Lecture – 59
Attribute Selection Measures in CART – II

In this lecture, we are going to see how to select attributes in CART model. In our previous lectures, we have seen how to choose attributes by using gain value. In this lecture, there are another two methods for choosing the attributes. One is gain ratio; another one is Gini index. This lecture we are going to see other two criteria for choosing the attributes.

(Refer Slide Time: 00:51)

First, we will look at the Gain ratio. The information gain measure is biased towards tests with many outcomes. That is, it prefers to select attributes having a large number of values. For example, consider an attribute that act as a unique identifier, such as product ID.

(Refer Slide Time: 01:41)

Gain Ratio

- Because each partition is pure, the information required to classify dataset D based on this partitioning would be $\text{Info}_{\text{product ID}}(D) = 0$
- Information Gain = Info D - $\text{Info}_{\text{product ID}}(D)$ = maximum
- Therefore, the information gained by partitioning on this attribute is maximal
- Clearly, such a partitioning is useless for classification
- Gain ratio is an extension to information gain which attempts to overcome this bias

A split on product ID would result in a large number of partitions as many as there values each one containing just one tuple because each partition is pure, the information required to classify dataset D based on this partitioning would be Info product ID will be 0, because there might be only one value in each partition. So Information gain is we know the formula, the formula for information gain is Info D – Info product ID for D because this value is going to be 0, so the information gain will be maximum.

Therefore, the information gained by partitioning on this attribute is maximal. Clearly, such partitioning is useless for classification, because there are going to be one element for each partition. So, the gain ratio is an extension to information gain, which attempts to overcome this bias.

(Refer Slide Time: 02:14)

Split information

- It applies a kind of normalization to information gain using a "split information" value defined analogously with Info(D) as:

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right).$$

- D_j = single partion
- D = Data set
- This value represents the potential information generated by splitting the training data set, D, into v partitions, corresponding to the v outcomes of a test on attribute A

Let us see what is split information. It applies a kind of normalization to information gain using a split information value defined analogously with Info D. So, how to find out the split info for level for $D = -\Sigma^v_{j=1}$, v is different levels; (lDjl divided by lDl), multiplied by log (lDjl divided by lDl) to the base 2. Here the Dj is the single partition. In the next example, I will explain the Dj. D is dataset. So, this value that is the split info represents the potential information generated by splitting the training data set D into v partitions, corresponding to the v outcomes of test on attribute A.

(Refer Slide Time: 03:11)

So, let us see the formula for the gain ratio. The gain ratio differs from the information gain, which measures the information with respect to classification that is acquired based on the same partitioning. So, the gain ratio is (Gain of A) / (Split Info of A). The attribute with the maximum gain ratio is selected as the splitting attribute. I have an example; with that I can explain how to use this gain ratio for choosing the attribute.

(Refer Slide Time: 03:43)

Gain Ratio example

- Consider the previous example for computation of gain ratio for the attribute income
- A test on income splits the data of the following Table into three partitions, namely low, medium, and high, containing four, six, and four tuples, respectively

Han, J., Pei, J. and Kamber, M., 2011. *Data mining: concepts and techniques.* Elsevier.

RID	age	income	student	credit.rating	Class: buys.computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle.aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle.aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle.aged	medium	no	excellent	yes
13	middle.aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

Look at this example. In my previous lecture, I was showing this dataset. Consider in our previous example for computation of gain ratio for the attribute income. So, we are going to take this variable first. We are going to find out gain ratio. I am going to say the procedure for only one attribute, like that you have to try for age attribute, for student attribute, for credit rating attribute.

Whichever is high that variable should be chosen for classification. A test on income splits the data into following table into three partitions. See, when you look at this income column, there are three levels, one is low, medium, and high containing in low, there are 6 element is there, in medium 6 is there, in high there is 4. This example is taken from the book data mining concepts and techniques.

(Refer Slide Time: 05:04)

Calculate Entropy for 'low'

- Low :

Low	Class: buys computer
Yes	3
No	1

- Calculate Entropy for Low:

$$= -(3/4)\log_2(3/4) - (1/4)\log_2(1/4)$$

RID	age	income	student	credit.rating	Class: buys.computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle.aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle.aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle.aged	medium	no	excellent	yes
13	middle.aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

Now, let us calculate the Entropy for level low. In level low, how many have answered yes to by computer when it is low. So how many people have answered low. When the level is low, so the entropy is –(3 / 4) log 3 / 4 to the base 2 – (1 / 4)log 1 / 4 to the base 2.

(Refer Slide Time: 05:44)

Calculate Entropy for buying class D

- Calculate information:
- $= -p_y \log_2 (p_y) - p_n \log_2 (p_n)$
- Where p_y is probability of yes and p_n is probability of no

$$Info(D) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) = 0.940 \text{ bits.}$$

For finding the information gain for the attribute income, first we need to know the entropy for the class D. We know that how to find the entropy for the class D. It is – py log py to the base 2 – pn log pn to the base 2. So, the py is in our class, how many people have answered yes. When you look at the table, there was 9 yes and there is 5 no. So, for s it is 9 / 14, log 9 / 14 to the base 2 minus no. There is 5 no, so it is 5 no, so 5 / 14, log 5 / 14 to the base 2 equal to 0.940 bits.

(Refer Slide Time: 06:43)

Gain of income

- The expected information needed to classify a tuple in D if the tuples are partitioned according to income is:
- $Info_{income}(D) = (4/14) (-(2/4)\log_2(2/4) - (2/4)\log_2(2/4)) +$
  $(6/14) (-(4/6)\log_2(4/6) - (2/6)\log_2(2/6)) +$
  $(4/14) (-(1/4)\log_2(1/4) - (3/4)\log_2(3/4))$
  $= 0.911 \text{ bits}$

Gain of income : $Info(D) - Info_{income}(D)$
  $= 0.94 - 0.911 = \boxed{0.029}$

Now, let us find out the information gain for the attribute income. What is the meaning is, if you use income as an attribute for the classification, how much information you can gain. So the expected information needed to classify a tuple in D if the tuples are partitioned according to income is, so Info for attribute income is 4 / 14, we got this 4 for the attribute income, there was three levels low, medium, and high.

In low, there was a 4, in medium there was 6, and in high there was 4 values. So, it is a kind of weighted attributed, because it is nothing but expected information needed. So, it is nothing but our weighted entropy. So, the weighted entropy is 0.911 bits. So, the gain of income is Info D – Info for the attribute income. So, we got this 0.94, which we got from this value, $0.94 – 0.911 = 0.029$.

(Refer Slide Time: 8:08)

Gain-Ratio(income)

- Calculation of split ratio:

$$SplitInfo_A(D) = -\frac{4}{14} \times \log_2\left(\frac{4}{14}\right) - \frac{6}{14} \times \log_2\left(\frac{6}{14}\right) - \frac{4}{14} \times \log_2\left(\frac{4}{14}\right)$$

$$= 0.926.$$

- Therefore, Gain-Ratio(income) = 0.029/0.926 = 0.031

Now, we will go for split ratio. So, the split ratio is equal to – (4 / 14) multiplied log (4 / 14) base 2 – (6 / 14) log (6 / 14) to the base 2 – (4 / 14) log (4 / 14) to the base 2. So, split info is 0.926. Therefore, the Gain-Ratio for the attribute income is 0.029 / 0.926 is equal to 0.031.

**(Refer Slide Time: 09:05)**

Interpretation

- Further we calculate the same for the rest 3 criteria (age, student, credit rating)
- The one with maximum Gain ratio value will results in the maximum reduction in impurity of the tuples in D and is returned as the splitting criterion

Further, we calculate the same for the rest of 3 criteria, what are the other three attributes, age is there, student is there, credit rating is there. The one with the maximum Gain ratio value will result in maximum reduction in impurity of the tuples in D and is returned as the splitting criterion. So, what we have to do, we have seen for one attribute that is income. There are another 3 attributes, such as age, student and credit rating. For these attributes also, we have to find out the information gained ratio.

That corresponding attribute should be taken as the splitting criterion. We have found the Gain ratio for the attribute income. The same way, there are other attributes like age, student and credit rating. For these attributes also, we have to find out the Gain ratio. One with the maximum gain ratio value will result in maximum reduction and impurities of the tuples in D and is written as the splitting criterion.

(Refer Slide Time: 10:09)
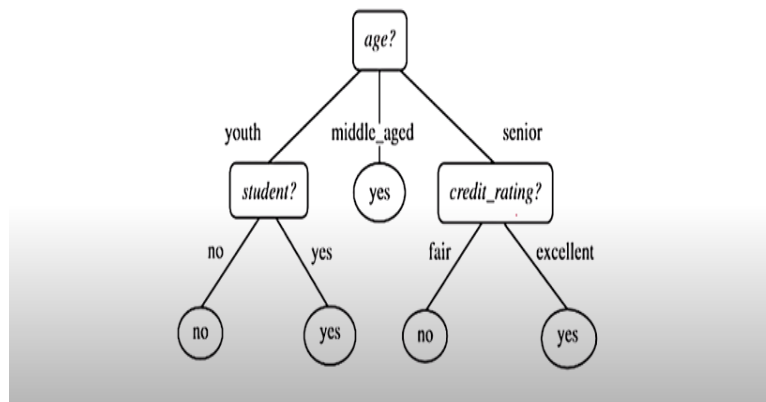
For example, assume that the attribute age is having maximum gain ratio, so that variable should be chosen as the splitting variable. Then from there, if there is a student, assume that the attribute age is having highest gain ratio, so that age is taken as the splitting variable. Then, there is a student. There are remaining other variables, for example, student, credit rating and so on. So, out of these, again you have to find the Gain ratio. So, out of this, which one is giving the maximum Gain ratio, that should be taken as the splitting criterion.

(Refer Slide Time: 10:50)

Decision tree using Gini index

- Let's take the Introduction of a decision tree using Gini index
- Let D be the training data of the following table

Han, J., Pei, J. and Kamber, M., 2011. *Data mining: concepts and techniques*. Elsevier.

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

Next, we will go to another criterion for choosing the attribute, that is Gini index. Let us take the introduction of decision tree using Gini index. Let D be the training data of the following table. So, this data also taken from the book data mining, concepts and techniques, the source is given here. Now, we are going to see how to find out the Gini index.

(Refer Slide Time: 11:12)

Example

- In this example, each attribute is discrete-valued
- Continuous-valued attributes have been generalized
- The class label attribute, buys computer, has two distinct values (namely, {yes, no}); therefore, there are two distinct classes (that is, m = 2)
- Let class C_1 correspond to 'yes' and class C_2 correspond to 'no'.
- There are nine tuples of class 'yes' and five tuples of class 'no'.
- A (root) node N is created for the tuples in D

In this example, each attribute is discrete-valued because all are in different category, so continuous-valued attributes have been generalized. We did not take continuous value. The class label attribute, buy computer, has two distinct values, yes or no, therefore, there are two distinct classes, m = 2. Let class C1 correspond to yes and class C2 correspond to no. There are nine tuples of class yes and five tuples of class no. A root n is created for the tuples D.

(Refer Slide Time: 12:07)

Calculation of Gini(D)

- We first use the following Equation for Gini index to compute the impurity of D:

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2,$$

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle aged	medium	no	excellent	yes
13	middle aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

$$= \quad Gini(D) = 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2 = 0.459.$$

Now, we go for calculation of Gini index. We first use the following equation for Gini index to compute the impurity of D. So, first we will find out the Gini of class D is $1 - \sum_{i=1}^{m}$, (m is number of levels), p_i^2. So, what is the p, how many s is there. So, $1 - (9/14)^2$, how many no is there, next level, so the Gini of class D = $1 - (9/14)^2 - (5/14)^2$, that is 0.459.

(Refer Slide Time: 12:54)

Gini index for income attribute

- Lets calculate Gini index for income attribute
- To find the splitting criterion for the tuples in D, we need to compute the Gini index for each attribute
- Let's start with the attribute income and consider each of the possible splitting subsets
- Income has three possible values, namely {low, medium, high}, then the possible subsets are {low, medium, high}, {low, medium}, {low, high}, {medium, high}, {low}, {medium}, {high}, and {}
- Power set and empty set will not be used for splitting

Let us calculate, Gini index, previously found the Gini, the Gini index for income attribute. To find the splitting criterion for the tuples in D, we need to compute the Gini index for each attribute. In this, you take an example, income. Let us start with the attributes income and consider each of the possible splitting subsets. Incomes has three possible values, namely low, medium, high.

The possible subsets are low, medium, high, then all possible combinations low and medium, low and high, medium and high, then values which is having one value in the set, low, medium, high and null set. Power set and empty set will not be used for splitting. What is the power set where all the element is there, for example, low, medium, high is the power set, the null set is nothing but the empty set.

(Refer Slide Time: 14:00)

Gini index for income attribute

- Consider the subset{low, medium}
- This would result in 10 tuples in partition D1 satisfying the condition "income ∈{low, medium}"
- The remaining four tuples of D (high) would be assigned to partition D2

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

So this will not be used for splitting. Now, we are going to split into two category. One is subset low and medium, because it is a binary classification, when you go for low and medium, another group. Suppose, there are two group is there, group 1 and group 2, in group 1, we have taken low and medium, obviously another set will be high. So, this would result in 10 tuples in partition D1.

We have to count how many low medium is there. We have got 10 tuples in partition D1, so this group is D1, that is why the condition income in the set low and medium. The remaining four tuples of D, the remaining is high, that is the remaining 4 that is D2, the remaining four tuples of D would be assigned to partition D2. What has happened, we have made a two subset, one is low medium and the remaining one is high.

(Refer Slide Time: 15:09)

Tuples in partition D1

- Low + Medium:

Medium + Low	Class: buys computer
Yes	3+4 =7
No	1+ 2 = 3

RID	age	income	student	credit.rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

For the low and medium, for the class variable buys computer, we are going to see how many yes is there. By looking at together, the level medium and low, there are 7 yes is there, and there are 3 no is there.

(Refer Slide Time: 15:28)

Tuples in partition D2

- High : (D₂)

High	Class: buys computer
Yes	2
No	2

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

For high, because that was group D1, this is for group D2, how many yes is there, 2, how many no is there, 2 no is there. So, two yes is there and 2 no is there.

(Refer Slide Time: 15:38)

Gini index for income attribute

- The Gini index value computed based on this partitioning is

$$Gini_{income \in \{low, medium\}}(D)$$
$$= \frac{10}{14}Gini(D_1) + \frac{4}{14}Gini(D_2)$$
$$= (10/14)(1 - (7/10)^2 - (3/10)^2) +$$
$$(4/14)(1 - (2/4)^2 - (2/4)^2)$$
$$= 0.443 = Gini_{income \in \{high\}}$$

The Gini index for income attribute. The Gini index value computed based on this partitioning is Gini income to the set low and medium, so (10 / 14) Gini D1 + (4 / 14) Gini D2, so how we got this four, by looking at the low and medium, there are 10 values is there, out of 14, so 10 / 14. The another set, the remaining is 4. In D2, there is only 4, so 4 / 14. So Gini D we have seen in the previous lecture, 1 – (7 / 10) whole square – (3 / 10) the whole square.

How we got this 7, you see that there are 7 yes and there are 3 no. That is why it is 7 / 10 whole square minus 3 / 10 whole square. For D2, there are two yes, there are two no. So, (4 /

14) (1 – (2 / 4) whole square – (2 / 4) whole square), so this value is 0.443, so this is Gini value for the income high. For example, if you found the Gini value for low and medium, that is equivalent to finding the Gini value for the next group, that is the high.

(Refer Slide Time: 17:00)

Consider the next subset, that is high and medium. When you look at high and medium, there will be 10 tuples in partitioning D1, so here also we are going for two group, one group is D1, and another group is D2. Here high and medium is in one set, that is D1, so when you go for high and medium, obviously the next will be low. Low will be in the set D2. High and medium, there will be 10 tuples satisfying the condition, remaining 4 tuples of D would be assigned to partitioning D2.

(Refer Slide Time: 17:57)

When you look at high and medium together, 6 yes for buying computer, 4 no for buying computer.

(Refer Slide Time: 18:10)

Then the next group, that is the tuples in partition D2, for low, there is one person has answered no, and 3 people has answered yes for buys a computer.

(Refer Slide Time: 18:22)

Now, we will find out Gini index value computed based on this partitioning. Gini index for income attribute, Gini index value based on the computed partitioning for Gini is for income, we had two category, one is D1 and D2. D1, we had high and medium, and D2 we had low. Now for high and medium, let us find out the Gini index because there was 10 out of 14 that comes high and medium, so the Gini for D1 plus 4, in this low, there was 4 out of 14, so Gini for D2.

So, what is the value of D1, because in D1, for high and medium together, there was 6 yes was there and 4 no was there. For D1, that is the Gini index value. For D2, the Gini index value is 4 / 14, 1 yes was there and 3 no was there. So 1 / 4 whole square minus 3/4 whole square, so the Gini index for high and medium group is 0.45, that is nothing but the Gini index for the group also.

(Refer Slide Time: 19:54)

Now, we go for another subset, high and low. Now, here what we are going to do. We are going to have two groups because it is a binary classification, so high, low is one group, so this is D1, and D2 obviously it is D2. So here for this, this would result that is 8 tuples in partition D1 satisfying the income is high and low, the remaining 6 tuples of D would be assigned to partitioning D2.

(Refer Slide Time: 20:27)

Tuples in partition D2

- Medium:

Low	Class: buys computer
No	2
Yes	4

RID	age	income	student	credit.rating	Class: buys.computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle.aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle.aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle.aged	medium	no	excellent	yes
13	middle.aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

For high and low, there was 5 yes is there, for high and low, there was 3 no is there. Another group is D2. In that, there was a medium. In that medium, there was a 2 no, and 4 yes.

(Refer Slide Time: 20:41)

Gini index for income attribute

- The Gini index value computed based on this partitioning is

$$\text{Gini}_{\text{income} \in \{\text{high, low}\}}$$

$$= (8/14) \, (1 - (5/8)^2 - (3/8)^2) +$$
$$(6/14) \, (1 - (2/6)^2 - (4/6)^2)$$
$$= 0.458 = \text{Gini}_{\text{income} \in \{\text{medium}\}}$$

We will find out the Gini index value computed based on the partitioning that is for this group, high and low. In high and low, totally there was 8 value, 8 / 14, in that 1 – 5 yes out of 8, so 5/8 whole square, there are 3 no, -3 / 8 whole square plus for group D2, that is medium that was in another group, in that 6 elements was there out of 14, (6 / 14) (1 – how many yes was there, two no was there, 2 / 6 whole square, four yes was there, 4 / 6 whole square). So, the Gini index for the group high and low that is equivalent to. For the medium group, Gini index value, that also 0.458.

(Refer Slide Time: 21:40)

Gini Index values

	Gini Index values
Gini $_{income \in \{high, low\}}$	0.458
Gini $_{income \in \{high, medium\}}$	0.45
Gini $_{income \in \{medium, low\}}$	0.443

We have completed all possible binary classifications, so Gini income high and low, Gini index value 0.458, the Gini group for high and medium is 0.45, for medium low, the Gini index value is 0.443. How to interpret this table, so this value 0.443, this is having the minimum Gini index value.

(Refer Slide Time: 22:11)

Interpretation

- The best binary split for attribute income is on {medium, low} (or {high}) because it minimizes the Gini index
- The splitting subset {medium,low} therefore give the minimum Gini index for attribute income
- Reduction in impurity = 0.459 – 0.443 = 0.016
- Further we calculate the same for the rest 3 criteria (age, student, credit rating)
- The one with minimum Gini index value will results in the maximum reduction in impurity of the tuples in D and is returned as the splitting criterion

The best binary split for the attribute income is on medium and low, otherwise high, because it minimizes the Gini index. When you look at the previous table, it is having the minimum Gini index value. The splitting subset, medium and low therefore gives the minimum Gini index for attribute income. So, the reduction in impurity is 0.459, this value which we got from slide number 18, this for the class D, - 0.443.

We got from this value, Gini index value for each subset. So the difference is 0.016. Further, we calculate the same for the rest of 3 criteria, so we got for reduction impurity for, this is income, there are another 3 attributes that is age, student and credit rating. For each attribute, we have to find out the reduction in impurity. The one with minimum Gini index value will result in the maximum reduction in impurity of the tuples in D and is returned as the splitting criterion.

(Refer Slide Time: 23:28)
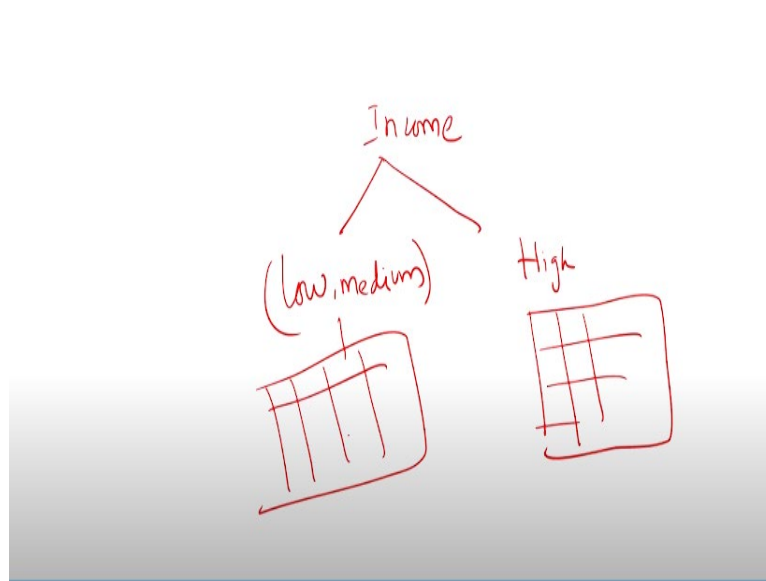
Now, how to have the classifications. For example, we have income is the splitting variable. There was a 2 binary split. The first one was low and medium is one group, and high is another group. Now, in the high also, you might have some other table like this. For each value, we have to find out the Gini index, for low and medium group also, we will have another table. For this value also, we have to find out the Gini index.

After finding out the Gini index, whichever is having highest level of reduction impurity, otherwise lowest value of Gini index should be chosen as the splitting criteria for further classification. In this lecture, I have explained how to choose an attribute for the decision tree model, by using two criteria. One is gain ratio and Gini index. For both the method, I have taken a numerical example. With the help of numerical example, I have explained how to choose an attribute. In the next lecture, we are going to use Python for making a CART model. Thank you.