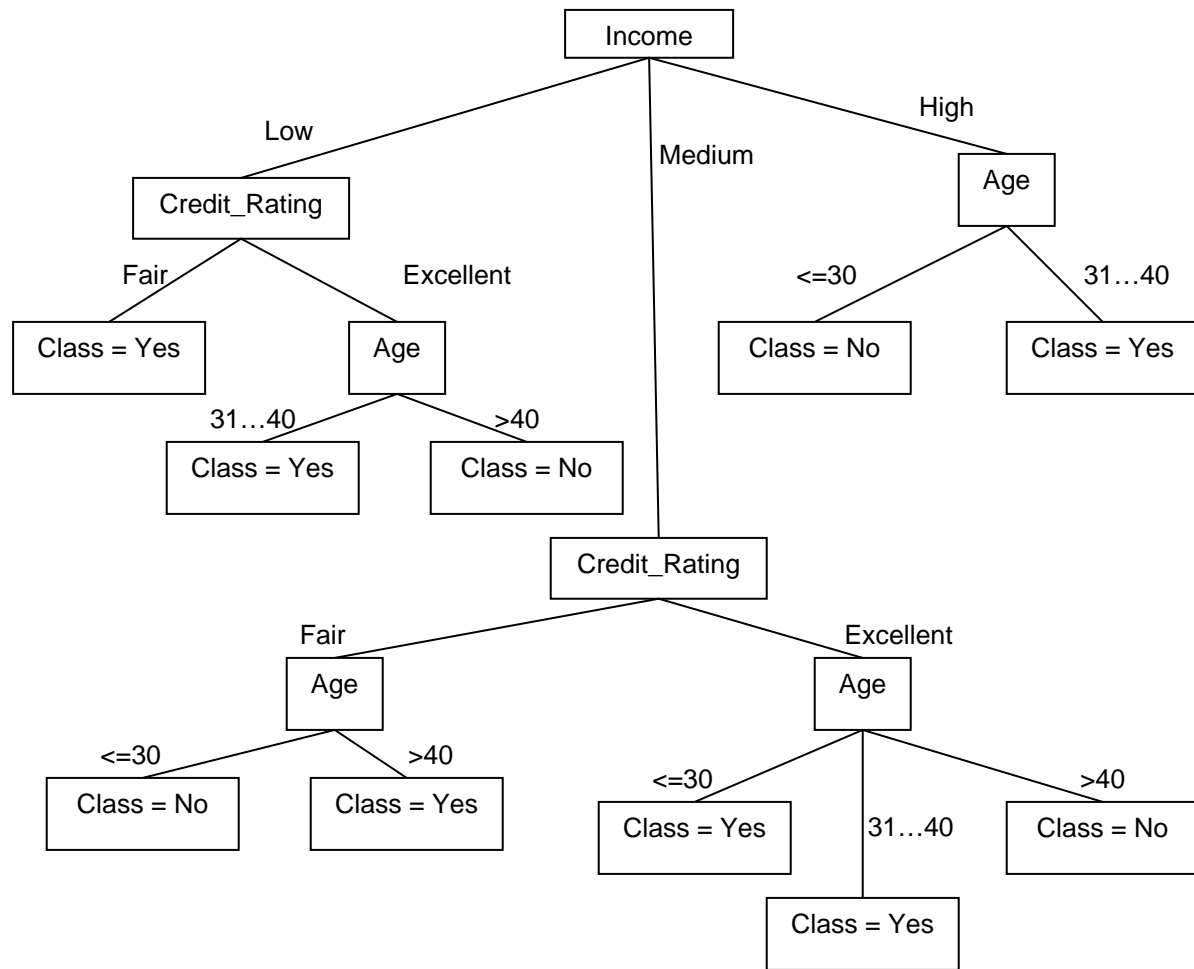
age	student	credit_rating	class
≤ 30	No	Fair	No
≤ 30	No	Excellent	No
$31 \dots 40$	No	Fair	Yes
$31 \dots 40$	Yes	Fair	Yes

For the “High” branch, we can branch on the attribute “Age”.



For the “ ≤ 30 ” branch, we set the class to “No”.

For the “31...40” branch, we set the class to “Yes”.



Discriminant rules determined by the decision tree.

IF Income = “Low” AND Credit_Rating = “Fair”

THEN Buys_Computer = “Yes”

IF Income = “Low” AND Credit_Rating = “Excellent” AND Age = “31...40”

THEN Buys_Computer = “Yes”

IF Income = “Low” AND Credit_Rating = “Excellent” AND Age = “>40”

THEN Buys_Computer = “No”

IF Income = “Medium” AND Credit_Rating = “Fair” AND Age = “ ≤ 30 ”

THEN Buys_Computer = “No”

IF Income = “Medium” AND Credit_Rating = “Fair” AND Age = “>40”

THEN Buys_Computer = “Yes”

IF Income = "Medium" AND Credit_Rating = "Excellent" AND Age = "<=30"
 THEN Buys_Computer = "Yes"

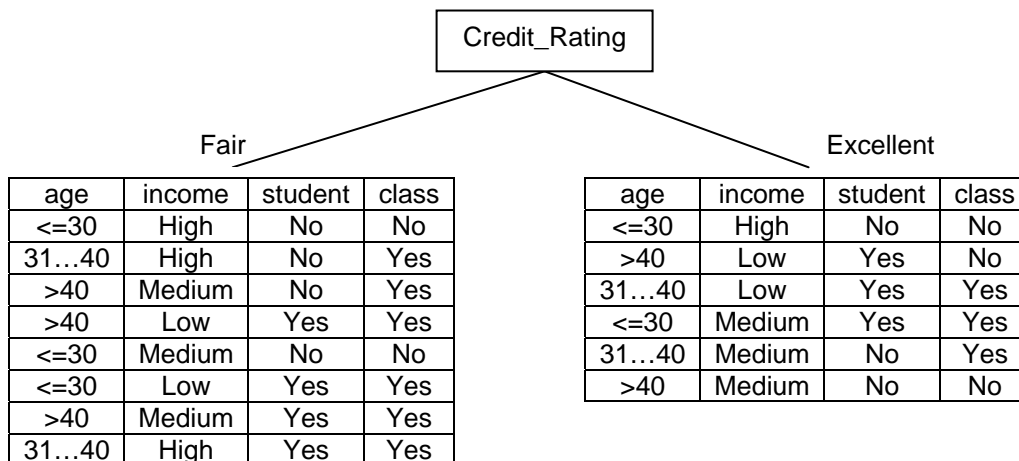
IF Income = "Medium" AND Credit_Rating = "Excellent" AND Age = "31...40"
 THEN Buys_Computer = "Yes"

IF Income = "Medium" AND Credit_Rating = "Excellent" AND Age = ">40"
 THEN Buys_Computer = "No"

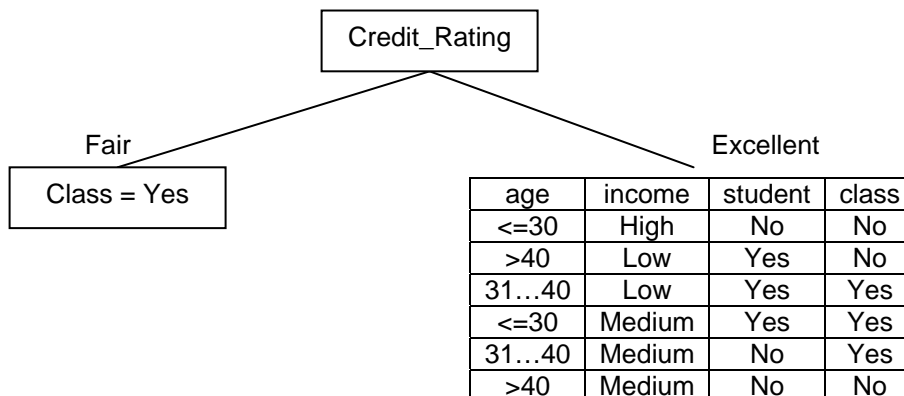
IF Income = "High" AND Age = "<=30"
 THEN Buys_Computer = "No"

IF Income = "High" AND Age = "31...40"
 THEN Buys_Computer = "Yes"

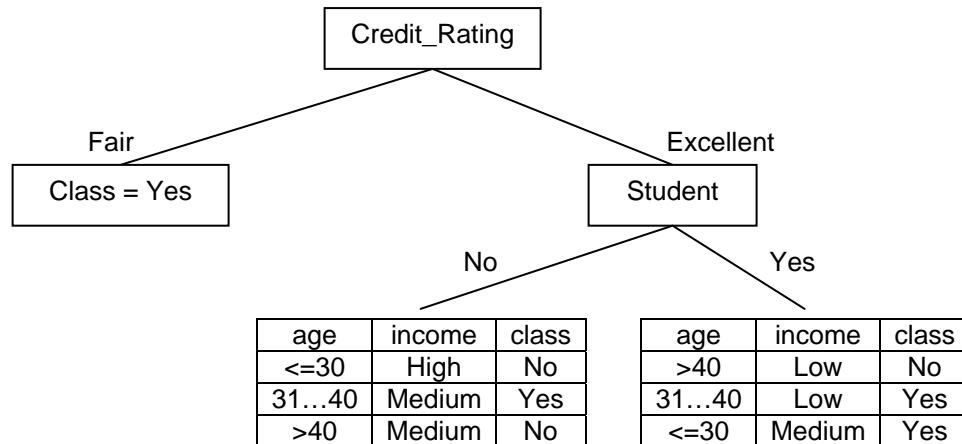
2. The decision tree with majority voting and root attribute "Credit_Rating" as the root.



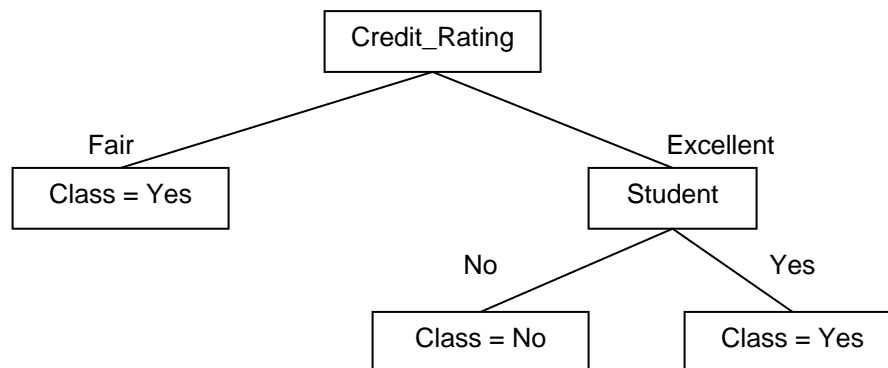
Let's do majority voting on the "Fair" branch. So, for the "Fair" branch, we set the class to "Yes".



For the “Excellent” branch, we choose the attribute “Student”.



Now we can do majority voting on the “No” and “Yes” branches. For the “No” branch, the class is set to “No”. For the “Yes” branch, the class is set to “Yes”.



Discriminant rules from the decision tree 2.

IF Credit_Rating = “Fair”

 THEN Buys_Computer = “Yes”

IF Credit_Rating = “Excellent” AND Student = “No”

 THEN Buys_Computer = “No”

IF Credit_Rating = “Excellent” AND Student = “Yes”

 THEN Buys_Computer = “Yes”

5. Predictive Accuracy Computation

First, we test the rules from the decision tree 1. (without majority voting)

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

The predictive accuracy of the set of rules from the decision tree without majority voting is $\frac{3}{6} = 50\%$.

Second, we test the rules from the decision tree 2. with majority voting.

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

The predictive accuracy of the set of rules from the decision tree with majority voting is $\frac{3}{6} = 50\%$.

Problem 2

Create test data for your rules that guarantees 100% predictive accuracy.

Here's test data that guarantees 100% predictive accuracy for the set of rules from the decision tree without majority voting.

Obj	Age	Income	Student	Credit_Rating	Class
1	<=30	Low	Yes	Fair	Yes
2	31...40	Low	Yes	Excellent	Yes
3	>40	Low	Yes	Excellent	No
4	<=30	Medium	No	Fair	No
5	>40	Medium	No	Fair	Yes
6	<=30	Medium	Yes	Excellent	Yes
7	31...40	Medium	Yes	Excellent	Yes
8	>40	Medium	Yes	Excellent	No

Here's test data that guarantees 100% predictive accuracy for the set of rules from the decision tree with majority voting.

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

Problem 3

Create test data for your rules that guarantees 0% predictive accuracy.

We do it for the tree with general majority voting; others are constructed similarly.

RULES:

IF Credit_Rating = "Fair"

THEN Buys_Computer = "Yes"

IF Credit_Rating = "Excellent" AND Student = "No"

THEN Buys_Computer = "No"

IF Credit_Rating = "Excellent" AND Student = "Yes"

THEN Buys_Computer = "Yes"

Test Data

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

Problem 4

Compute predictive accuracy of the rules from the lecture notes.

Verification of the predictive accuracy for “Lecture Notes”:

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

Predictive accuracy is $5/6 = 83.3\%$.