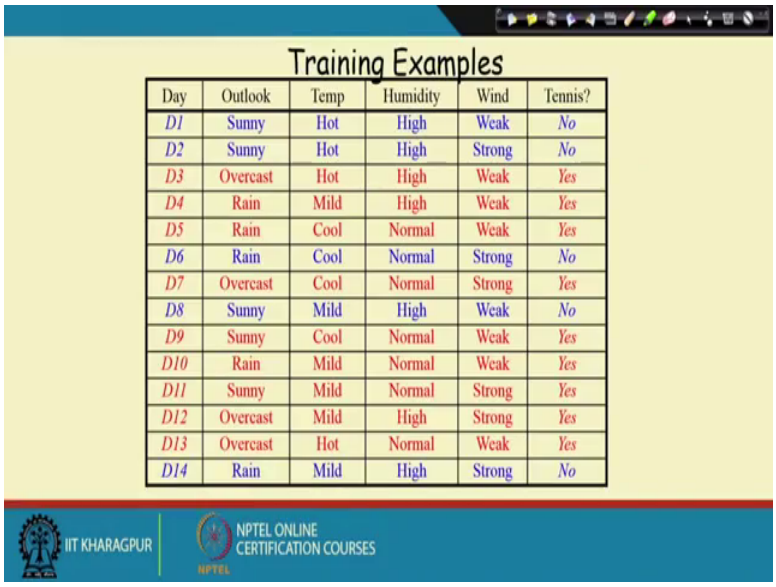
Data Mining
Prof. Pabitra Mitra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 11
Decision Tree - IV

Let me explain the process of Decision Tree construction with the training examples we consider.

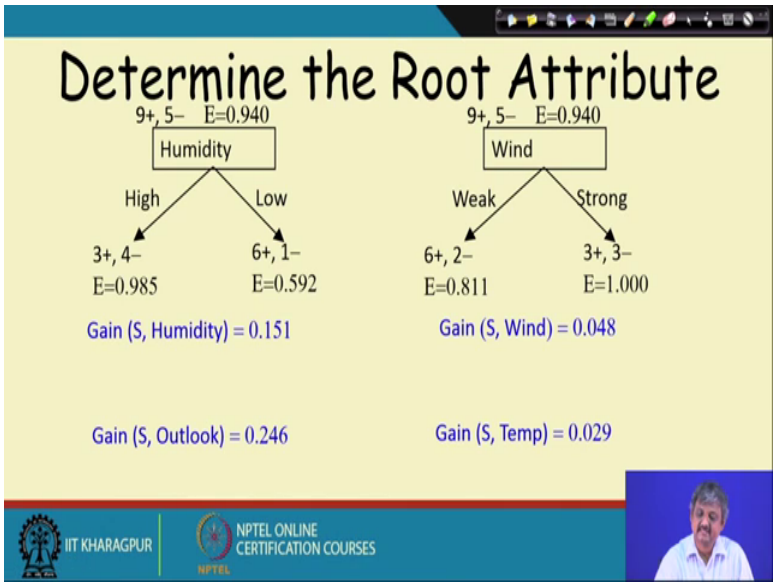
(Refer Slide Time: 00:29)



Day	Outlook	Temp	Humidity	Wind	Tennis?
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

So, we come back to our table of examples that we had seen earlier, I have 14 examples with belonging to 2 classes. So, what I do is that in order to consider I mean construct the decision tree initially I consider all the 14 examples ok.

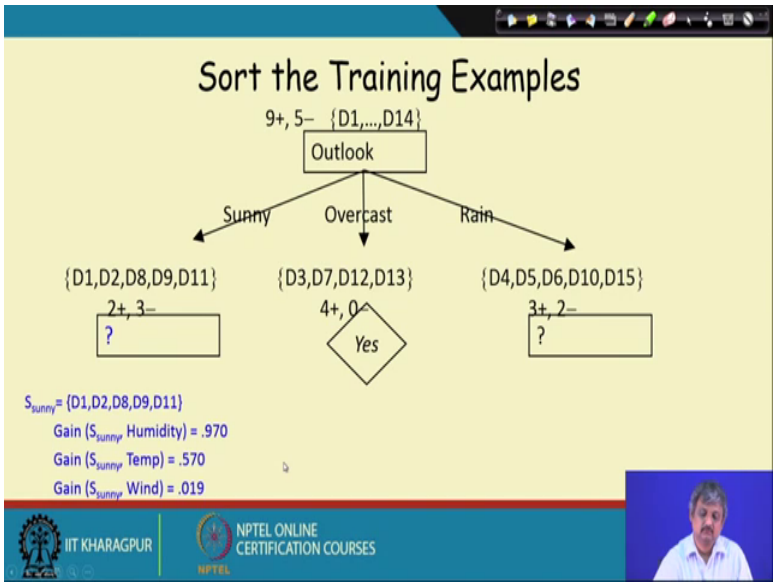
(Refer Slide Time: 00:52)



And then for each of the attributes I perform a split for example, if you see. So, there are 14 examples of the out of the 14 examples 9 belong to plus class S class 5 belongs to no class. And if we split on humidity let us say there are 2 groups high and normal and if you count 7 examples have humidity high, and 8 have humidity low. So, if we split along that and using these values of n plus and n minus, if you con compute the entropy these are the values we get.

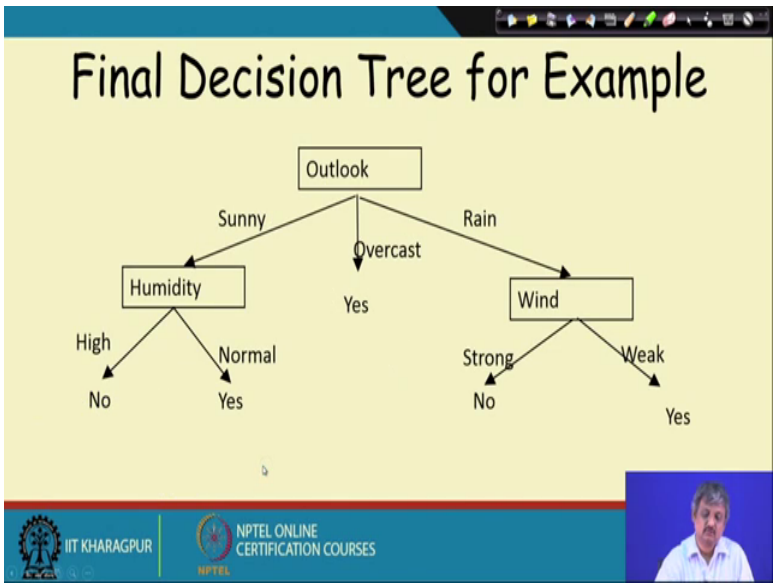
So, the gain is this much 0.985 plus 0.592 subtracted from that 0.940, so what do you do similarly for the wind we calculate, this is the gain similarly for the outlook this is the for the outlook, this is the gain, and temperature this is again. So, we see that outlook has the highest gain.

(Refer Slide Time: 02:27)



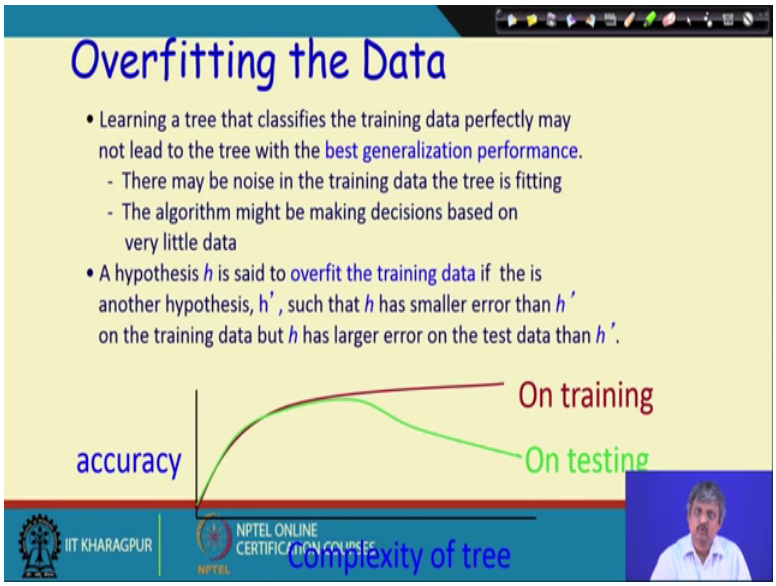
In order to construct the decision tree ill first split on outlook. So, outlook equal to sunny I have these 5 examples overcast this, and 2 of them are plus t minus you see overcast is pure only plus class so it is a leaf. So this is my new S these 5 examples, on this I recursively split again. So, gain on humid is this gain on S sunny not the full set only on S sunny wind is this, humidity is highest.

(Refer Slide Time: 03:20)



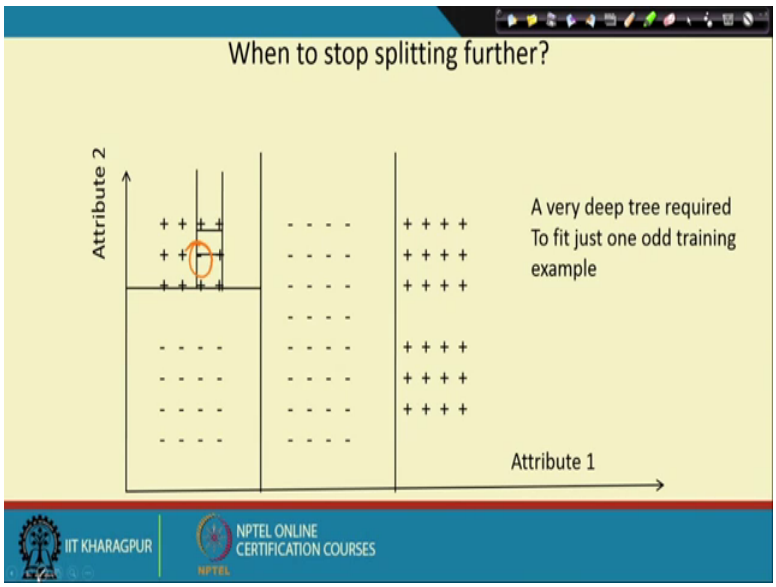
So, I split on humidity and I repeat the exercise for this and I get pure leaf nodes and this is what I get.

(Refer Slide Time: 03:35)



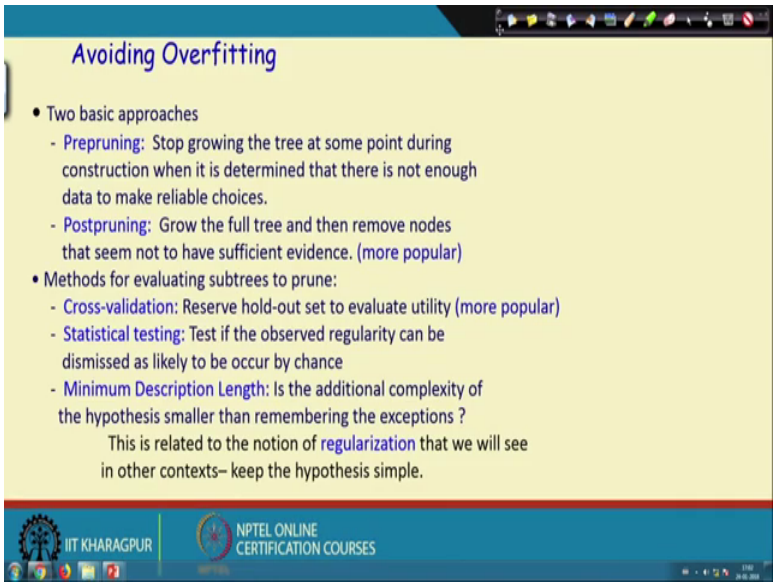
Now, one problem that happens in this kind of classification algorithm is you get, over fitting. So, if you make the training set quite small and if there is some noise in the training set if you get a deeper and deeper tree, your training set accuracy will keep on increasing, but test set accuracy may decrease after some time. So, x axis is height of the tree.

(Refer Slide Time: 04:16)



I explain this suppose there is a noise this point is a noise. If I keep on splitting it if I keep on splitting it to fit that, I have to draw few more lines you correctly get a pure leaf and I will have a deeper tree.

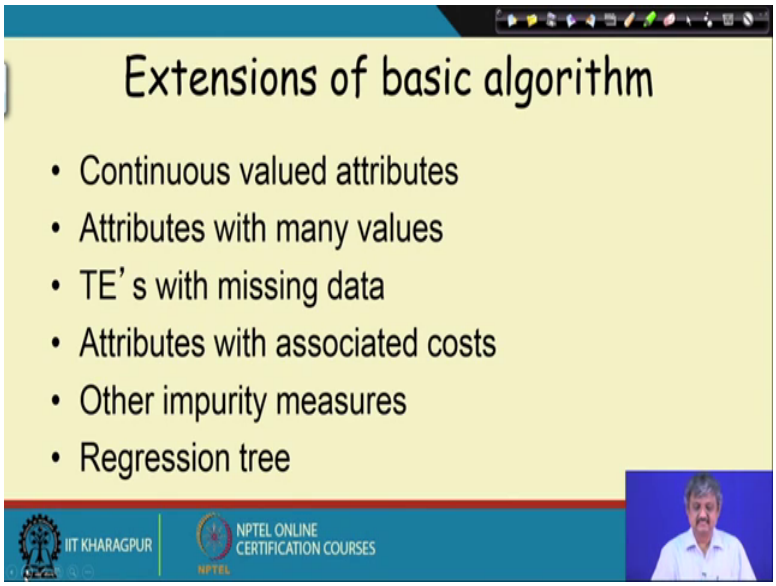
(Refer Slide Time: 04:50)



So, to solve this problem what one does is that you build a tree, then sometimes instead of pure node 100 percent pure node maybe 95 percent pure node is enough like maybe this much impurity you can allow and not split further. So, this means I am clipping the tree and not building it further this process is known as pruning. There are 2 approaches pre pruning while growing that tree you stop or post pruning you grow fully, then cut some branches which branches to cut you do something called a cross validation; that means, you take the training set error test consider sorry the test set error.

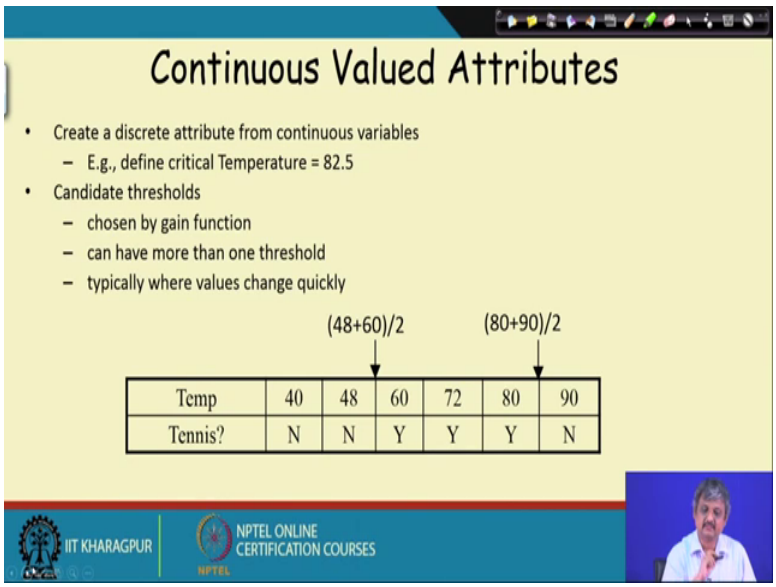
And see how much error you are getting. If it is high you stop do not you prune that, but do not build it further grow that branch. Further otherwise there is something called a description length; that means, description length of a tree is the height of the tree plus the number of examples it misclassifies, I want to build a tree which has the minimum description length ok.

(Refer Slide Time: 06:41)



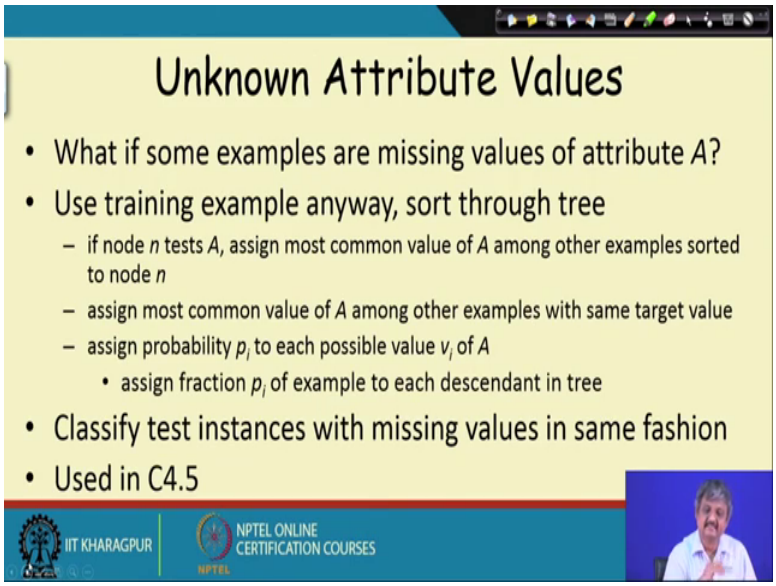
So, this is about tree pruning I will few extensions of the basic tree algorithm what about continuous values.

(Refer Slide Time: 06:53)



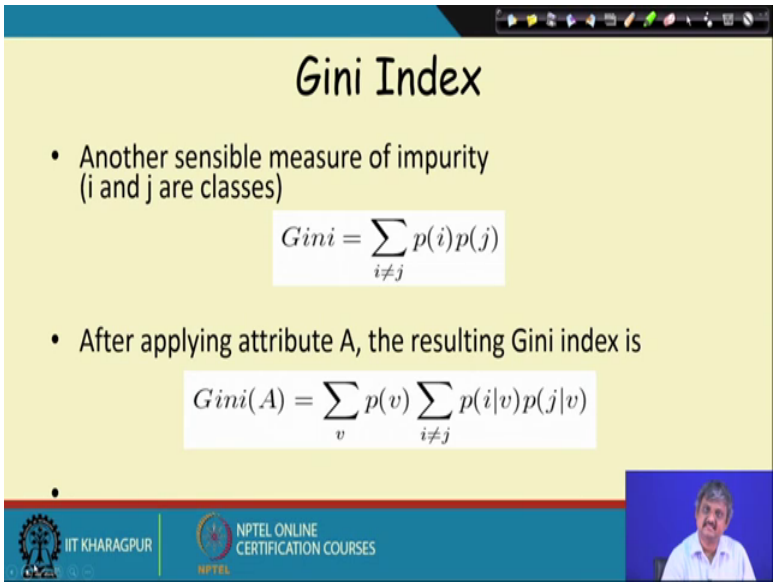
Often what we do you first discretize the continuous value into intervals by some method; some method you discretize, and then consider design that is the most common.

(Refer Slide Time: 07:18)



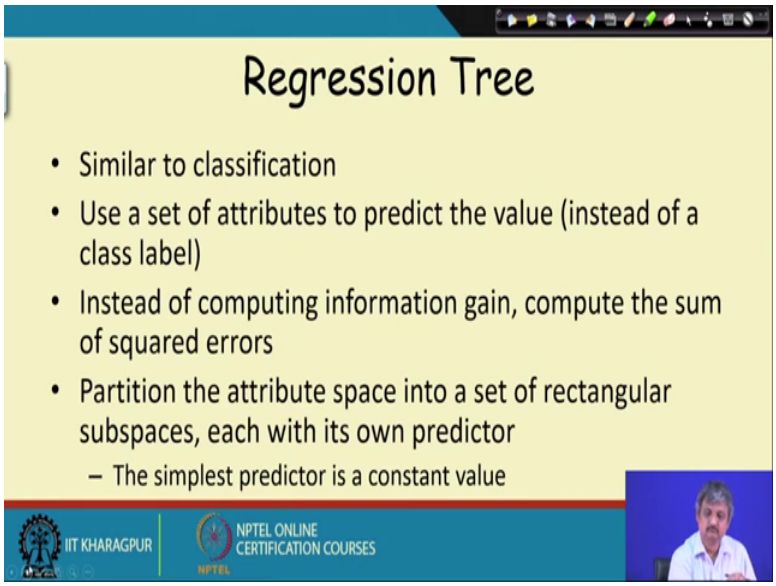
I will let us discuss some properties of discretize some algorithms for discretization. Similarly missing attributes you take often these are the strategies. So, C4.5 is popular software for decision tree which uses all these strategy.

(Refer Slide Time: 07:41)



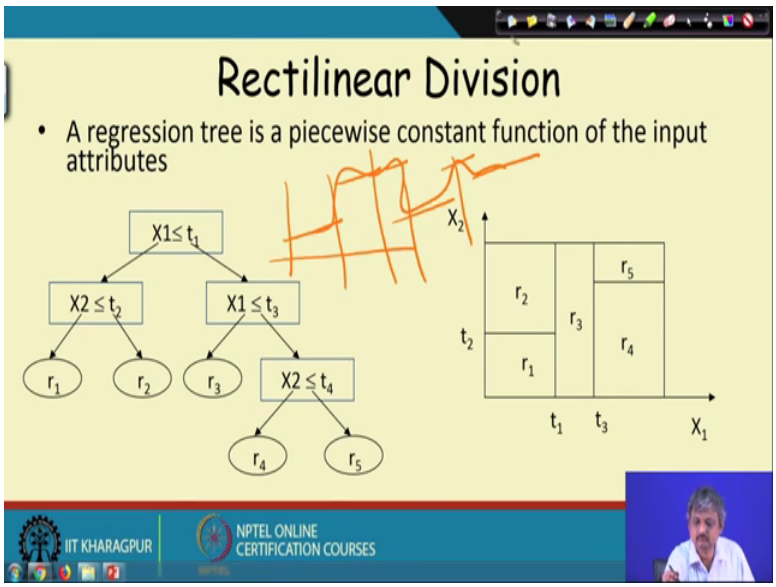
Sometimes instead of the entropy an alternate index is used called the Gini index which is defined like this.

(Refer Slide Time: 07:54)



One more extension is a regression tree, what we do here is that instead of leafs being classes each of leafs are some regression line.

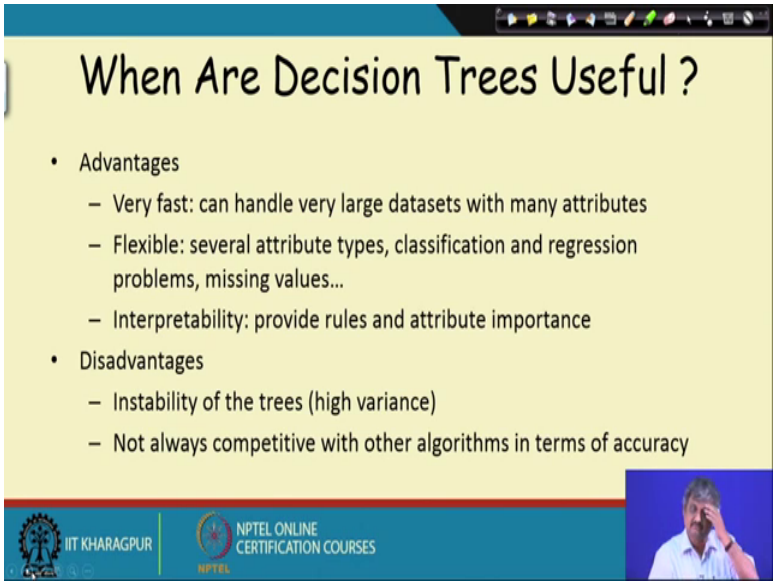
(Refer Slide Time: 08:07)



Some constant value or some constant piecewise, constant function some values or 1 or 2 instead of class their values. So, how do I use this? So suppose I want to fit a regression to this kind of a function I will piecewise fit constant values, and these boundaries will be decided by the decision tree. So, this algorithm is a popular algorithm called the cart

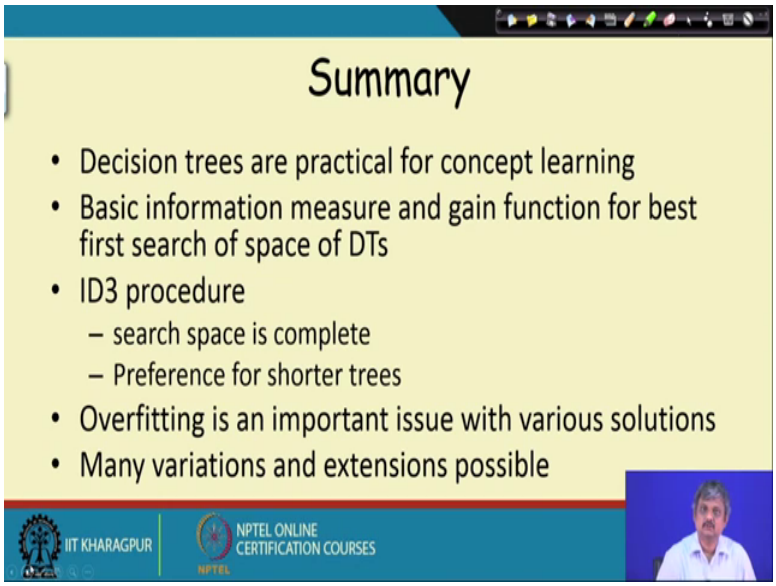
algorithm classification and regression tree, when I covered regression later I will cover more details of this.

(Refer Slide Time: 09:11)



These are some of the advantages of decision tree they are very fast flexible interpretable you can find rules. Disadvantage it depends very much on the training set you use some and also sometimes get less accuracy to the algorithms. So, we will consider in our next lectures here.

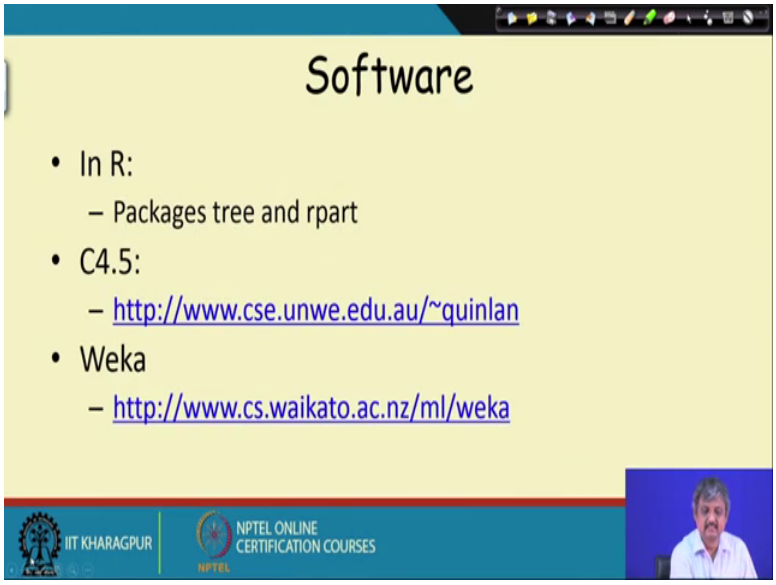
(Refer Slide Time: 09:36)



Is the summary of the topic of decision tree note that again, ID3 is all this method I discussed is they are usually clubbed together in an algorithm called ID3, which is a popular decision tree algorithm, so it is a method of learning classes classification uses information gain it prefer shorter tree over fitting is an important issue, and people have used various variation extensions to different cases.

So, if we are face to it and classification problem with say discrete attribute or few values. The first thing to try is a decision tree if it does not give good accuracy you go for other algorithms and decision is implemented in most of the software's.

(Refer Slide Time: 10:35)



So, with this these are some of the softwares, I which you can use to implement a decision tree C4.5 is a popular Weka is a machine learning toolbox which you can use and R has all the decision tree algorithms with this, I close my discussion on decision tree. In our next lecture we will go to other algorithms by classification.

Thank you.