Machine Learning of Soil and Crop Management
Professor Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology, Kharagpur
Lecture 18
Applications of Classification and Clustering Methods in Agriculture (contd.)

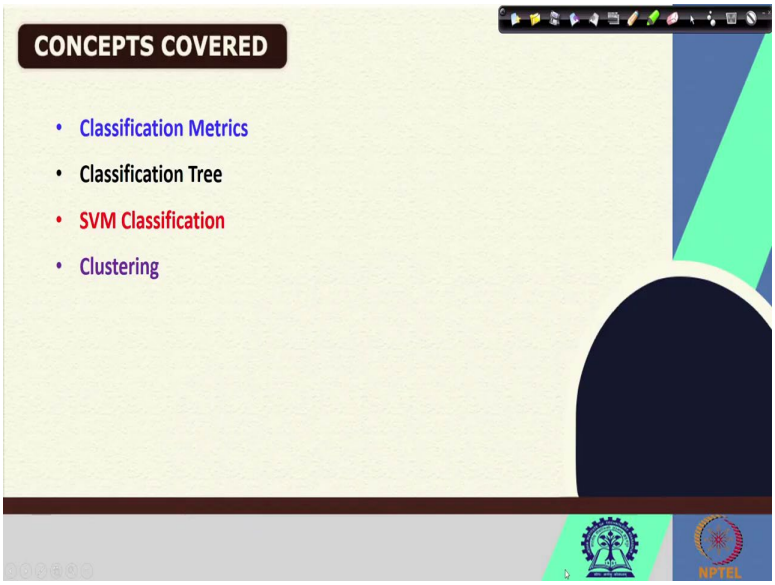
Welcome friends to this lecture number 18 of NPTEL online certification course of Machine Learning for Soil and Crop Management, and in this is week 4, and in this week 4 we are discussing the Application of Classification and Clustering Methods in Agriculture. So, in previous 2 lectures, lecture number 16 and lecture number 17, we have discussed what is classification and what is clustering?

What is the difference between classification methods and clustering methods? Remember, classification is a supervised method whereas clustering is an unsupervised method and we have seen some good examples of classification methods. We have seen the seen the linear classification methods like linear discriminant analysis.

And also we have seen the logistic regression, also, we have seen the another classification method that is K nearest neighbor classification and also we have seen very briefly the performance metrics or confusion metrix and different types of performance metrics calculated based on those based on that perform confusion metrix.

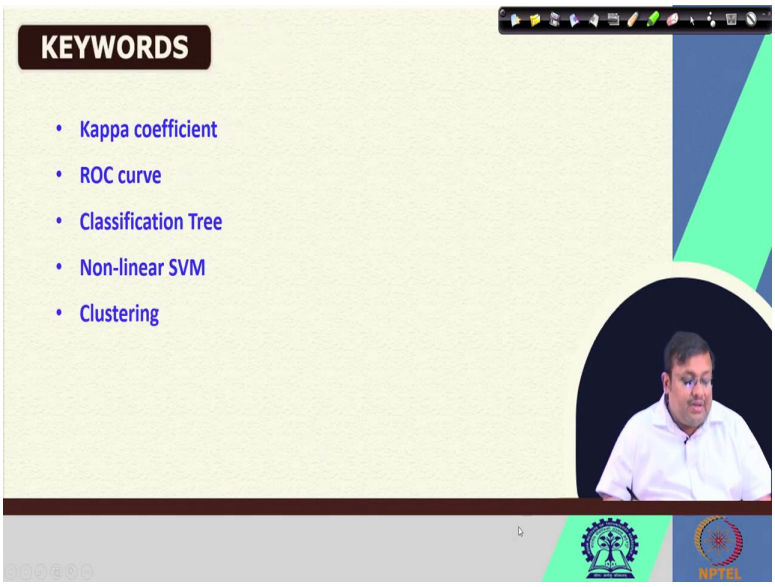
Now, these performance metrics there are different types of performance metrics like ROC recall Cohen's kappa, we have discussed very briefly, since, we have a very limited time. Now, today in this lecture, we are going to discuss these terms in details, so, that we can have a clearer idea about these terms. And we are going to see an example based on that example, we are going to discuss these performance metrics of any classification algorithm. Also, we are going to discuss some other classification and clustering methods.

(Refer Slide Time: 2:38)



So, these are the concepts which we are going to cover in this lecture, we are going to first discuss different classification metrics in details and also we are going to discuss the classification tree and also is SVM classification and also clustering, we are going to start the discussion on clustering and what are the basic concept of, what is the basic concept of clustering and how clustering is generally done, we are going to discuss in this lecture.

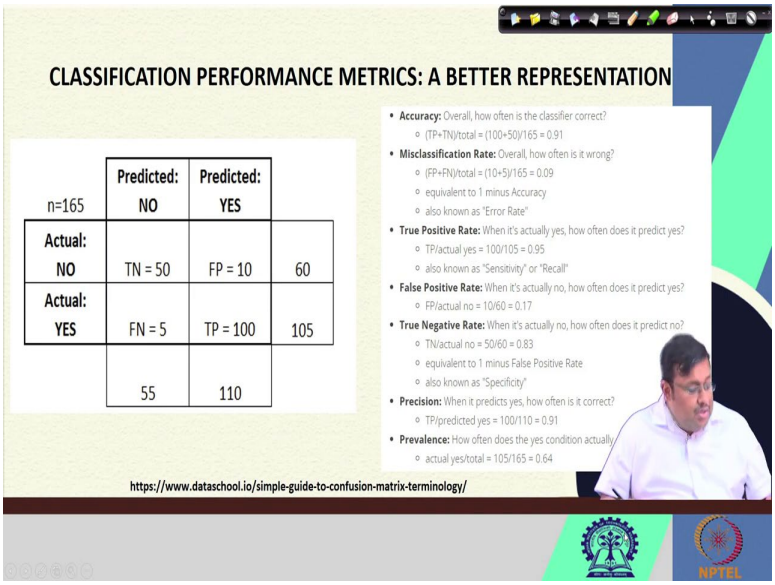
(Refer Slide Time: 3:16)



And these are the major keywords, which we are going to discuss in this lecture. First of all, we are going to discuss the Kappa coefficient, then ROC curve, then we are going to discuss

about the classification tree, then nonlinear SVM, linear SVM and also we are going to talk about the clustering. So, these are the important keywords for this lecture.

(Refer Slide Time: 3:41)



So, let us start with a discussion of classification performance metrics and I am going to use this example which I have found in this source. And we have also seen this confusion matrix in our previous lecture. So, we are going to discuss this in detail suppose, there are total 165 people who are being classified based on whether they have contacted any disease or not or whether we have predicted their disease occurrence, rightly or wrongly.

So, we can see that these 165 observations are divided into 4 categories and we call them true negative, true positive, false positive and false negative. So, if we see here this is the predicted when there is a predicted no that is there is no disease and here there is a prediction when there is the confirmation of the disease, positive test.

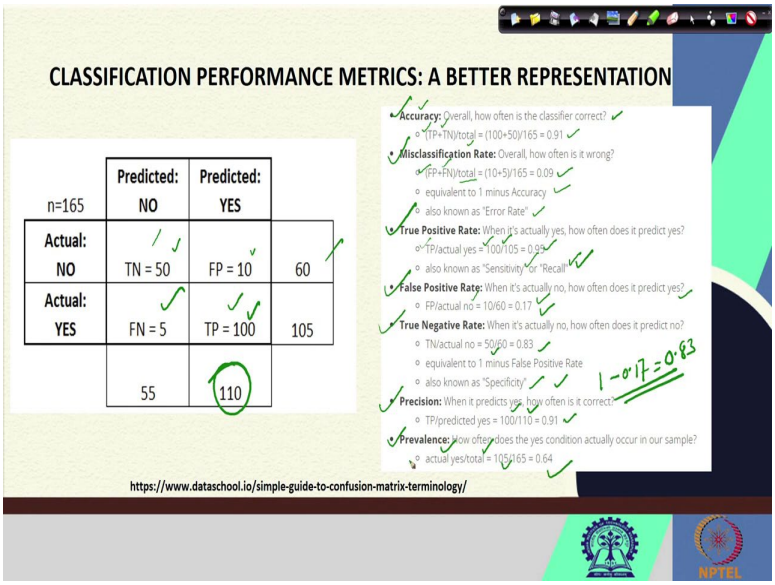
So, here this is an actual and this is actual no and this is actually yes. So, this is the predicted values, these are the actual values and we can see the total 165 observations are divided into these 4 categories. So, if we see the predicted when there is no disease, when actually there was no disease and predicted there was no disease also, so we can see this is called true negative and the 50 samples are going to this true negative class.

And when this when there was no disease, but our test was this predicted that there is a disease that is called false positive and there are 10 number of samples in the false positive category. Also, you can see when the actual with actually the subjects have disease

occurrence, but we have predicted that there is no disease. So, this is called false negative and the true positive that means, actually there is disease and our test also correctly predicted that disease occurrence. So, this is called true positive.

So, we can see true positive false positive, false negative true negative. So, this is how this whole 165 samples are divided into 4 different categories, and if we sum up these columns 55 and here 110 and here, summing up the rows, you can see here all the elements in a single row, so 50, 110, 60 and 500 there is 105.

(Refer Slide Time: 6:47)



So, the performance metrics which we are going to discuss here are mentioned here, you already know this the accuracy term. Accuracy says overall how many often is the classifier correct. So, of course, the classifier is correct, when we are getting the true positive and true negative values. So, true positive plus to negative divided by the total number of observations. So, we can here we can see here that is 0.91.

So, it is 91 percent accuracy, you can see, what is the misclassification rate? Misclassification rate that means overall how often it is wrong, so of course, it will be just opposite. So, false positive plus false negative divided by the total number of samples. So, you can see 10 plus 5 by 165 that is 0.09. So, 9 percent misclassification and 91 percent classification accuracy, so basically these misclassification rate generally is also equivalent to 1 minus the accuracy of course, 1 minus 0.91 equal 2.09. So, also we call it as an error rate.

So, now, it is clear that what is accuracy and what is the misclassification rate. Now, the next important point is the true positive rate. So, when it is actually yes, how often does it predicted yes. So, actually yes, so, true positive was 100 and actually predicted yes, if you see if you just sum up these elements in this row of the you know of yes, so, you can see 5 plus 100 that is 105. So, that means, that it is predicted yes for total 100 plus 5, 105 observations.

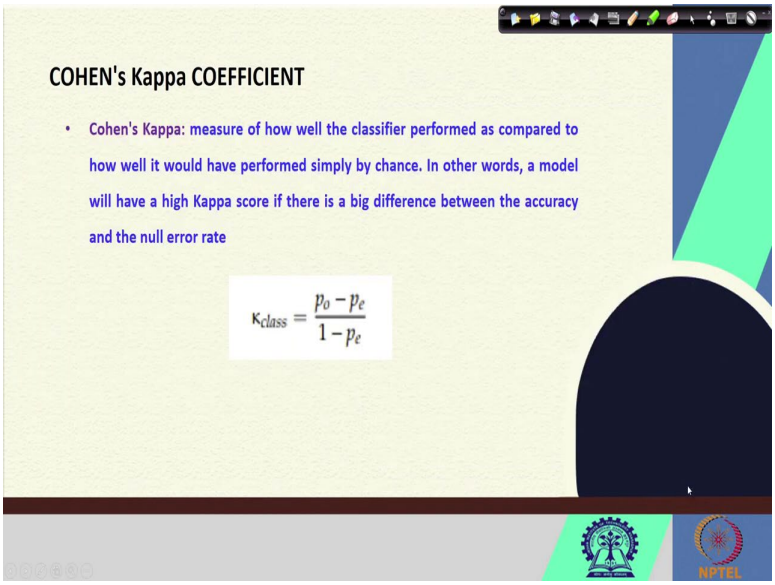
So, the true positive rate is 0.95 also known as the sensitivity or recall. So, this is called sensitivity or recall. False positive rate is when it is actually know how often it does it predicted yes. So, it is actually know in case of 10 observations and actual know was total 10 plus 50 that is 60, 0.17. Also, the true negative rate is when it is actually know, how often does it predicted?

So, here basically through negative by the actual number of negative, so here we know the true negative is 50 and actual negatively predicted actual negative was 50 plus 10 equal to 60. So, we can get 10 plus by 60 that is 10 by, sorry, 50 by 60 equal to 0.83, so, it is also equivalent to 1 minus false positive rate. So, false positive rate it is 0.17. So, 1 minus 0.17 stands for 0.83.

So, this is the true negative rate. Also this is equivalent to or also known as synonymous less specificity. So, we now know that the call this is the specificity, what is precision? When it predicts yes how often it is correct? So, the correct is true positive that is 100 and the total number of predicted yes is 100 plus 10 equal to 110. So, 100 plus 100 by 110 equals 2.91, so 91 percent precision.

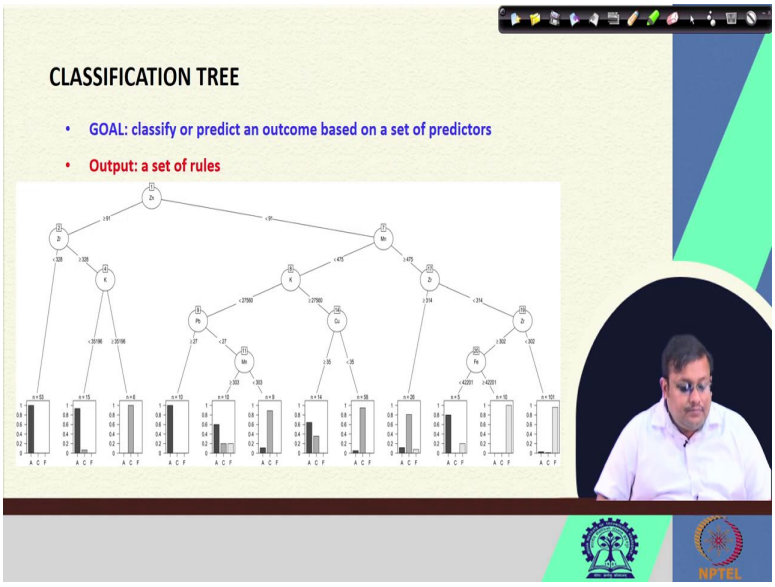
And prevalence, how often does the yes condition actually occur in our sample, so, actually yes verse by total, so, actually yes is 100 plus 5, 105 divided by 165 so, it is 0.64. So, this is how we calculate these different performance metrics like accuracy, misclassification, true positive rate, false positive rate, true negative rate, precision and also prevalence. So, now, I hope the things are much clearer to all of you.

(Refer Slide Time: 11:24)



Now, another important is performance metrics is Cohen's Kappa, this Cohen's kappa basically measures how well the classifiers perform to as compared to how well it will it would have performed simply by chance. So, in other words, a model will have higher kappa score if there is a big difference between the accuracy and the null error rate. So, here, this kappa coefficient can be calculated by this formula and we have already discussed this formula in our previous lecture.

(Refer Slide Time: 12:02)



So, we are going to first all we know today, we are also going to discuss the classification tree now, you all know what is classification tree? From our previous week discussion, we

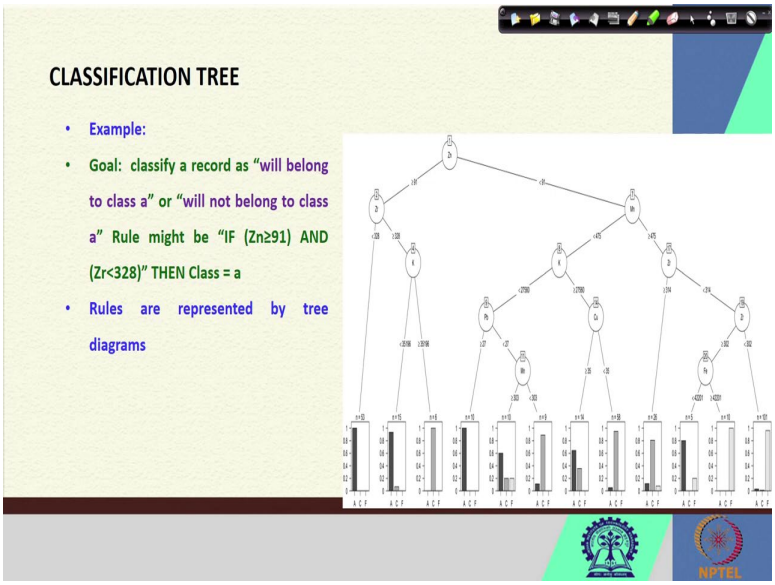
know the classification and regression tree when our target is a categorical variable, then we call it a classification tree and when our target is a numerical continuous variable, then we call it a regression tree sometime classification tree is also known as decision tree with an LE decide the class of a sample based on certain rules and ultimately using the recursive partitioning of the data we assign the test sample into one of the target groups.

So, as we have seen the same example, in our previous lecture also in the week 3. So, this is that, this is the same classification of land use type using the elemental content measured by some handheld sensor called portable XRF or extra fluorescent sensor and you can see that these are the spitting nodes and these are the terminal nodes.

And in this terminal nodes, the distribution of the samples in each of these 3 classes like A, C, F. A stands for agriculture, C stands for converted lands and F stands for forest land and their relative distribution is given here. So, if it is sample is having the zinc content of greater than 0.91 unit and less than but 3 to 8 unit of zircon, zirconium then we can assign them into this terminal node.

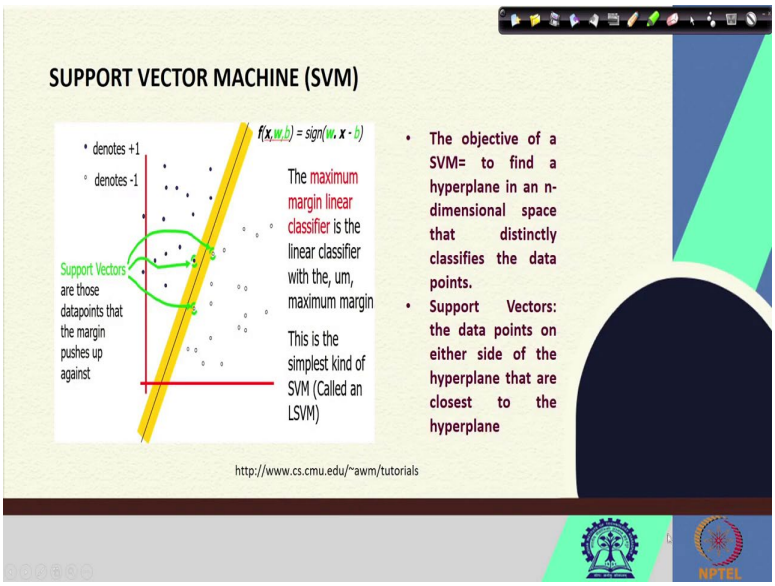
And also if the zinc content is greater than 91 and zeta content is greater than 328 and potassium content is less than 35196, this 3.196 is ppm value and ultimately, we can classify them into this terminal node. So, this is how we can differentiate any sample based on these rules to one of these terminal nodes. So, this is called the classification tree. We have seen this already in our previous week.

(Refer Slide Time: 15:14)



And our goal is to classify if any unknown sample will belong to class a or will not belong to class a, it is just an example. So, here the rule is Zn continued, Zn is greater than 0.91 and Zr is less than 3 to 8 then the class is a. So, rules are represented by these tree diagrams.

(Refer Slide Time: 15:46)

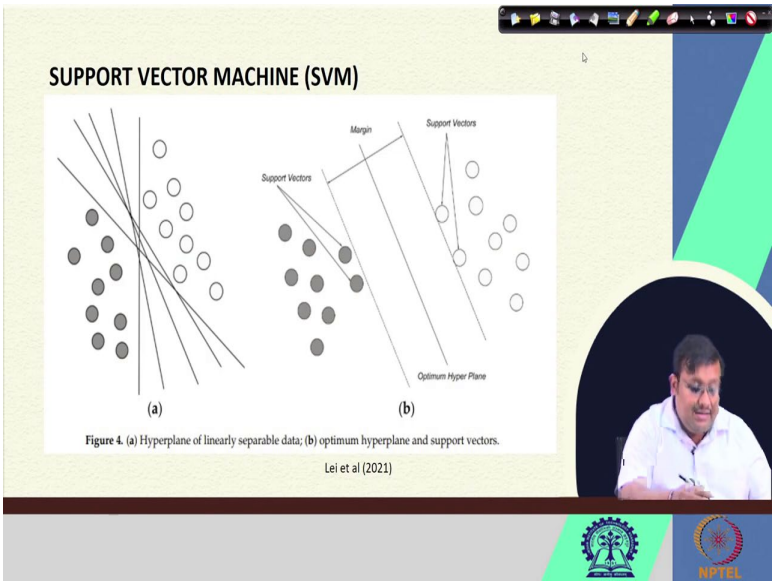


So, once we have completed this classification to discussion, let us see the support vector machine this is also we have discussed in details in our previous week. Now, we know that the objective of SVM is to find a hyperplane. So, here this is a hyperplane, this is a hyperplane and so, the whole objective of the SVM is to find the hyperplane in an n dimensional space that distinctly classifies the data points.

So, here there are 2 types of data belongs to 2 different classes and this hyperplane is differentiating the samples into the 2 classes and also the, what are the support vectors? So, the data points on either side of the hyperplane that are closest to the hyperplane are known as the support vectors. So, these points are known as the support vector which are either side of the hyperplane and closest to the hyperplane.

So, basically the idea behind the support vector machine is to define the maximum margin linear classifier. So, which can show the maximum margin to separate these 2 classes, so this is the simplest kind of SVM also known as linear SVM and we are also going to discuss the non-linear SVM. So, this is the simple representation simplest kind of support vector machine representation, this is the linear support vector machine regression.

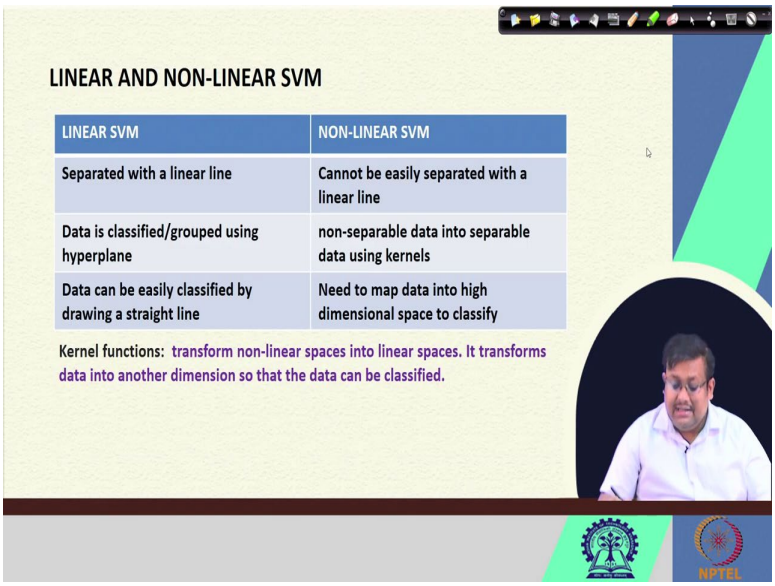
(Refer Slide Time: 17:44)



And so, again some examples are given here you can see that hyperplane have linearly separable data. So, there are 2 classes of data you can see and we can draw different hyperplanes to linearly separate the data. However, this is the optimum hyperplane because it is reducing the maximum margin from the closest point of 2 classes.

So, these are the support vectors this is support vector these are the support vectors also. So, this is an optimal hyperplane because these are maintaining the maximum margin distance from this in this hyperplane. So, this is how it is called the linear support vector machine algorithm.

(Refer Slide Time: 18:32)

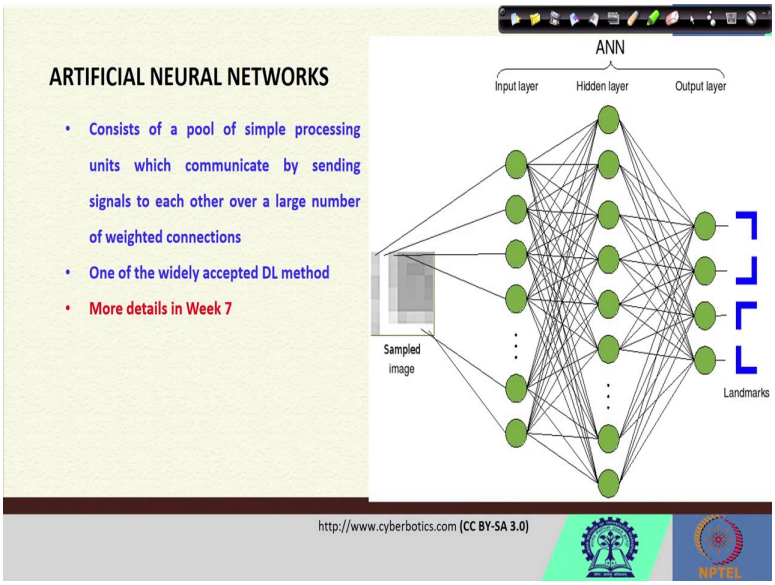


Now, the difference between linear SVM and non-linear SVM, Linear SVM is when we separate with a linear line that is called a linear SVM we have seen that and in case of non-linear SVM it the data set cannot be separated easily with a linear line. So, also in case of linear SVM the data is classified grouped using hyperplane.

However, in case of non-linear SVM the data is classified based on the kernels and we use a tree called kernel trick, what is that we will see later. And then the data can in case of linear SVM, data can be easily classified by drawing a straight line, whereas, in case of nonlinear SVM it need to map the data into high dimensional space to classify the samples.

So, what is the kernel function? Kernel function is basically a mathematical operation which transform the data into the transforms that nonlinear space into linear space. So, if the data is nonlinear and any want to transform them into linear space, you have to use the kernel function. So, it transformed data into another dimension, so that data can be classified.

(Refer Slide Time: 20:09)

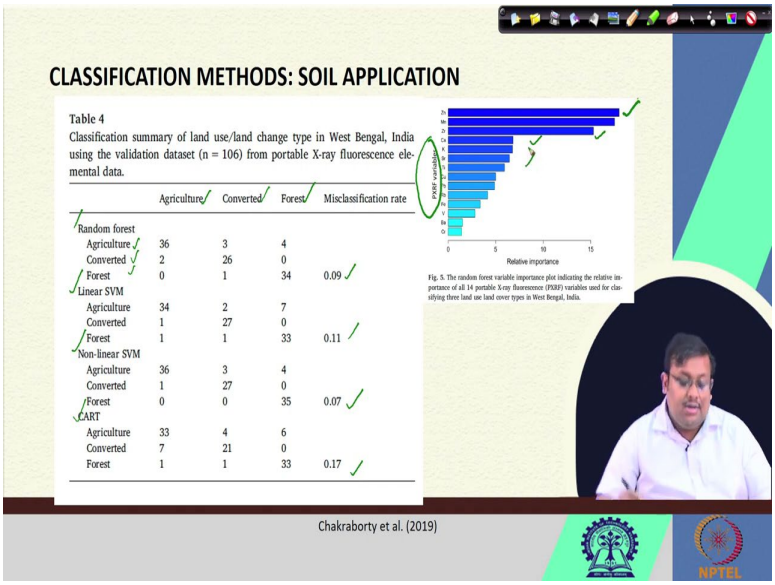


Now, we also going to discuss what is artificial neural network. Now, this is one of the world's most widely accepted deep learning method and it is it consists of a pool of simple processing units which communicate by sending signals to each other over a large number of weighted connections. And so, here you can see that it has the input layers, where we input where the data is getting incorporated and there are multiple hidden layers and these hidden layers are assigned different weights and then there is an output layer.

So, the whole representation of input layer, hidden layer and output layer resembles the nerve system of human body where the neurons which are the main conductor of nerve impulse are conducting the information from one part of the body to another part of the body. So, similar type of representations are there in case of neural network and that is why it is called artificial neural network, we are going to discuss this algorithm in details in week 7 when we are going to discuss the image processing using deep learning methods.

So, I am not going to discuss in detail about these artificial neural network in this lecture, we are going to see these in detail in our upcoming week 7 lectures, but remember that these artificial neural network can be utilized both in regression as well as classification problem and it is widely used method for a for data classification.

(Refer Slide Time: 22:09)



So, let us see some example of classification methods we have seen that in this example this shows the classification summary of land use, land change type if West Bengal. Also actually in this research, 2 years back, 2 to 3 years back, so, what we did we classify samples coming from 3 different land use type like agriculture converted land and forested land using different types of classification algorithm into 1 of 2, one of these 3 classes.

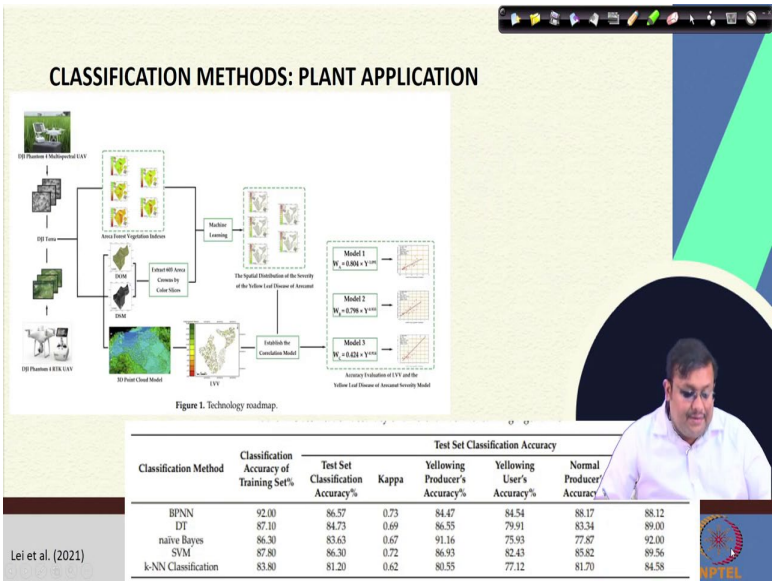
So, you can see here agriculture converted and forest and random forest. So, these are the actual and these are the random forest based predictions of agricultural samples belong to agricultural converted and forest areas. So, we have compared random forest we have compared linear SVM we have compared non-linear SVM and also cart or classification tree.

So, this is the confusion metrix and from this conservation metrix, we have calculated the misclassification rate. So, in case of random forests, we got the misclassification rate of 9 percent in case of linear SVM, we got the classification misclassification rate of 11 percent. In case of non-linear SVM, we got the club misclassification rate of 7 percent and in case of a classification tree, we got the misclassification rate of 17 percent.

So, this is and then so, from there we can see that both random forest and also nonlinear SVM gave very good very low misclassification rate. In other words, they are highly accurate. Not only using the random forest, random forest we have discussed in details, but remember, the random forest can be used both in classification as well as in regression. Now, in case of random forests classification and regression, we can select the variables based on their relative importance and we can plot them.

So, here you can see we have plotted the variable importance of random forest input variables here, we have incorporated the elements like zinc, manganese, and we can see the highest contribution we are getting from zinc followed by manganese then Zr then Ca then K then CSR and so on so forth. So, we can arrange them we can solve them.

(Refer Slide Time: 25:02)

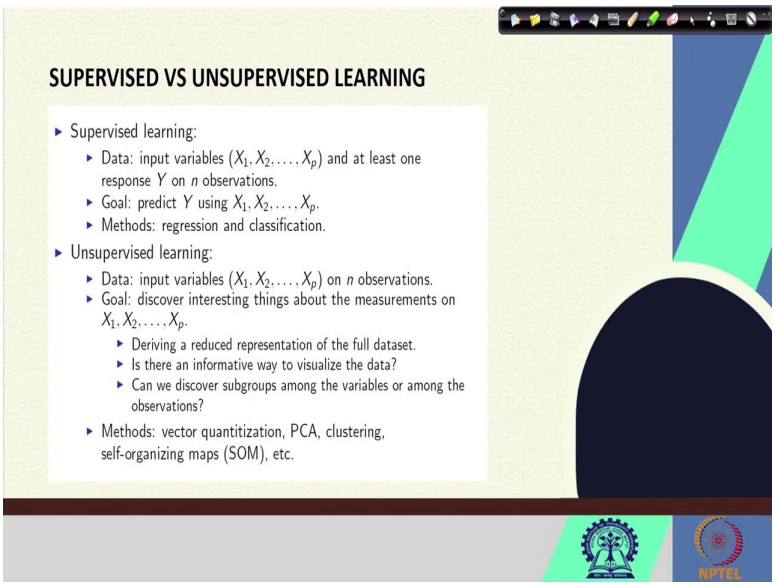


Classification Method	Classification Accuracy of Training Set%	Test Set Classification Accuracy					
		Test Set Classification Accuracy%	Kappa	Yellowing Producer's Accuracy%	Yellowing User's Accuracy%	Normal Producer's Accuracy	
BPNN	92.00	86.57	0.73	84.47	84.54	88.17	88.12
DT	87.10	84.73	0.69	86.55	79.91	83.34	89.00
naïve Bayes	86.30	83.63	0.67	91.16	75.93	77.87	92.00
SVM	87.80	86.30	0.72	86.93	82.43	85.82	89.56
k-NN Classification	83.80	81.20	0.62	80.55	77.12	81.70	84.58

Another application we can see here it is a application of classification method for plant. So, here, you can see that in this example, Lel at al in 2021 they have used the UV images for collecting the multispectral data and high resolution, remote sensing data and using that data or images, they have developed some kind of vegetation indices for vegetation indices for classifying the severity of the yellow leaf disease of arecanut in China.

And they have used different types of machine learning classification algorithm like BPNN and then back propagation neural network and then decision tree or in other words, there is a classification tree then knave bias modeling then support vector machine classification and k nearest neighbor classification and they have calculated the classification accuracy of the training set and test set classification accuracy is also calculated based on the performance metrics like classification accuracy Kappa coefficient and so on So, forth.

(Refer Slide Time: 26:31)



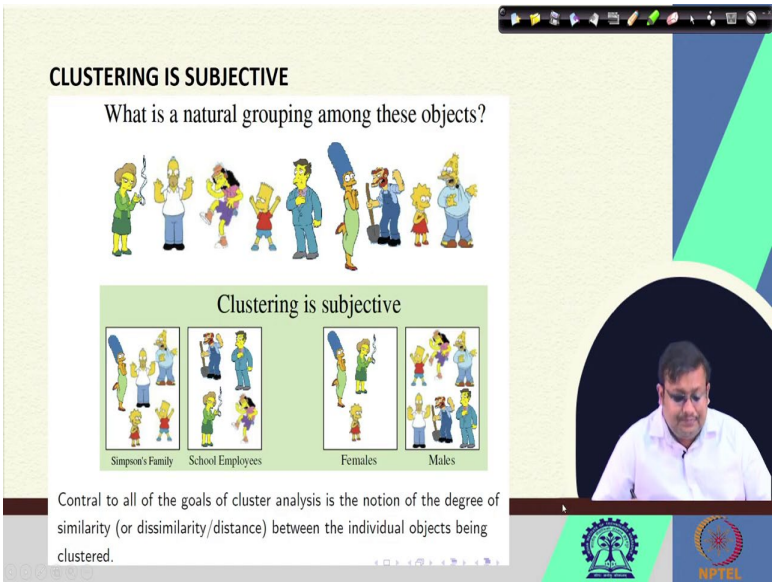
So, we know that the major difference between supervised and unsupervised learning is, in case of supervised learning, we have both the input variables and at least one response Y on the inner observations and in case of supervised learning, our predict is to our goal is to predict the Y based on these X1, X2 up to Xp variable a feature space and the methods which we generally use are either regression or classification.

Whereas, in case of unsupervised method, the input variables lie from X1 and X2 that is total n observations are there and our goal is to discover the interesting thing about them of the measurement of X1 to Xp that means, without the labeling of the data without the outcome of the without the depend sorry, dependent variable, what is the interesting trend among the feature space so, that we calculate in case of unsupervised learning.

Now, ultimately, in case of unsupervised learning, we derive a reduced representation of the full data set and we ask ourselves is there any information or informative way to visualize the data? And the third question we have in our mind in case of unsupervised learning is can we discover subgroups among the variables or among the observations?

So, these 3 we generally try to answer using unsupervised learning and the methods which we use are vector quantization, and then PCA, PCA is unsupervised method you know that then clustering themselves organization maps etc. We have already discussed PCA in our previous lecture. So, I am not going to discuss PCA anymore, we are going to discuss the clustering.

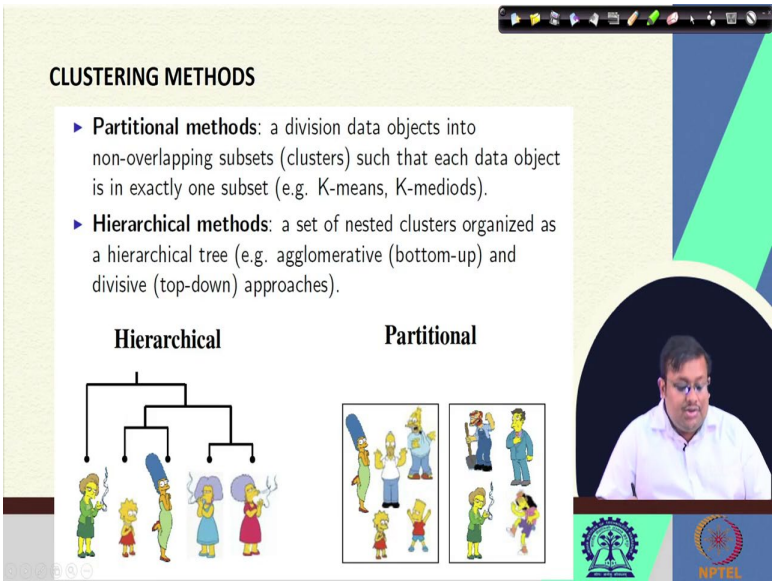
(Refer Slide Time: 28:31)



So, remember clustering is always subjective, because what is the natural grouping among these subjects, we have these a collection of different people. So, we can do the clustering from different perspective, we can either class these based on the Simpson family, we can class them into 2 categories like Simpsons family and school employees or we can classify them based on females and males.

So, central to all of these goals is to cluster analysis is the notion of degree of similarity or dissimilarity between the individual object being clustered. Now, what are the similarity and dissimilarity measures we are going to learn in our next lecture?

(Refer Slide Time: 29:16)



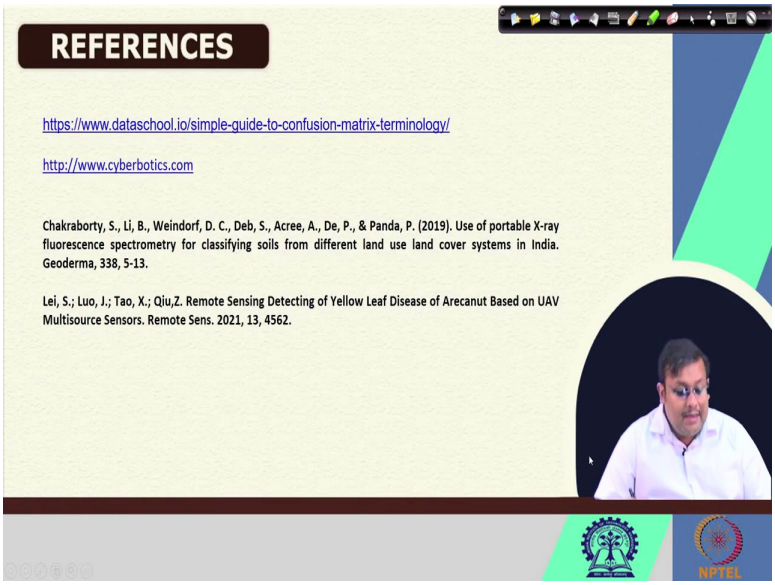
So, what are the different types of clustering methods? One is called the one major category is called partitioning methods and other is hierarchical methods. So, in case a partitioning method a division of the data objects into non-overlapping subsets or clusters such that each data object is not exactly one subset.

So, basically we divide the data into the non-ever overlapping subsets or clusters and some of the organ some of the methods are K-means clustering, K-means clustering and also the second category is hierarchical methods. In the case of hierarchical, so, this is an example of partitioning method. So, we are classifying the objects into non-overlapping subsets. So, they are not overlapping.

So, they are individual clusters and their method is hierarchical methods, which could be either bottom up or bottom or top down approach, the bottom upper approach is known as agglomerative approach and the top down approach is known as divisive approach. So, here basically, we try to in this hierarchical method is basically a set of nested cluster organizes a hierarchical tree.

So, you can see, hierarchically, we try to divide the objects into multiple classes based on some similarity and dissimilarity measures. So, these are 2 clustering methods, which are generally broad clustering categories. And we are going to discuss these K-means and K-meteoroids and also these agglomerative hierarchical clustering in our upcoming lectures.

(Refer Slide Time: 31:00)



So, I hope, I have covered on I have given you some important information regarding classification and clustering and these are the references which I have used in this lecture. And let us meet in our next lecture to discuss the clustering metrix, the clustering and also some similarity, dissimilarity measures in details. So, let us meet in our next lecture to discuss those things. Thank you.