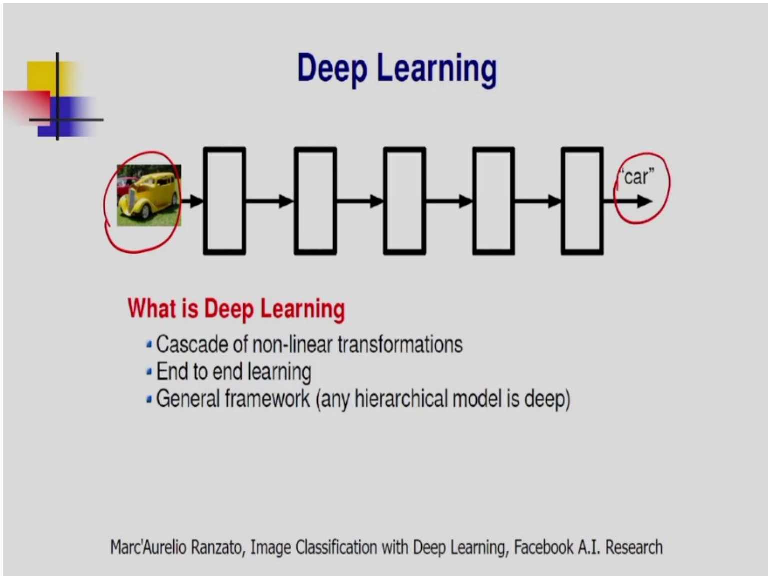
Computer Vision and Image Processing- Fundamentals and Applications
Professor Doctor M.K. Bhuyan
Department of Electronics and Electrical Engineering
Indian Institute of Technology, Guwahati India
Lecture 31
Introduction to Machine Learning

Welcome to NPTEL MOOCs course on Computer Vision and Image Processing- Fundamentals and Applications. In my last class, I discussed the concept of pattern classification and I briefly explained the concept of the supervised learning. So, today I am going to continue the same discussion, first I will explain the concept of deep learning, how it is different from the traditional pattern recognition system and after this, I will discuss the concept of supervised learning and unsupervised learning. So, what is deep learning?

(Refer Slide Time: 01:11)

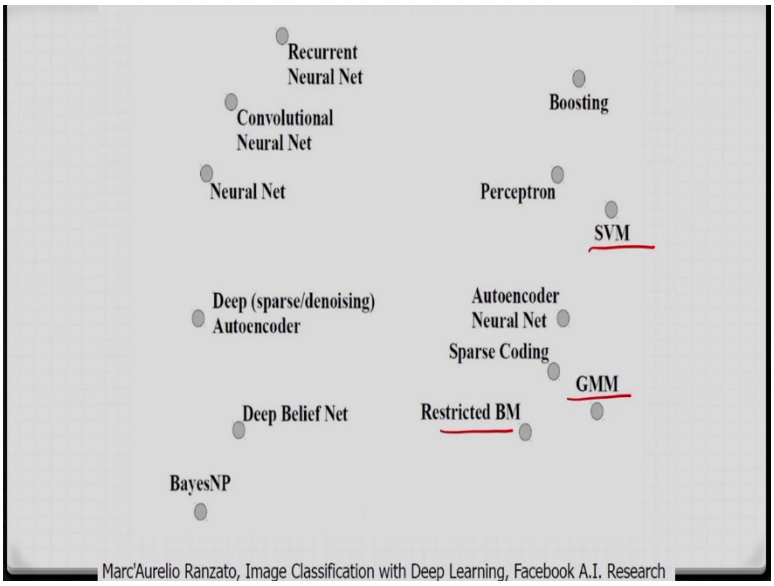


Here you can see in my next slide, now already I have discussed about the definition of the deep learning the deep learning is nothing but the subset of the machine learning. So, mainly it is the extension of the artificial neural network. So, in this figure you can see what I am considering, I am considering one image and I want to recognise whether it is a car or not.

So, in this case in case of a deep networks, it is nothing but the cascade of nonlinear transformations and in this case, I am considering the hierarchical model that means, I want to extract more and more information from the input signal, input image. So, by considering the cascade of nonlinear transformations and also, we are considering the end-to-end learning, we want to extract more and more information from the image, from the signal that means, the image has the low frequency information, the high frequency information, maybe some diagonal edges may be there.

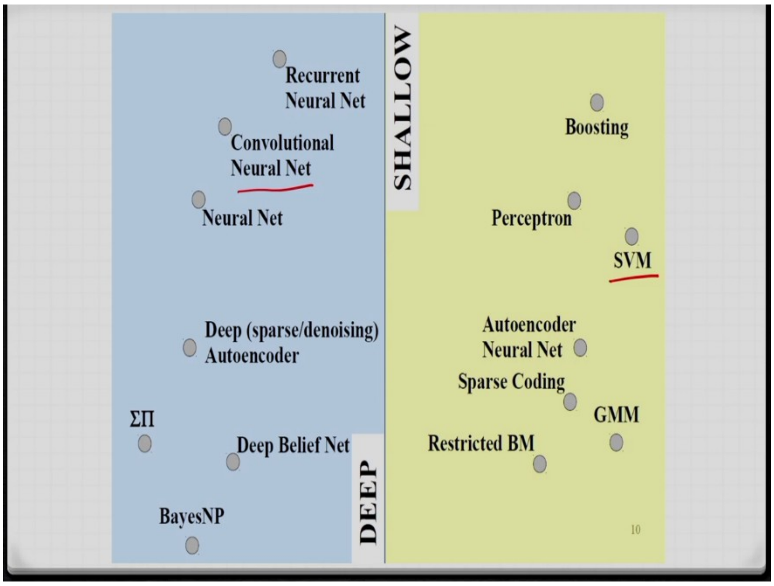
So, all this information I want to extract from the image, I will explain later on about the deep learning techniques, but mainly it is a cascade of nonlinear transformations and the end-to-end learning and we want to extract more and more information from the input image, from the input signal and based on this we can do the classification. So, the output is the car, that means it is identified a car is present in the image. The space of machine learning methods. So, I will show some popular machine learning methods in the next slides.

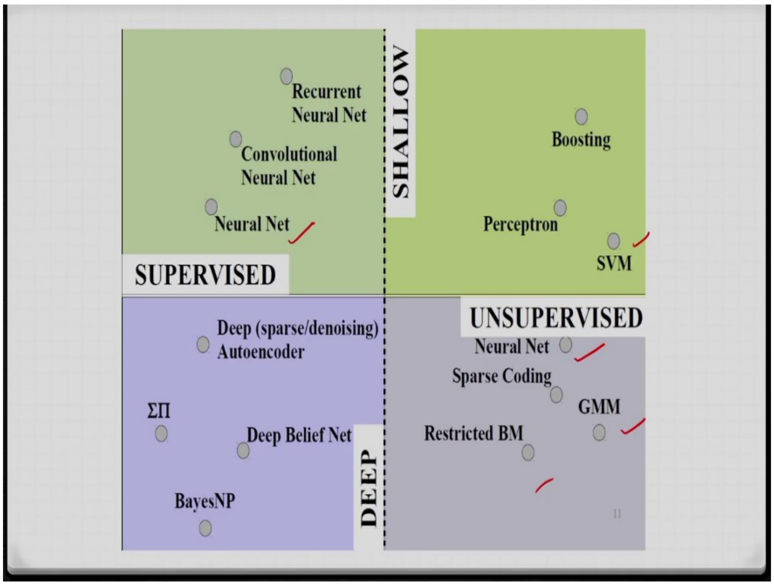
(Refer Slide Time: 02:42)

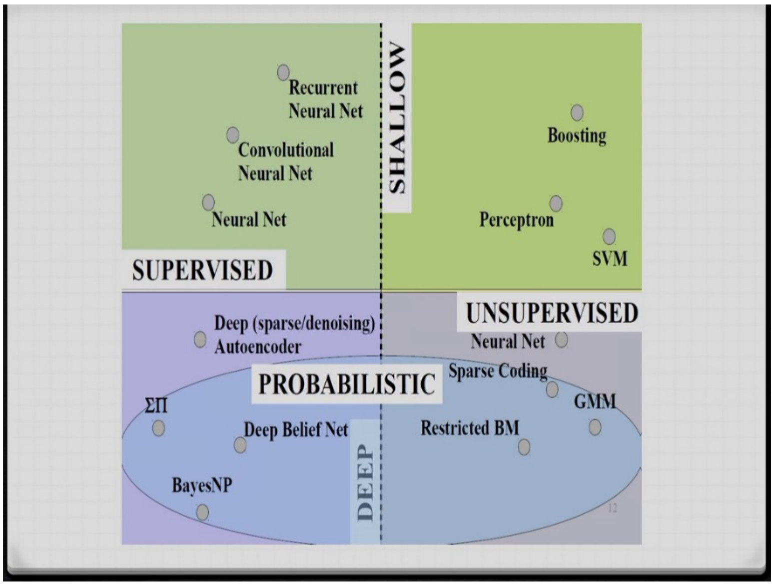


So, first you can see there are many algorithms like recurrent neural networks, convolution neural networks, neural networks, auto encoders, support vector machines, SVM is the Support Vector Machines, GMM is the Gaussian Mixture Model, restricted Boltzmann machine, sparse coding. So, perception. So, these are the techniques, the popular techniques.

(Refer Slide Time: 03:06)





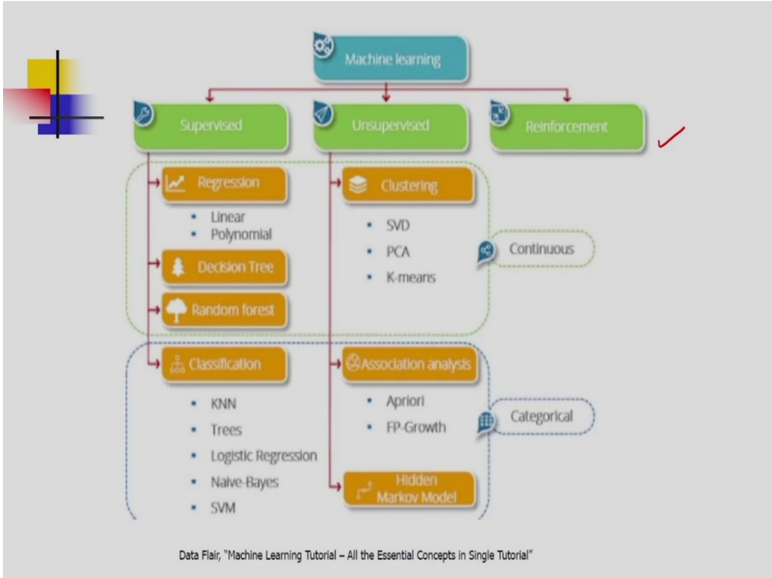


And you can see the left side is the deep structure, the right side is the shallow structure. So, if I consider support vector machine that is nothing but the shallow structure, but if I consider the convolution neural network, the recurrent neural networks that is the deep learning techniques and again I am classifying, you can see some techniques are supervised and some techniques are unsupervised.

So, if I consider, I can consider the supervised neural networks, also the unsupervised neural networks are also available, like the competitive neural networks or the unsupervised neural networks and you can see like Gaussian mixture model that is unsupervised restricted Boltzmann machine that is also unsupervised, but support vector machine you can see this side, that is the shallow learning, but if you consider the like, the convolution neural networks, recurrent neural networks, these are the deep learning techniques.

And some are the probabilistic methods you can see if I consider a Bayes NP, deep belief networks, restricted Boltzmann machines, GMM all these other probabilistic methods. So, here in this diagram, I have shown the popular machine learning algorithms.

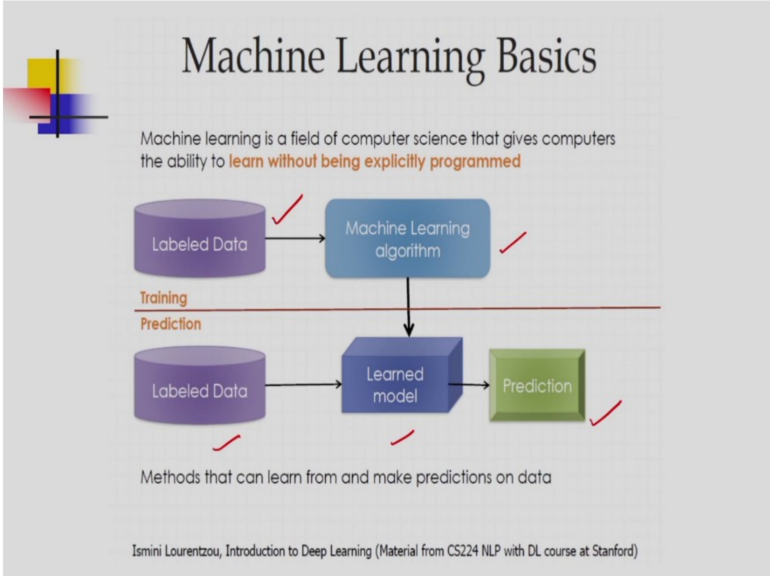
(Refer Slide Time: 04:24)



And in this case, you can see I am considering the machine learning techniques, some techniques are supervised you can see the machine learning books, because already I told you that it is not possible to discuss all the machine learning concepts in this computer vision course. So, that is why I am considering only the discussion, the brief discussion about these techniques.

So, one is the supervised technique. If you see maybe the regression, you can consider decision trees, random forests and classification like the K nearest neighbour technique, trees, logistic regression, naïve bayes, support vector machines. In case of the unsupervised clustering, I can consider the singular value decomposition that PCA, the Principal Component Analysis. The K means clustering like this we can consider. So also, we have the reinforcement learning that I will explain later on. So, these are some machine learning techniques.

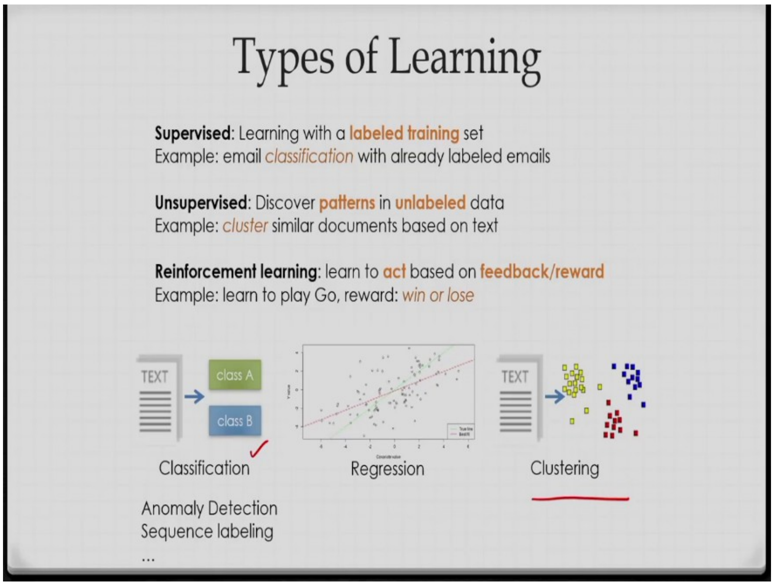
(Refer Slide Time: 05:21)



And in case of the machine learning techniques you can see, first I have to do the training and after the training, we have to do, go product testing. So, here you can see, I have the label data and we are considering the machine learning algorithms that is mainly for the training. So, in case of the supervised technique, already I have explained, we know that class labels and the corresponding training samples are available.

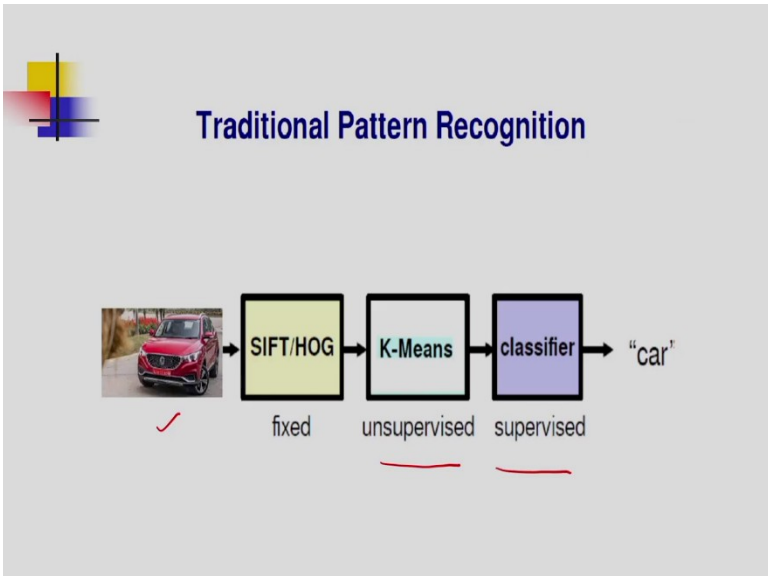
So, based on these training samples, what we can consider, we can do the training of the system, the training of the algorithm and after this, what we can do that suppose the level data is available and already we have the learn model. So, by using the learn model, we can do the prediction, we can do the decision making, this is about the machine learning basics.

(Refer Slide Time: 06:11)



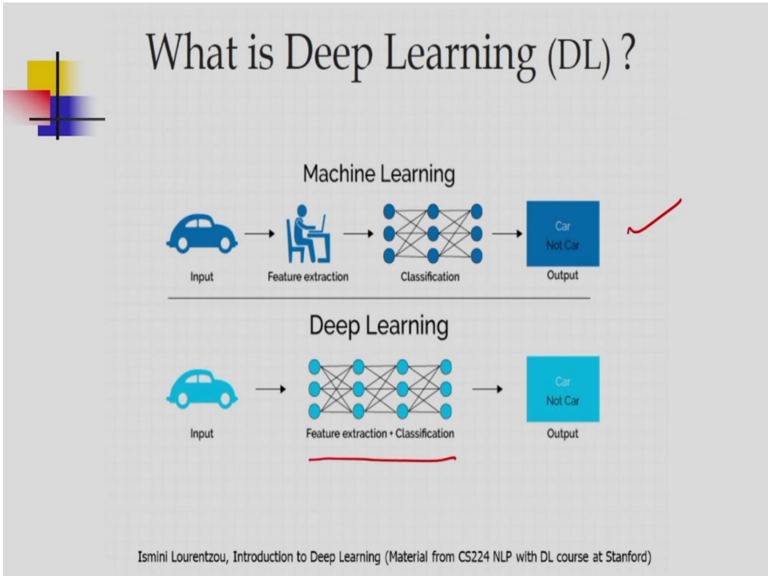
And here I have shown that concept of the supervised unsupervised and the reinforcement learning. The reinforcement learning, I will explain later on. So, here you can see, first one is the classification example I have shown. So, two classes I have shown the class A and class B and in case of regression you can see it is nothing but the fitting of a line between the sample points that concept also I will explain later on and what is the unsupervised, unsupervised is nothing but the clustering. So, the class levels are not available and only the feature vectors are available and based on some similarity, we can do the grouping of the feature of vectors that is the clustering.

(Refer Slide Time: 06:54)



And in case of a traditional pattern recognition system input image I am considering, I can extract the features like sift I can consider or maybe the HOG feature I can consider and after this we can apply some pattern classification techniques, maybe the unsupervised classification techniques or also the supervised classification technique we can apply for classification, for object recognition this is about the traditional pattern recognition system. That means from the input image, from the input signal, I have to extract the features and based on these features, we can do the classification.

(Refer Slide Time: 07:32)

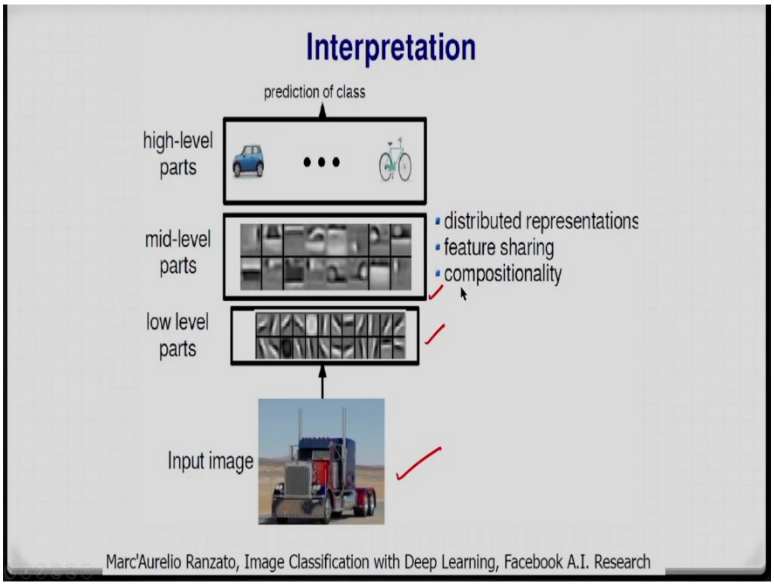


And you can see the difference between the deep learning and the machine learning. So, first if you see the first one is machine learning, so I have the input image, first I have to extract the features and after this I have to apply the classification algorithms and based on this, I can do the classification.

In case of the deep learning the feature of extraction is not important, but what I am considering, I am considering one network that is the suppose artificial neural network, which can directly extract important features from the input image and simultaneously it can do the classification. So, that is about our deep learning.

So, from the input emails directly, we can extract some important features and based on these features, we can and go for classification that is about that deep learning, but in case of the machine learning first we have to extract the features and after this we have to do go product classifications.

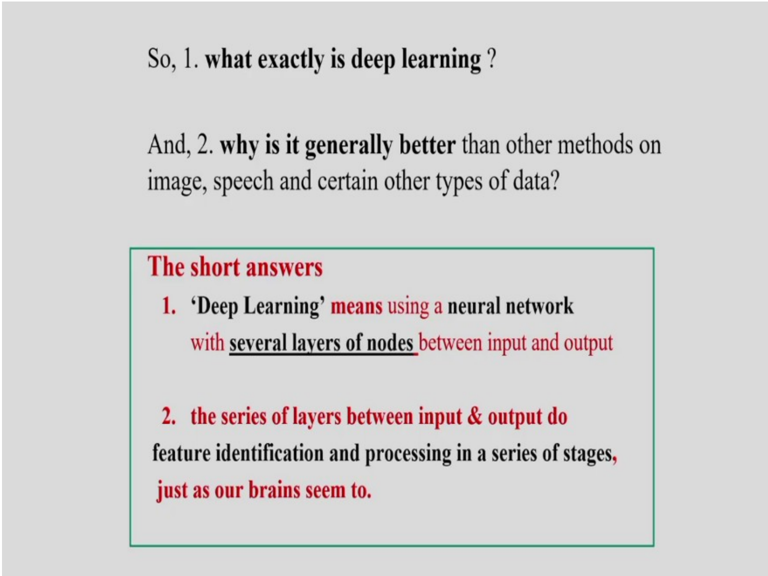
(Refer Slide Time: 08:29)



So, here you can see in case of the deep learning what is available here you see, suppose this is my input image. So, from the input image, I can extract some information, maybe the low-level information I can consider, mid-level information, high level information, I can consider, maybe the low frequency information, high frequency information all the information I can extract.

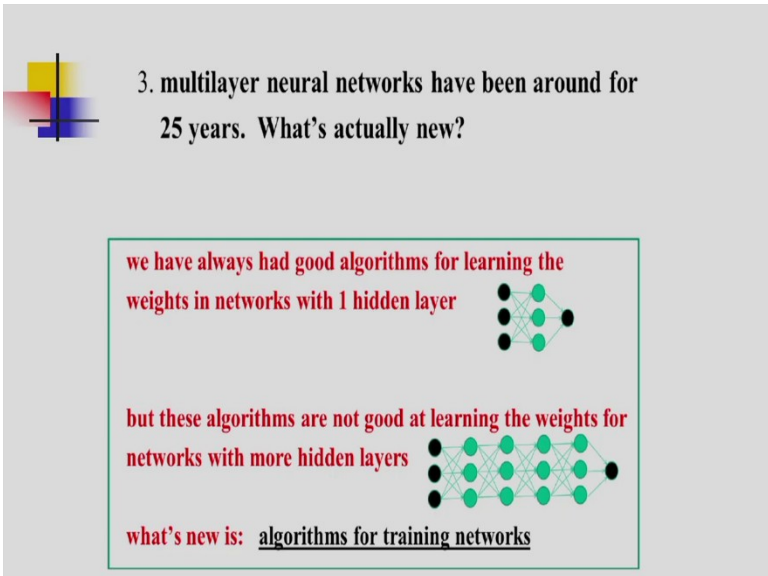
So, first I can extract the low-level parts like this and after this I can extract the mid-level parts and finally, I can extract the high-level parts. So, that is why it is called a deep learning, because I want to extract more and more information from the input image and that is nothing but the cascade of nonlinear transformations and it is the end-to-end learning and considering. So, this is the deep learning techniques and I have shown the interpretation of the deep learning techniques.

(Refer Slide Time: 09:26)



So, what exactly is deep learning and why it is generally better than other methods on image speeds and certain other types of data. So, deep learning means we are considering the artificial neural networks, having several layers of node between input and output and the series of layers between input and output do features identification and processing in a series of stages, just as our brain seems to do.

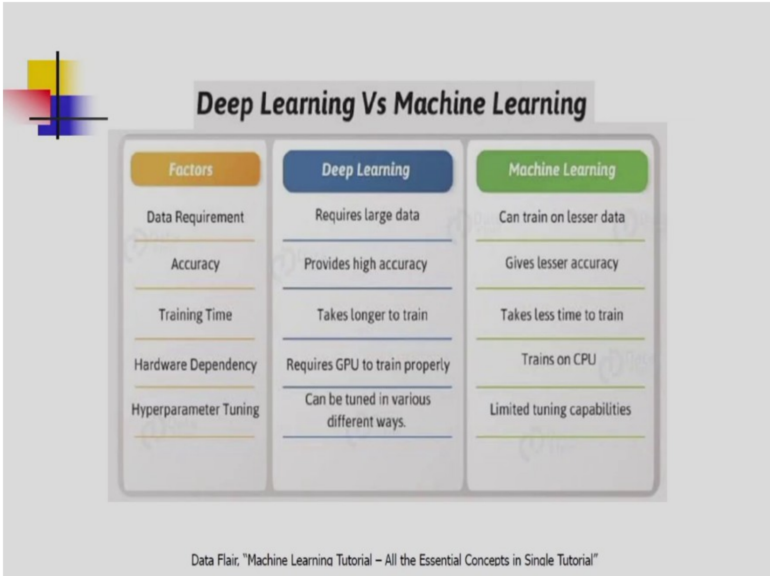
(Refer Slide Time: 09:56)



Multi-layer neural networks have been around for 25 years. So, what is new actually consider the difference between the artificial neural networks and the deep networks. You can see in case of the artificial neural networks; good algorithms are available for learning that width in networks with one hidden layer.

But if I consider more and more hidden layers with the existing algorithms, which are used for the artificial neural networks, that may not be sufficient to train the artificial neural network. So, that is why we have to consider the deep networks in which there are some good learning algorithms are available. So, that we can train the hidden layers.

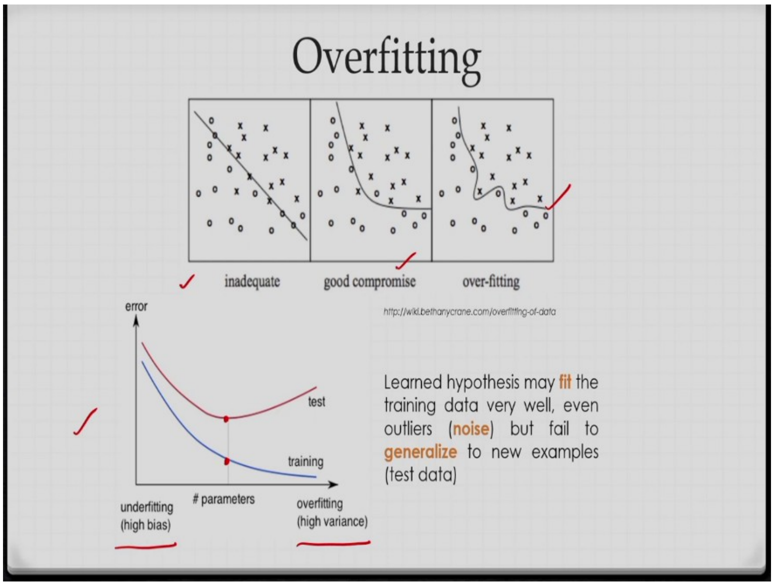
(Refer Slide Time: 10:45)



And if I want to compare the deep learning versus machine learning, you can see I can show some comparisons like this, one is the data requirement. In case of deep learning, it requires large amount of data, in case of the machine learning we can train on lesser data and accuracy provides high accuracy in case of the deep learning.

But in case of the machine learning accuracy is less, the training time another parameter for deep learning, we need the longer training time, but in case of machine learning takes less time to train and hardware dependency if I consider requires GPU to train properly, but in case of the machine learning it can be trained on CPU and also the hyper parameter tuning. So, in case of the deep learning, can be tuned in various different ways. But in case of the machine learning, limited tuning capabilities.

(Refer Slide Time: 11:45)

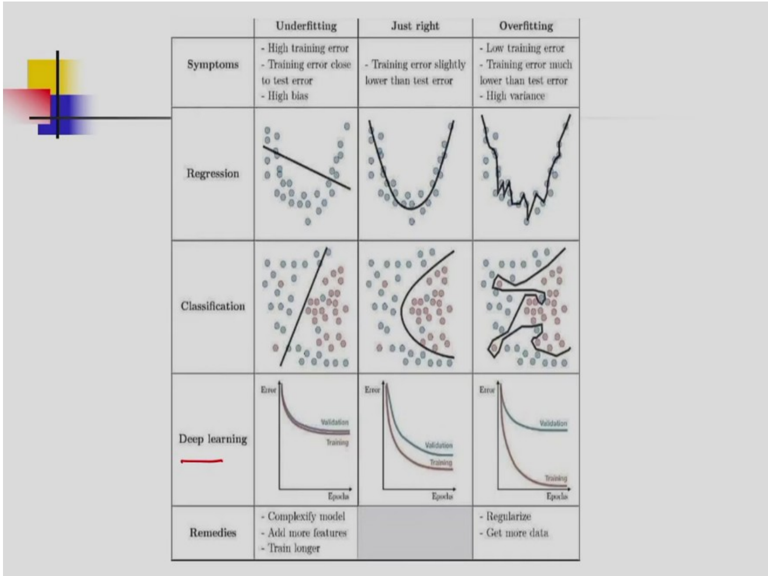


And I want to show the concept of the overfitting So, first figure you can see, in the first figure I am placing the decision boundary between the classes, then in this case you can see it is not the perfect placement of the decision boundary. If I consider these decision boundaries there may be misclassification, but if I consider a second figure that is a good compromise. So, I can consider the decision boundary like this, but in the third case what I am considering that is the overfitting, overfitting in the training.

So, if you see these diagram, if I consider the error versus these different cases, one is the inadequate, one is a good compromise, one is the overfitting corresponding to the underfitting that corresponds to high Bayes you can see the error is maximum during the training and also it is maximum during the testing. But if I consider the good compromise, you can see during the training, the error is less than the underfitting case and also in the testing the error is less than the underfitting case.

So, that is, that this case you can see and in case of overfitting that corresponds to high periods during the training the error is very less, then corresponding to this I am getting the decision boundary between the classes, but in the testing the error is very high that is that overfitting. So, underfitting is also not desirable and overfitting is also not desirable. So, that is that is why we have to consider that case, the middle case, the middle case is a good compromise, we have to consider. So, that is the concept of the overfitting.

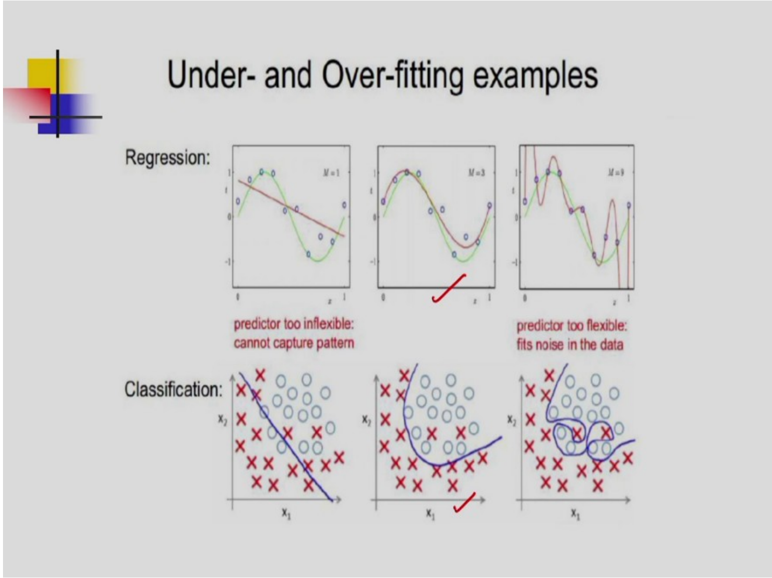
(Refer Slide Time: 13:38)



And in this case, I have shown this case, one is the regression, one is the just compromise and another only the overfitting, corresponding to regression. Second one is the classification, corresponding to classification also I have shown one is the underpinning that means underfitting means high training error and also high Bayes, the training error close to the test error and after this I am considering the just compromise, that is the training error slightly lower than the testing error that is the just right.

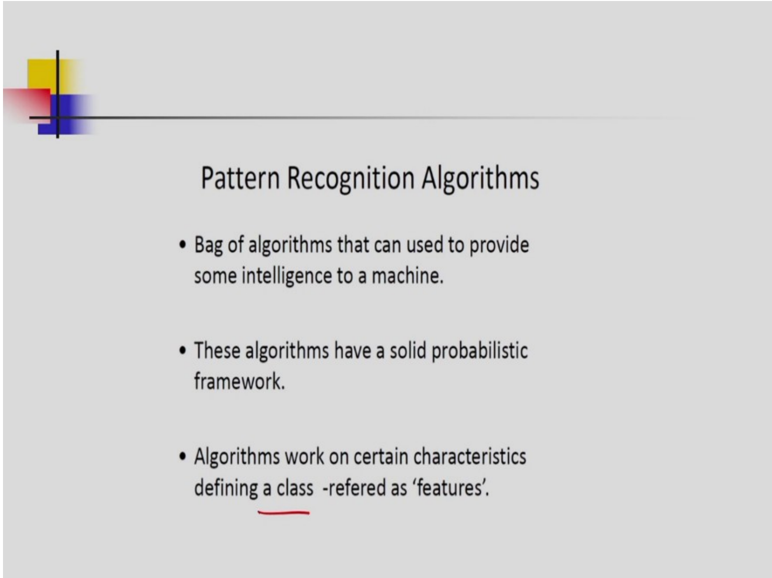
And after this I am considering the overfitting case. So, in the overfitting case low training error and the training error must be lower than the test error and that corresponds to high variance and after this I have shown the deep learning. So, in the deep learning also I have shown the training error and the testing error in all the cases. One is the underfitting, one is the just right and one is the overfitting. So, this is about the underfitting just compromise and overfitting.

(Refer Slide Time: 14:52)



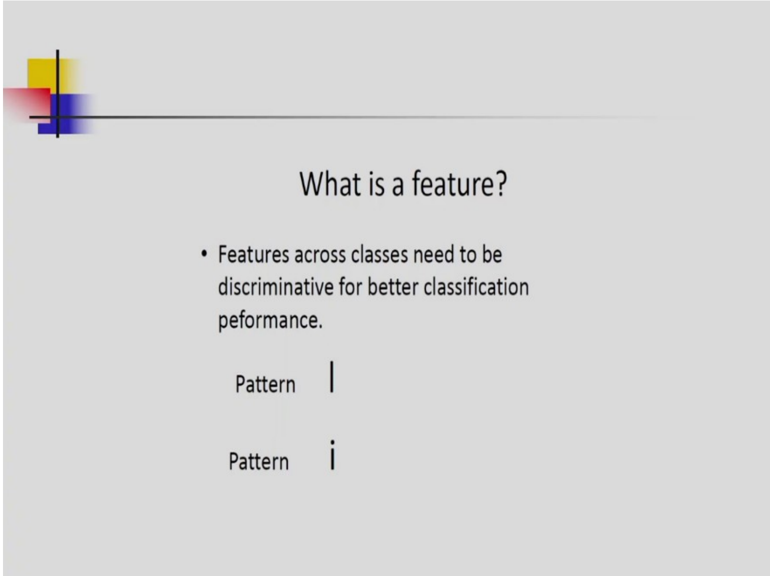
Here also I have shown this all the cases. One is put a regression. One is put a classification. One is the underfitting and one is the overfitting and I am considering the just right in the middle one is the just right I am considering. Regression is nothing but the fitting of the line between the simple points.

(Refer Slide Time: 15:12)



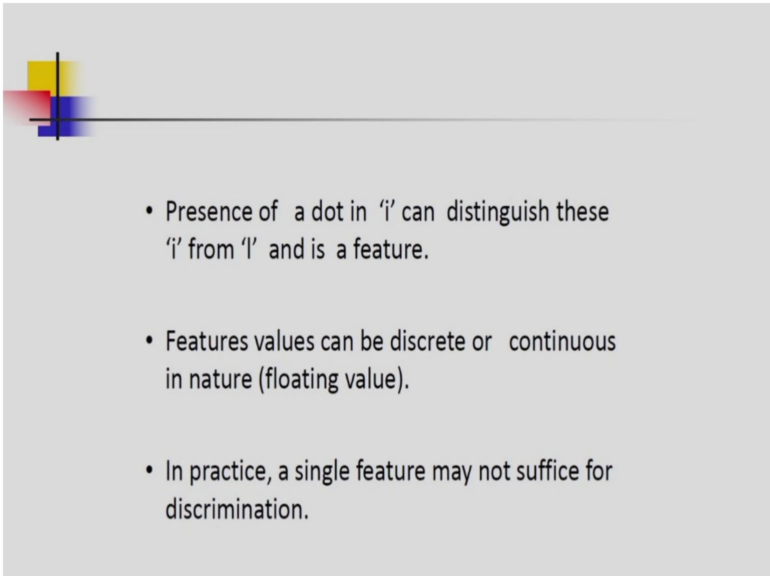
In case of the pattern recognition algorithms, the bag of algorithms that can be used to provide some intelligence to a machine and these algorithms have a solid probabilistic framework and for this we are defining the classes and also, we need the features. So, algorithms work on certain characteristics defining a class, refer as features.

(Refer Slide Time: 15:39)



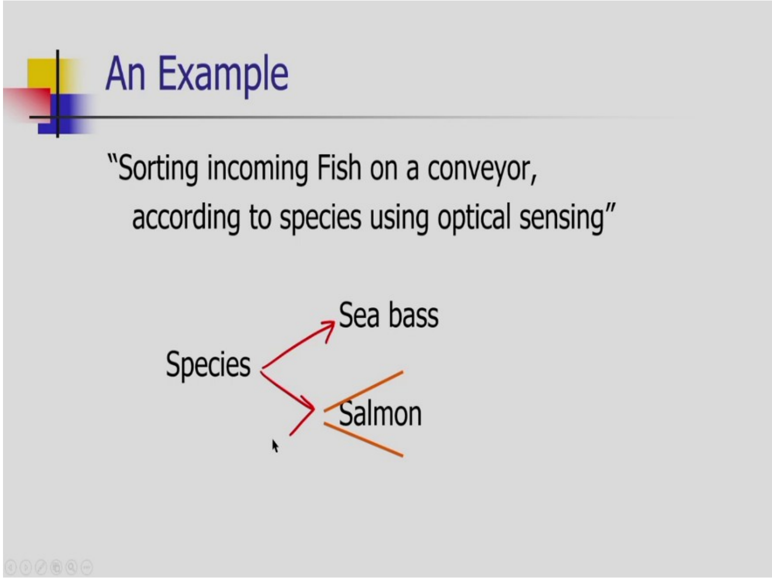
And what is a feature of a feature if I consider, suppose if I want to do the classification, between these two alphabets, one is the I, I another one is i. So, this I is the pattern I can consider. So, what will be the feature for this classification, one is the uppercase and lowercase.

(Refer Slide Time: 16:02)



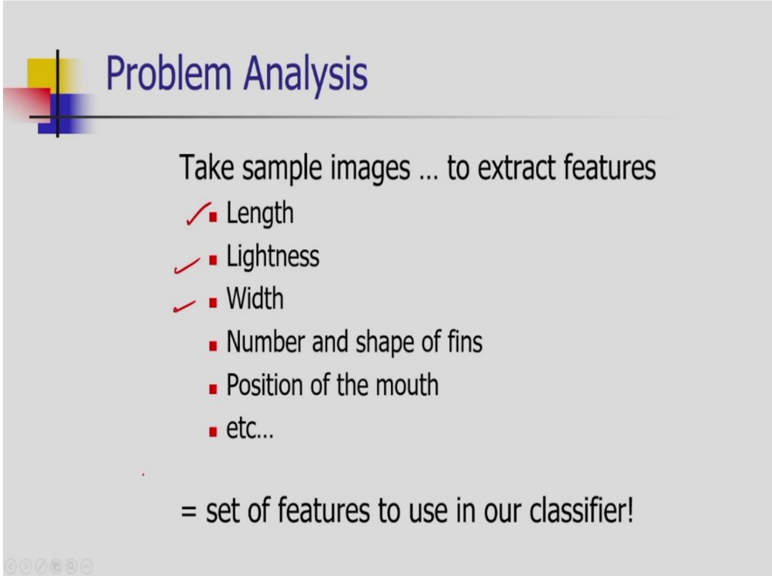
The presence of a dot in lowercase i can distinguish this i the small i from capital I and it is a feature. Feature's value can be discrete or continuous in nature and in most of the cases, the single feature may not be sufficient for discrimination. So, I may need more features for discrimination between the classes.

(Refer Slide Time: 16:25)



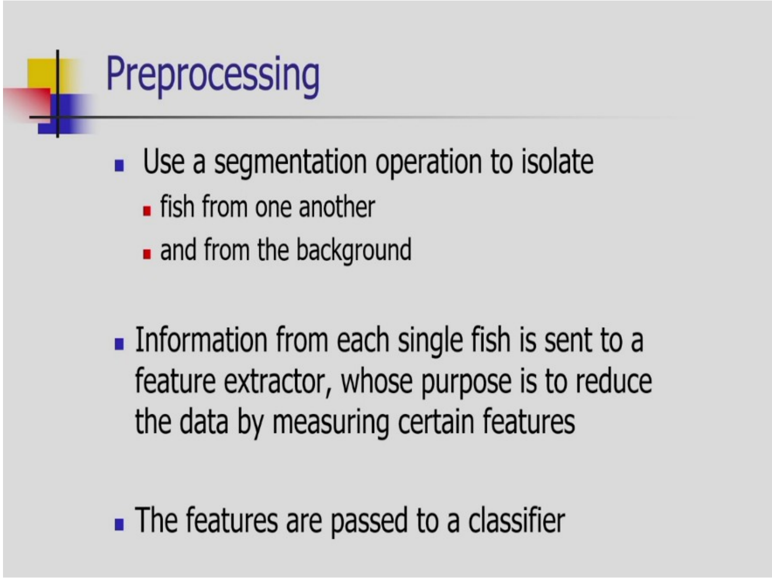
So, in this case, I am giving one example the classification of a fish. So, I am considering two features, one is the sea bass and other one is the salmon. So, how to consider this classification problem.

(Refer Slide Time: 16:40)



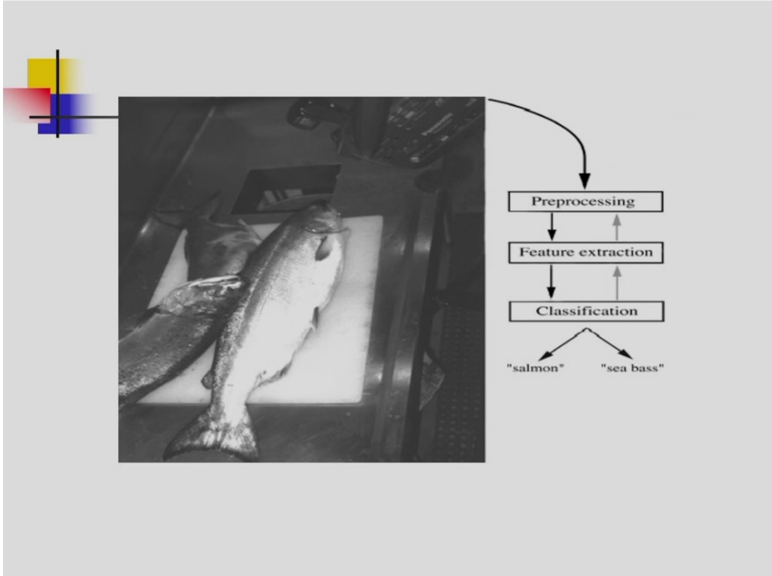
So, maybe we can consider the input images we can consider and we can extract some features, the features may be the length of the fish, the lightness of the fish, the weight of the fish like this, we can extract these features from the input image.

(Refer Slide Time: 16:56)



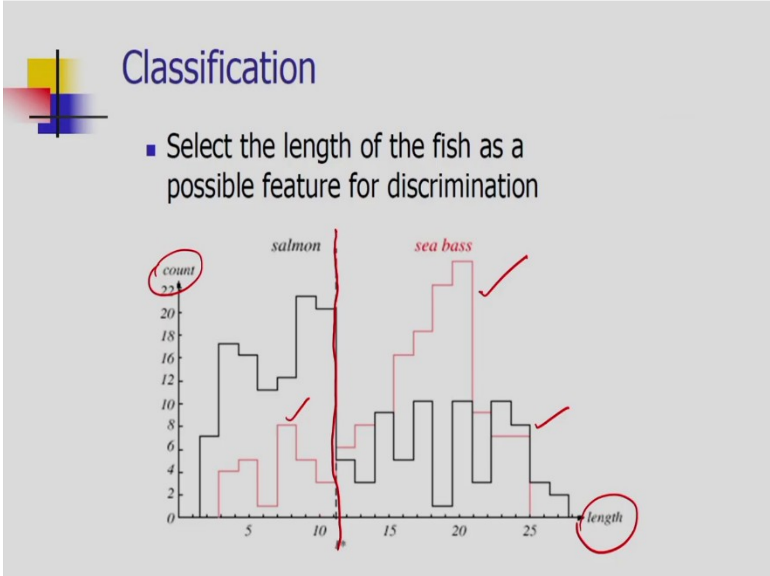
So, after getting the image we have to do some pre-processing image processing we have to do and also, we have to do the segmentation, the image segmentation, that is the segmentation is important to isolate the fish from the background. So, in the image we have the background and the foreground the foreground is nothing but the fishes. So, that means we have to use a segmentation algorithm to isolate fish from one another and from the background and after this we have to extract features.

(Refer Slide Time: 17:26)



So, this is the technique you can see. So, first I have to do the pre-processing input images are available, after this we have to extract features and after this, we have to do the classification. One is the salmon, another one is a sea bass.

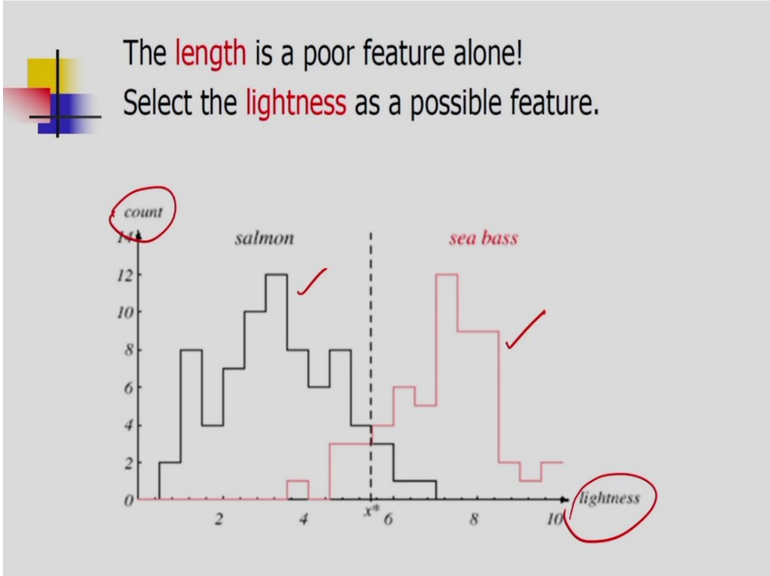
(Refer Slide Time: 17:41)



And in this case you can see I am considering only one feature, the feature is the length, length of the feature I am considering and based on this I am just counting you can see and here you can see, I can recognise salmon and sea bass. But there are some misclassifications, the misclassification is shown by the red one.

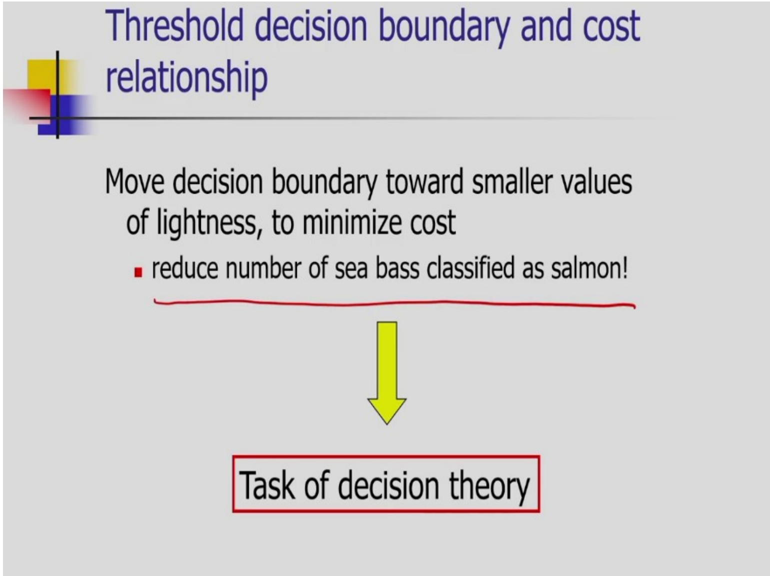
That means, the sea bass is recognised as salmon here you can see, a sea bass is recognised as salmon that is the red one and if I consider that this black one, you can see the salmon is recognised is sea bass and you can see this is the decision boundary, if I consider this is the decision between two fishes, one is a salmon and another one is the seabass. But the misclassification is happening, the seabass is recognised as salmon and salmon is recognised as sea bass. This is by considering only one feature.

(Refer Slide Time: 18:41)



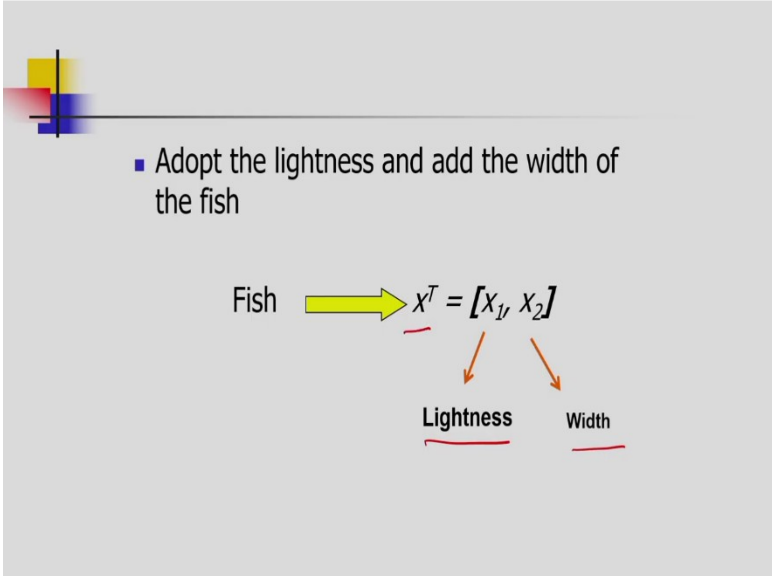
Next figure you can see I am considering another feature, that is the lightness of the fish, I am considering and again I am counting. So, counting means either I am just determining the probability. In this case also you can see there are some misclassifications, the red one is the sea bass and the black one is the salmon. The sea bass is recognised as salmon and salmon is also recognised as sea bass. So, this is the misclassification is taking place, because I am considering only one feature.

(Refer Slide Time: 19:16)



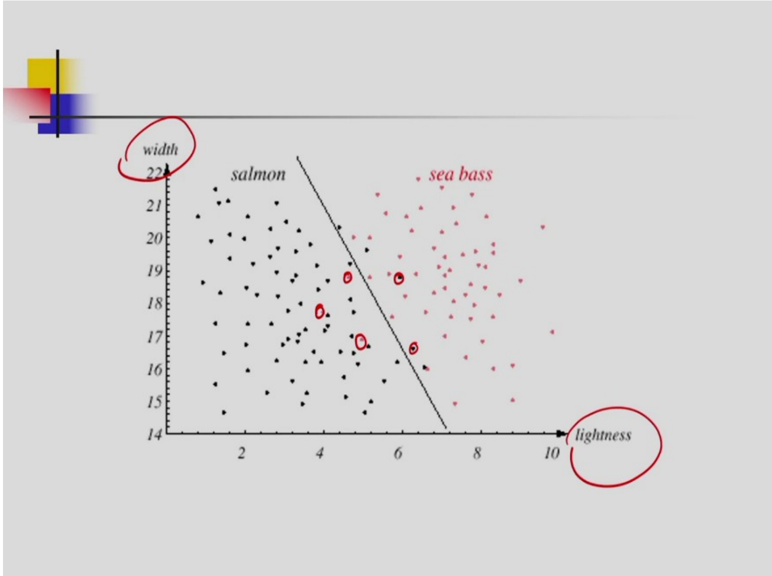
So, that is why we can consider maybe or two features or maybe more features for this classification problem. So, we have to minimise the misclassification, we have to increase the accuracy. So, we have to reduce the number of sea bass classified is salmon, that we have to reduce. So, for this we have to consider the decision theory. What decision theory we can consider.

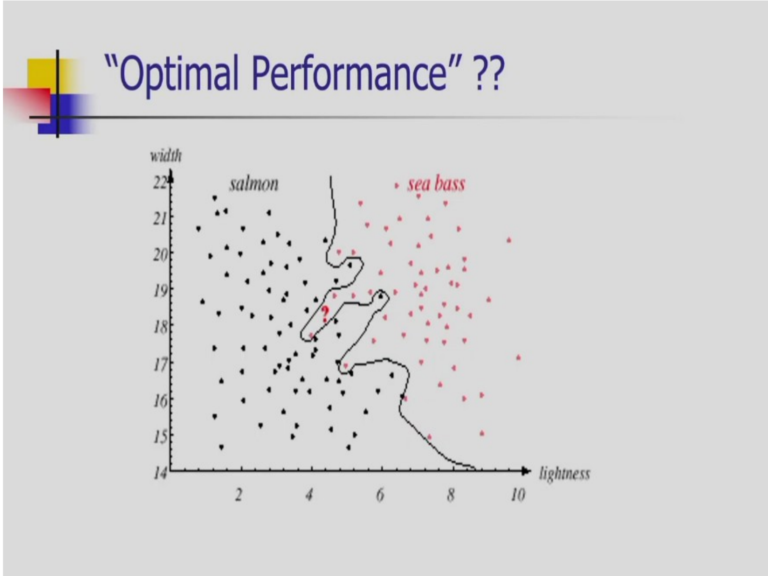
(Refer Slid Time: 19:42)



You can see here, so I am considering now two features, one is the lightness of the fish, another one that width of the fish. So, that is the feature of vector I am considering, X is the feature vector.

(Refer Slide Time: 19:54)

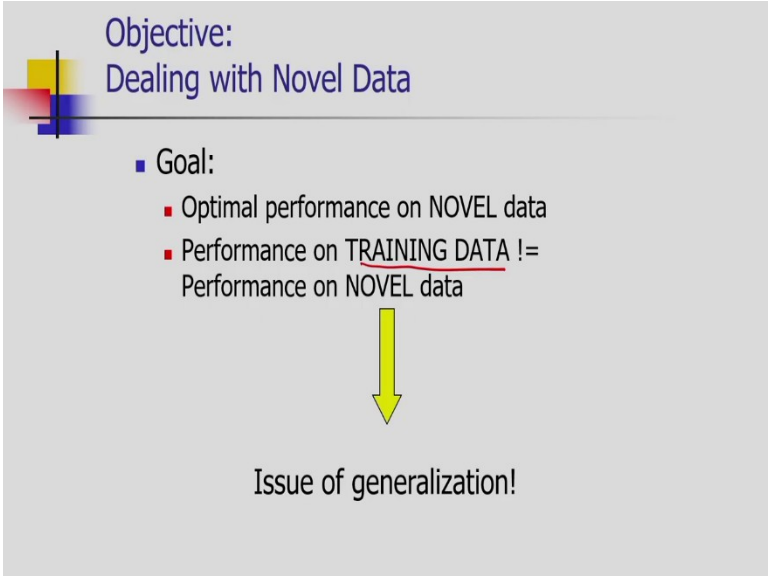




And if I consider these two features, here you can see it is a two-dimensional feature space. I am considering one feature is the lightness, another feature is that width and you can see the decision boundary between these two classes, one is the same one and another one is the seabass.

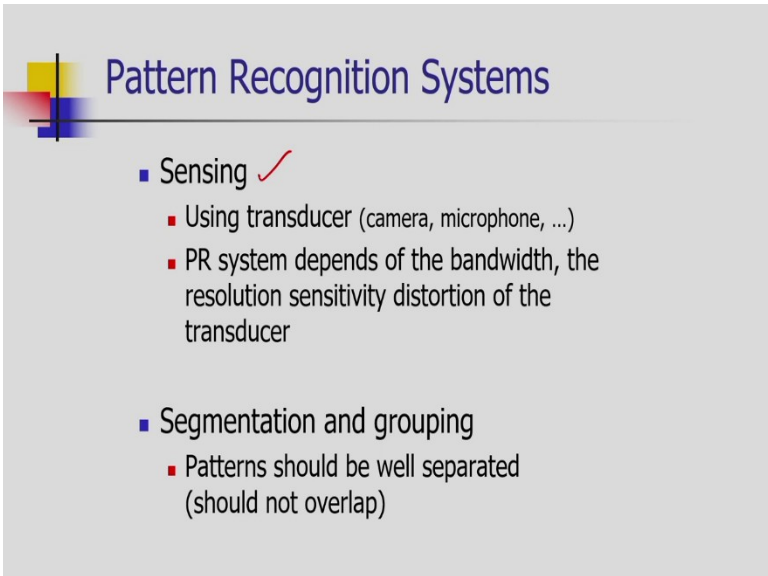
Then in this case the classification accuracy increases, because of considering two features, but in this case there are still some misclassification you can see, some misclassifications are happening in this case also, but because of this two features, the accuracy increases as compared to previous cases and in this case, you can see for optimal performance, we have to select the best decision boundary between the classes, if I consider these decision boundary and that is a nonlinear decision boundary and that classification error will be less, between these two classes.

(Refer Slide Time: 20:51)



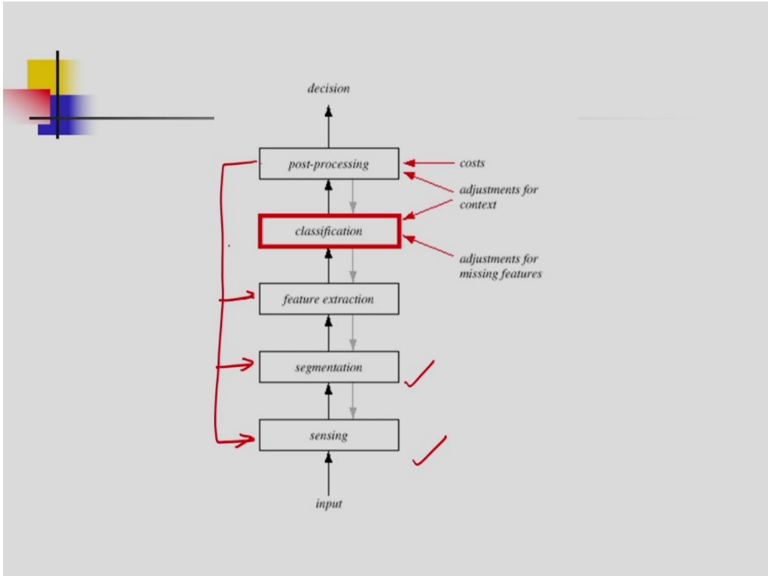
The objective is dealing with novel data and in this case, we have to the performance of the training data and that is important. First, we have to do the training and we have to see the performance on the training data and after this the performance on the new data novel data we have to consider, that should be good, the performance for the training and the testing should be good.

(Refer Slide Time: 21:13)



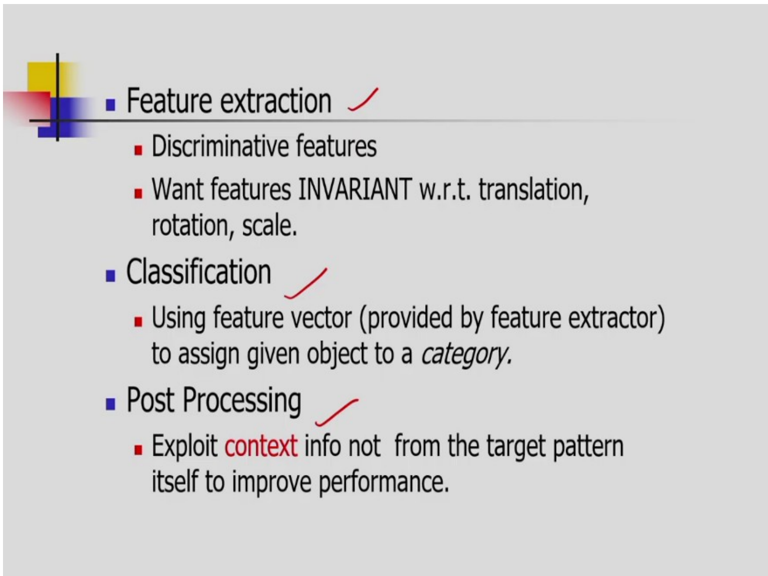
So, in case of the pattern recognition systems, what are the main steps? The first one is the data acquisition that is nothing but the sensing, that is I can consider in case of the image classification, I have to consider image acquisition or maybe the sensors we can consider for signal acquisition and after this we have to do the pre-processing and after this we have to do the segmentation in that grouping. So, patterns should be well separated and should not overlap that we have to consider.

(Refer Slide Time: 21:49)



Here you can see the block diagram of the pattern recognition systems, first one is the sensing, after this we have to do the segmentations. The next step is the feature extraction and all the features may not be discriminative. So, that is why we have to go for features selection and after this the classification and the post processing, the post processing is nothing but the system evaluation. So, we can determine accuracy, we can determine the false positive rate that true positive rate. So, all these rates we can determine and based on this we can adjust the parameter of the system that is nothing but the feedback. So, we can just do the feedback.

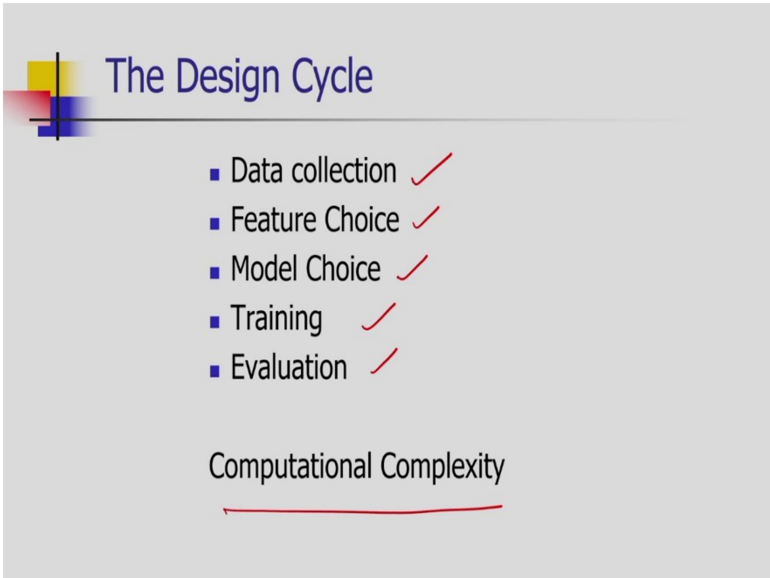
(Refer Slide Time: 22:31)



We can see here the next step is the feature extraction and we have to consider the discriminative features and one important point is the features should be invariant to a point transformation. Like if I consider translation rotation scale. So, in this case also the features should be invariant to a point transformation.

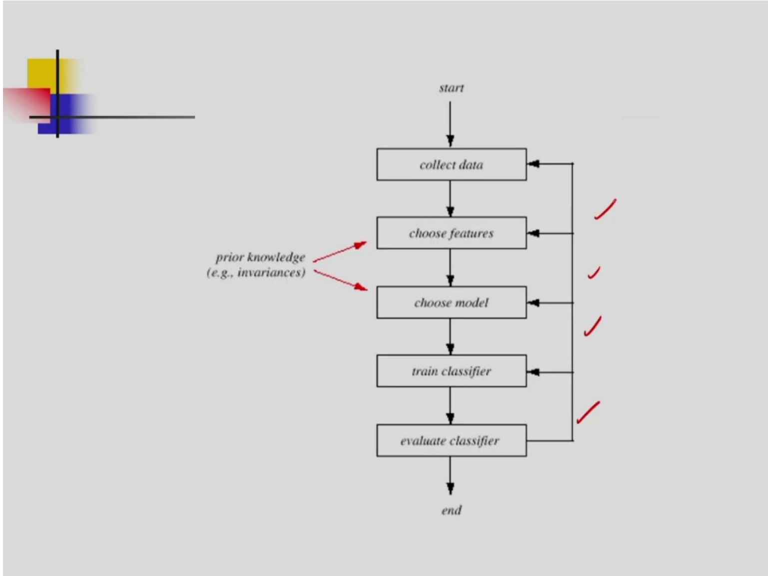
After extracting the discriminative features, we can go for classification and after this the post processing that already I have explained. So, we can improve the performance. So, for this, we can give the feedback to all the steps like a feature extraction, feature selection. So, that feedback is important from the output to the inputs.

(Refer Slide Time: 23:13)



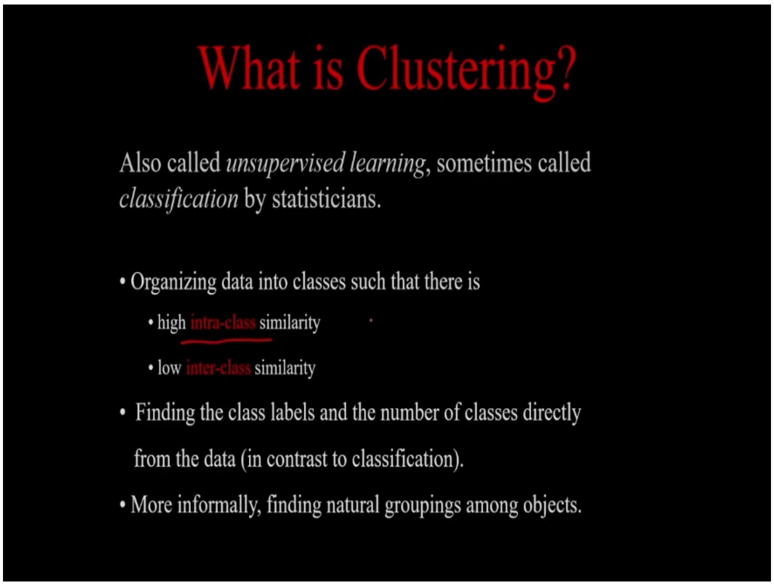
So, now the design cycle is first one is the data collection. Next one is the feature selection, after this the models election, we have to do the training and after this the evolution of the pattern recognition system and one important point is the computational complexity, we have to determine the computational complexity of the system that we also, this is important.

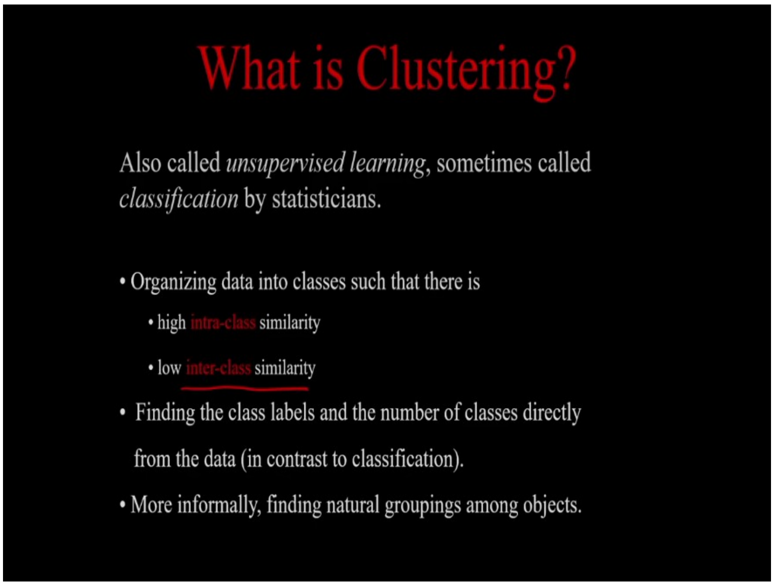
(Refer Slide Time: 23:37)



So, same thing I am showing here, first one is the collection of data after this the selection of the features, selection of the model and after this we have to train the classifier and after this, the evaluate the classifiers and here you can see the feedback. So, I am considering the feedback, so that we can improve the performance of the pattern recognition system.

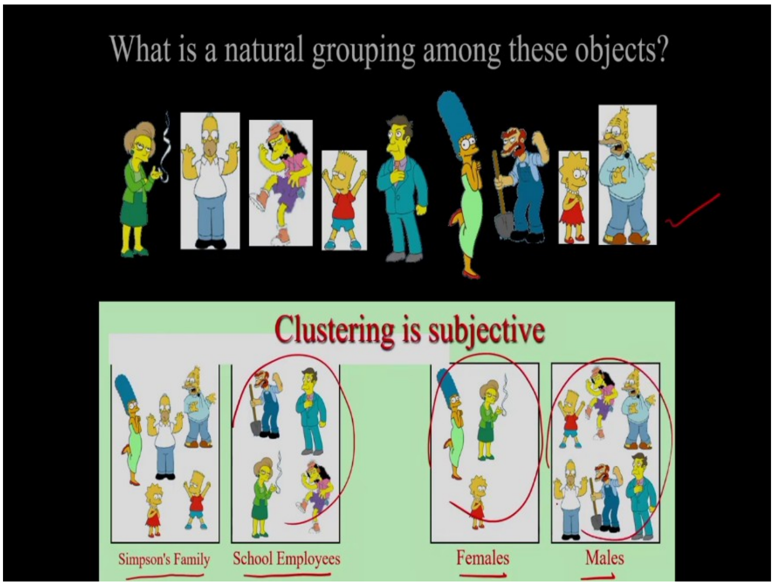
(Refer Slide Time: 24:00)





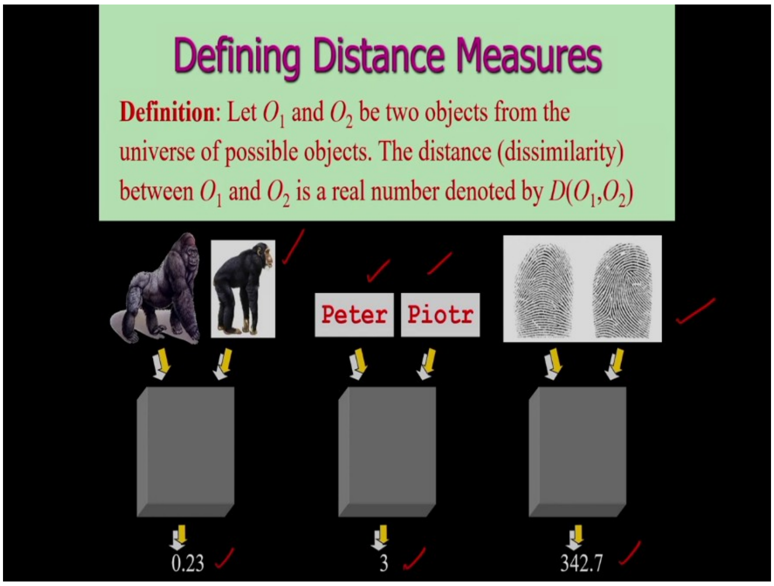
And what is clustering, clustering is the unsupervised learning and, in this case, you can see that we have to do the grouping and one important point is we have to consider high intra-class similarity we have to consider and also we have to consider the low inter class similarity. In case of the grouping, this is very important, one is the high intra class similarity, another one is the low inter class similarity that is important. So, based on this we can do the clustering.

(Refer Slide Time: 24:31)



So, in this case, you can see I am considering these images suppose and based on this we can do the groupings, but the clustering is subjective. If you see the second figure you can see, if I consider the family, the class, the class will be like this. If I consider the school employees, that class will be like this, if I consider only females, then the grouping will be like this. If I only considered males, then the grouping will be like this. So, that is why the clustering is subjective.

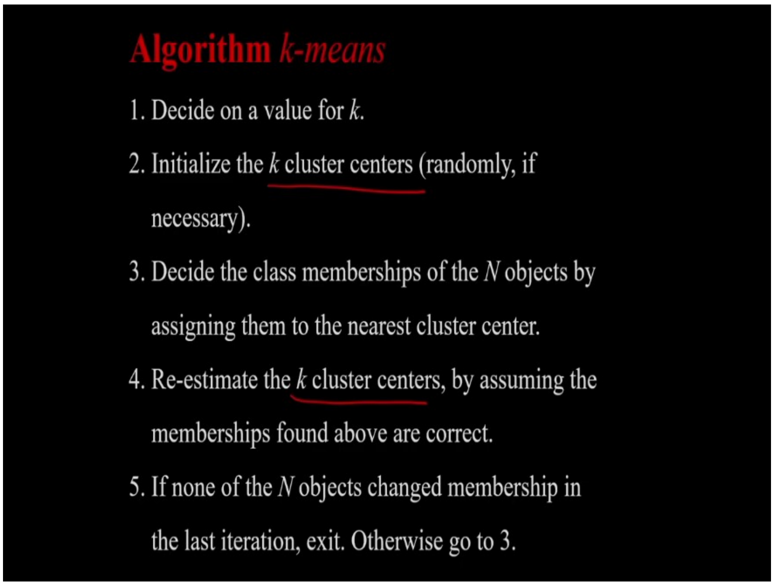
(Refer Slide Time: 25:03)



And in this case, already I have explained that the feature vector is available and based on some similarity, we have to group this feature of vectors. Here you can see I am giving one example, I am considering the, in the first example I am considering these two images and I want to find a similarity between these two.

So, maybe some equivalent distance measure we can consider for finding the similarity between these two images and I am getting the similarity score of 0.23. Similarly, in the second case also if I consider these two cases, the similarity score is 3, a third case I am considering the similarity between two fingerprints and the similarity score is 342.7 I can determine and these values can be normalised. So, this is the by using the distance measure I can find the similarity between the patterns.

(Refer Slide Time: 25:57)

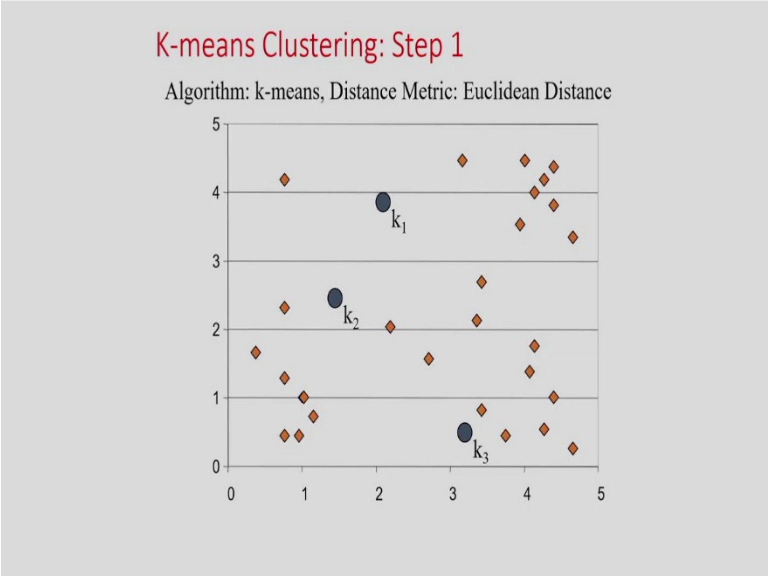


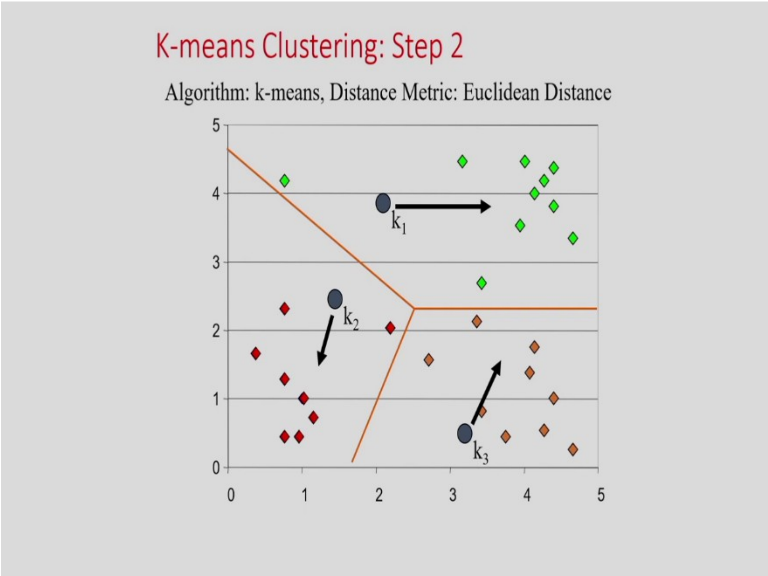
And for this clustering one popular algorithm is, the k means clustering, the k means clustering already I have discussed in one of my classes, let me image segmentation class I have discussed about the k means clustering. So, for this I have to consider the k number of cluster centres that I can select randomly and after this what we can consider and decide the class membership of the N objects by assigning them to the nearest cluster centre.

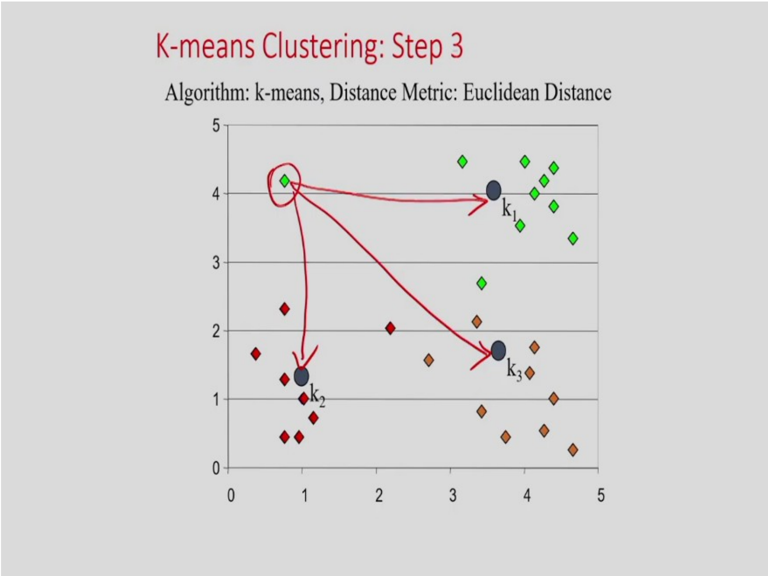
So, for this by measuring the distance between the sample points and the cluster centre, I can decide that one. So, whether their particular sample points belong to a particular cluster that I can decide based on this distance measure that is the minimum distance after this I have to re-estimate the k cluster centres, I have to re-estimate and after this, this process I have to repeat again and again and finally what will happen, this clusters centre will not be moving too much.

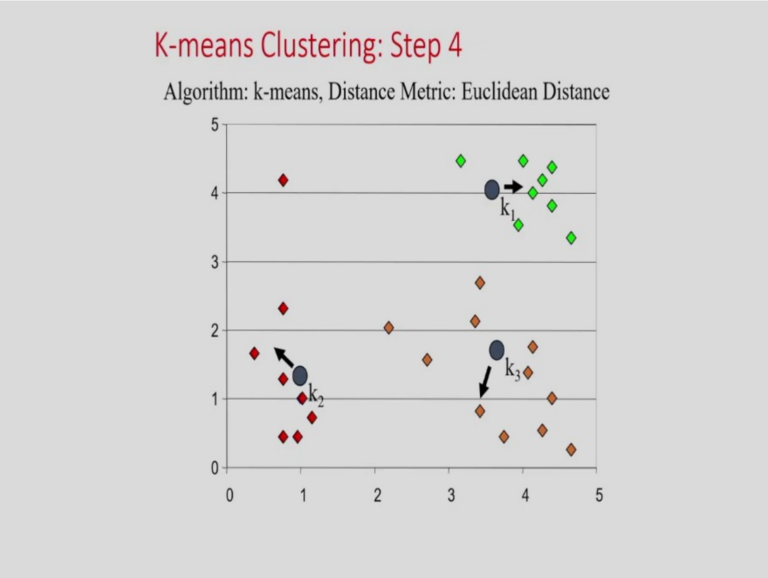
It will be fixed after some iteration and corresponding to this, I will be getting the cluster centres corresponding to k number of or maybe the c number of classes. So, pictorially I can show this algorithm like this.

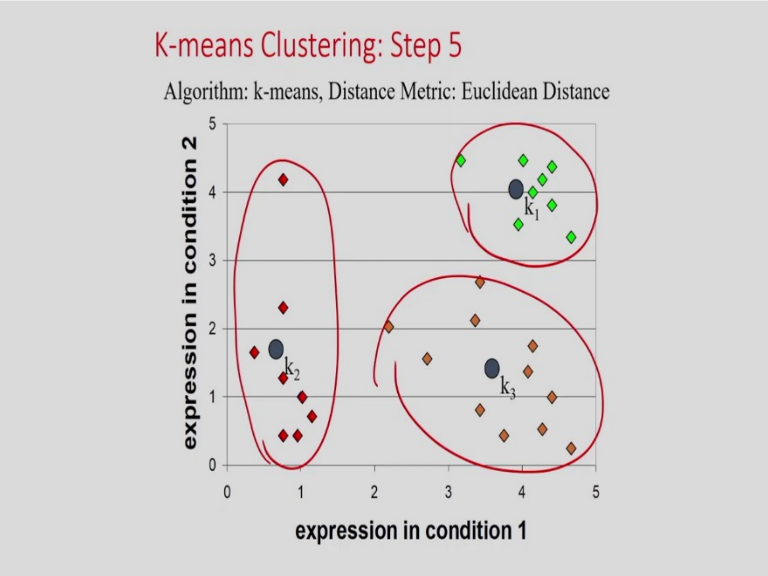
(Refer Slide Time: 27:16)











So, you can see I am considering some sample points and I am considering randomly selecting these centroids k_1, k_2, k_3 centroids I randomly am selecting and after this you can see I am assigning the sample points to a particular centroid based on the distance measure, I can consider the Euclidean distance.

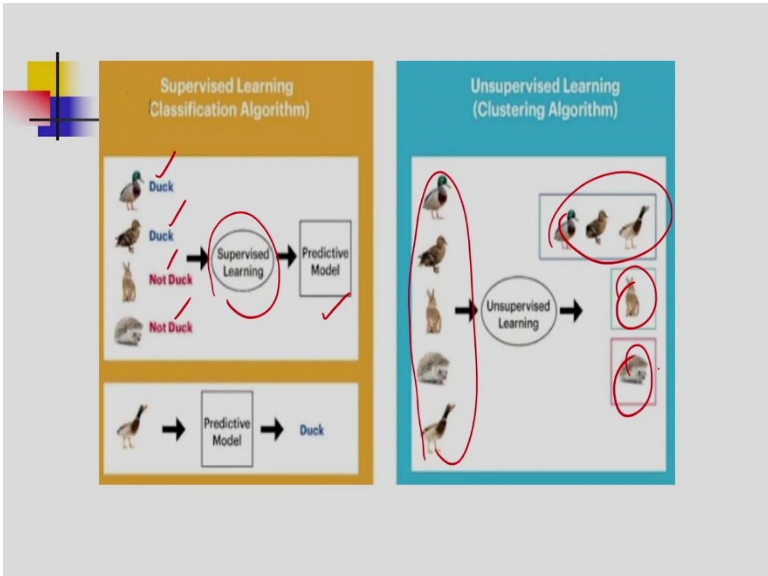
So, corresponding to this I am getting three classes one is the green, another one is the red, another one is orange and after this assignment, this cluster centre will be moving because I have to recompute the cluster centre that is the centroid. So, after the competing you can see the k_1 is moving, k_2 is moving, k_3 is moving, so that I have to recompute the cluster centres.

And after the computing you can see I am getting the new position of k_1, k_2 and k_3 and after this again I have to decide the sample points which will belong to a particular centroid. So, now, you can see these green suppose this point now belong to the cluster centre k_1, but if I find the distance between this point and this k_1 suppose if I find a distance between this one k_1 distance between k_2 distance between k_3 I have to find a minimum distance.

Now, in this case the minimum distance will be with k_2. So, that means that this point should be assigned to the cluster centre k_2. So, that means, you can see initially it was the green point and after this after finding the distance you can see this point is assigned to the cluster centre k_2 and after this again I have to recompute the centroids k_1, k_2, k_3 it will be shifted like this and this process I have to do it iteratively.

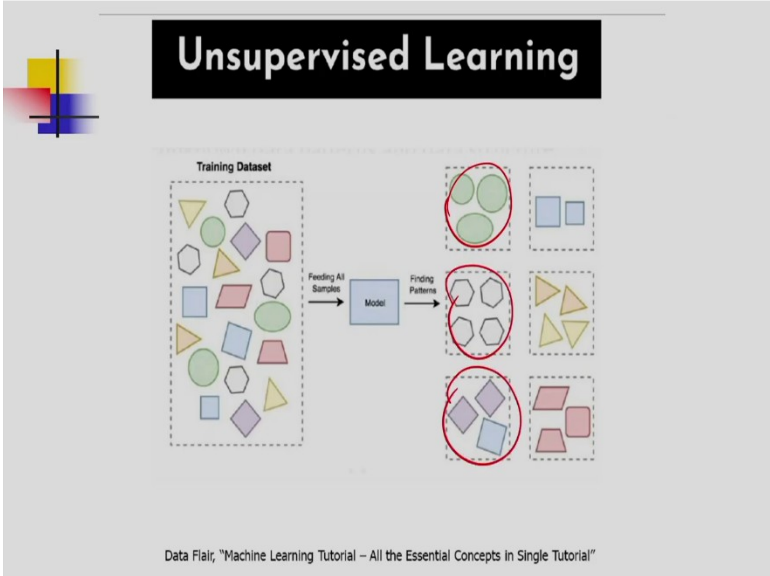
And finally, if there is no significant change of the values of k_1, k_2, k_3, then I have to stop the iteration and there will be the final position of the centroids belonging to the cluster. So, in this case I will be getting the cluster, one cluster is this another cluster is this and another cluster is this, so k_1, k_2, k_3.

(Refer Slide Time: 29:40)



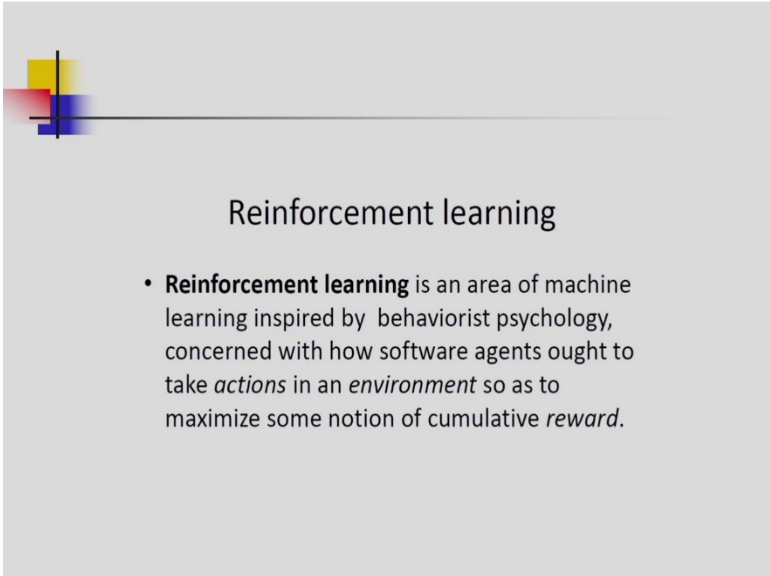
And in this case, you can see I have shown the concept of the supervised learning and unsupervised learning. In case of supervised learning, we know the class levels and also, we have the training samples for all the classes and you can see we have the classes like this duck, duck, not duck, not duck like this and we have to go for the supervised learning and we can do the prediction that is that supervised learning algorithms. In case of the unsupervised learning, what we can do, we can do the clustering. So, this is one cluster, this is one cluster, this is one cluster, based on some similarity, we can do the clustering.

(Refer Slide Time: 30:22)



Again, I am showing this one. So, how to do the clustering. So, we have the training data set and after this based on some similarity, we can do the clustering. So, you can see, so this is one cluster, like this we have the cluster, the groups. This is about the unsupervised learning.

(Refer Slide Time: 30:45)



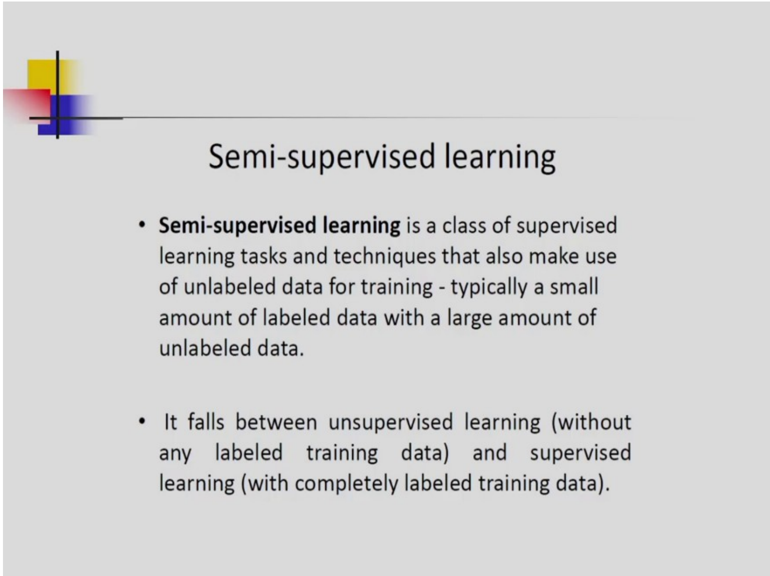
Now, I will discuss the concept of the reinforcement learning. So, concept is like this, suppose in some applications, the output of the system is a sequence of actions in such a case, a single action is not important. So, what is important, the important is, is the policy that is the sequence of correct actions to reach the goal and there are no such things as the best action in any intermediate state. So, that means, in any intermediate step, I cannot say that this is the best action.

An action is good, if it is a part of a good policy and machine learning programme should be able to assess the goodness of the policies and learn from past good action sequences to be able to generate a policy. This is the concept of the reinforcement learning. So, briefly I will explain one example.

So, if I consider the playing of chess. So, where a single move by itself is not that important, it is the sequence of right moves, that is important, a move is good, if it is the part of a good game policy and another example, I can give in robot navigation.

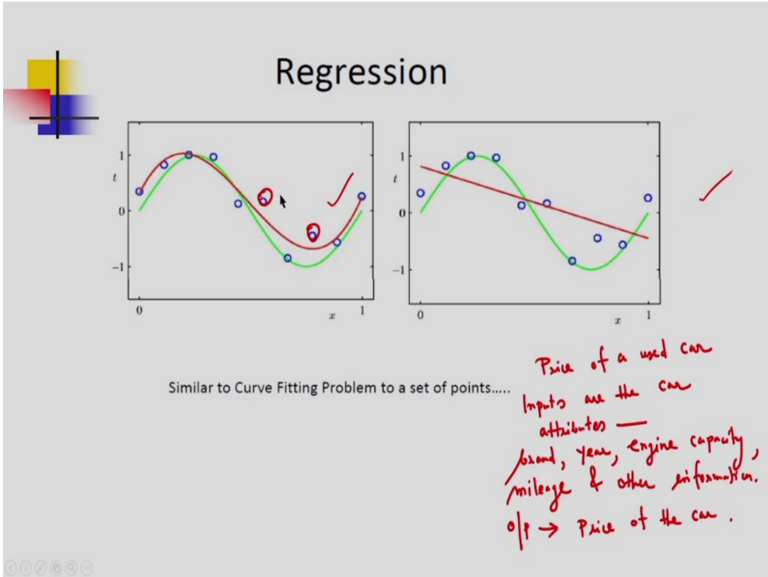
Now, a robot can move in one of a number of directions and after a number of trial runs, it should learn the correct sequence of actions to reach to the goal state from an initial state and in this case, it should not hit any of the obstacles. So, I have given these two examples. So, for this we can apply reinforcement learning. So, one action is not important, the group of actions or the policies more important that is the main concept of data reinforcement learning.

(Refer Slide Time: 32:48)



After these the next concept is semi supervised learning. Semi supervised learning is a class of supervised learning tasks and techniques that also make use of unlabelled data for training. Typically, a small amount of label data with a large amount of unlabelled data. So, that means, in case of the semi supervised learning, we have a small amount of the label data and also, we have the large amount of unlabelled data. So, that is it is in between the supervised learning and the unsupervised learning.

(Refer Slide Time: 33:27)



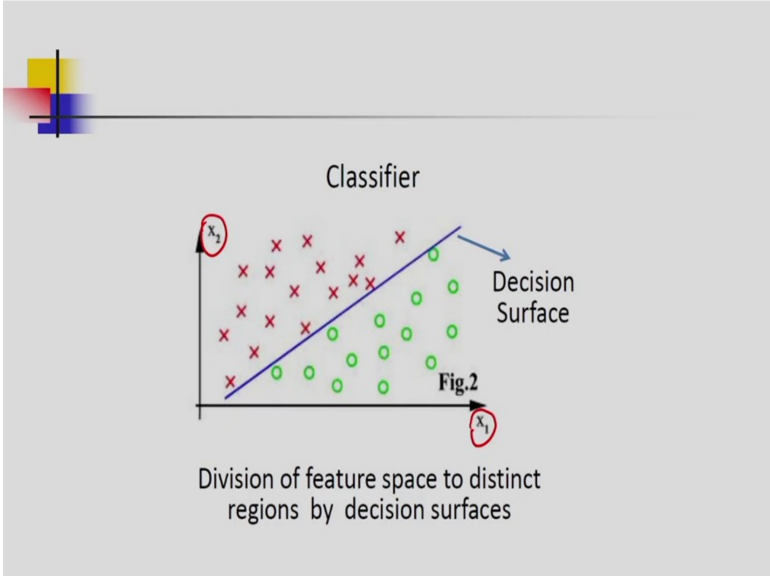
Next concept of the regression means, you can see similar to curve fitting problem to a set of sample points. So, here I am considering some sample points and I am hitting a curve. So, you can see in the figure, first figure and the second figure, I am doing the regression that is, it is similar to curve fitting problem to a set of sample points.

And briefly I can say the linear regression is a statistical method that allows us to model a relationship between a dependent variable and one or more independent variables and this can be done by fitting a linear equation to the observed data. So, that is the definition of regression. So, it is a statistical method that allows us to model the relationship between a dependent variable and one or more independent variables.

Suppose I can give one example, suppose we want a system so that it can predict the price of a used car. What I am considering the price of a used car that I am considering. So, what are the inputs? So, inputs will be, inputs are the car attributes. So, what are the attributes? I can consider suppose the brand, brand is one attribute maybe the year of manufacturing that I can consider or maybe I can consider engine capacity, engine capacity I can consider or maybe I can consider the mileage, I can consider or maybe some other information I can consider.

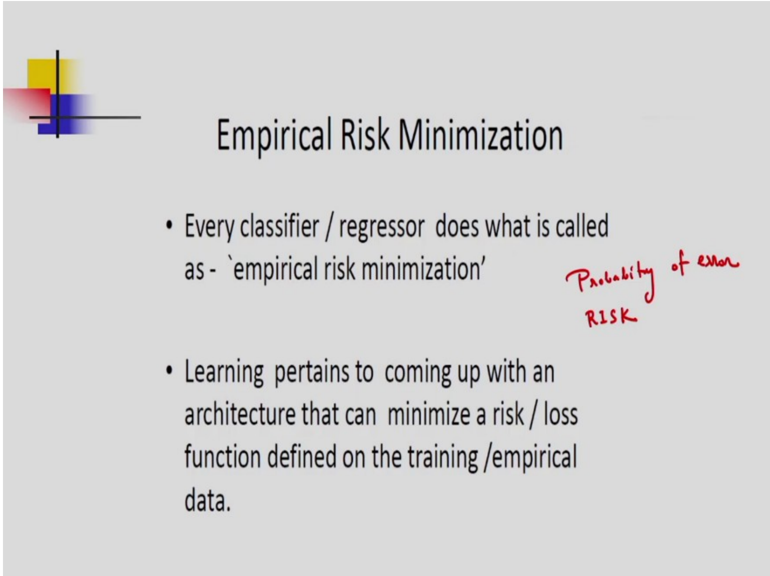
So, based on these attributes, I can determine the price of a used car. So, the output is the price of the car. So, output will be, the output will be the price of the car. So, this is the regression problem. So, briefly I have explained the concept of the regression. So, it is a statistical method that allows us to model the relationship between a dependent variable and one or more independent variables and this can be done by fitting a linear equation to the observed data. So, here you can see I am considering some observed data and I am fitting a curve.

(Refer Slide Time: 36:18)



Next, I am considering the classifier. So, in this example, you can see I am considering a feature of space, the two-dimensional feature space. So, x_1 and x_2 I am considering the features and you can see I am considering two classes, one is the red plus another one is the green plus and corresponding to this you can see the decision boundary between the classes. So, this is one example of a classifier that already I have explained in my last class.

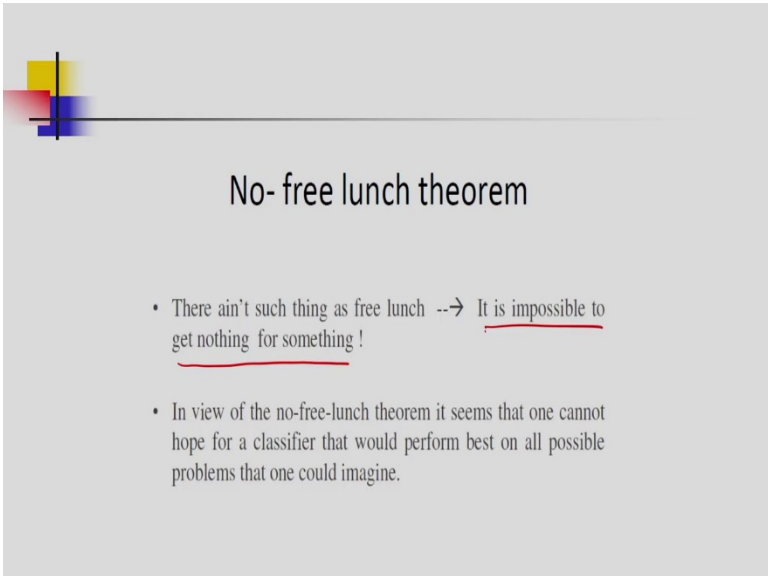
(Refer Slide Time: 36:49)



And one is the empirical risk minimization. So, based on this principle and that one is the probability of the error I can determine. So, in my next class, I will define what is the probability of error, probability of error I can determine and another one is the risk. So, these are two techniques by using this, we can take a particular classification decision.

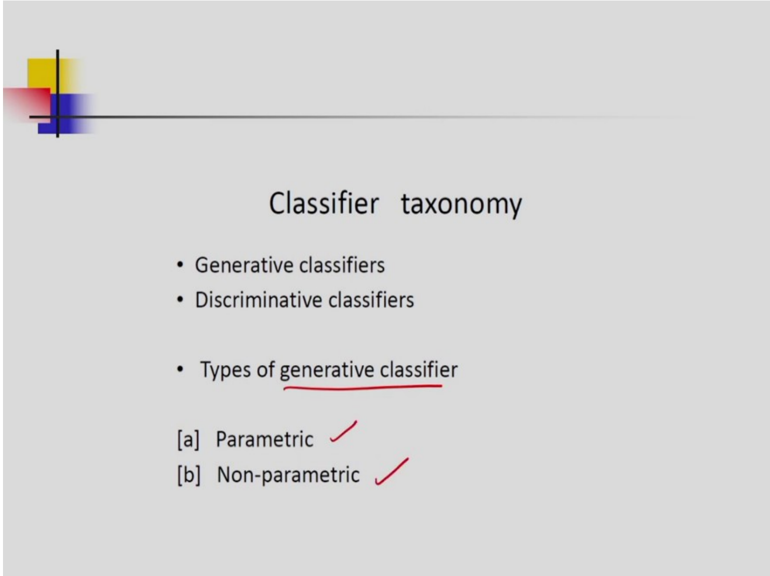
So, we can determine probability of error and also, we can determine risk. So, risk means, suppose if I consider a particular action corresponding to this particular action, I can consider a risk. So, this conditional risk I can determine and based on this conditional risk, I can take a classification decision. So, that is the empirical risk minimization. So, I have to minimise the risks.

(Refer Slide Time: 37:48)



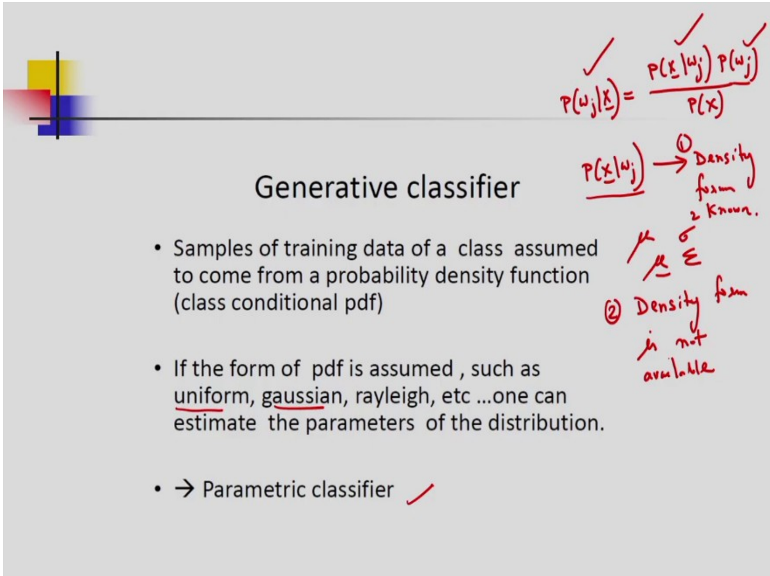
And one is that no free lunch theorem What is the meaning the meaning of this, so it is impossible to get nothing for something. The meaning is that in view of the no free lunch theorem, it seems that one cannot hope for a classifier that would perform based on all possible problems that one could imagine. That means, suppose you consider one classification algorithm that may be suitable for a particular application, but that may not be suitable for many other applications. So, it cannot be generalised. So, that is the concept of the no free lunch theorem.

(Refer Slide Time: 38:30)



And classifier taxonomy. So, there are two types of classifier, one is the generative classifiers, another one is discriminative classifiers. So, first I will be discussed about the generative classifiers, in case of generative classifiers we have two categories, one is the parametric classifier and other one is the nonparametric classifier. So, let us see what is the generative classifier.

(Refer Slide Time: 38:58)



So, in case of the generative classifier, a samples of training data are a class assumed to come from a probability density function that is the class conditional pdf. So, if you remember that in case of the Bayes theorem, we have shown like this, the probability of w_j given x that one, that I want to determine is equal to the P (x $|w_j$) and the prior is probability of w_j and this is the evidence.

So, evidence has no role in classification. So, it is simply a normalising factor. Now, suppose this is the class conditional density, the class conditional pdf. So, suppose the density form is known, density from its known. So, that means density form maybe it is the uniform density, Gaussian density like this we can consider and, in this case, if I know the density form, we can determine the parameters.

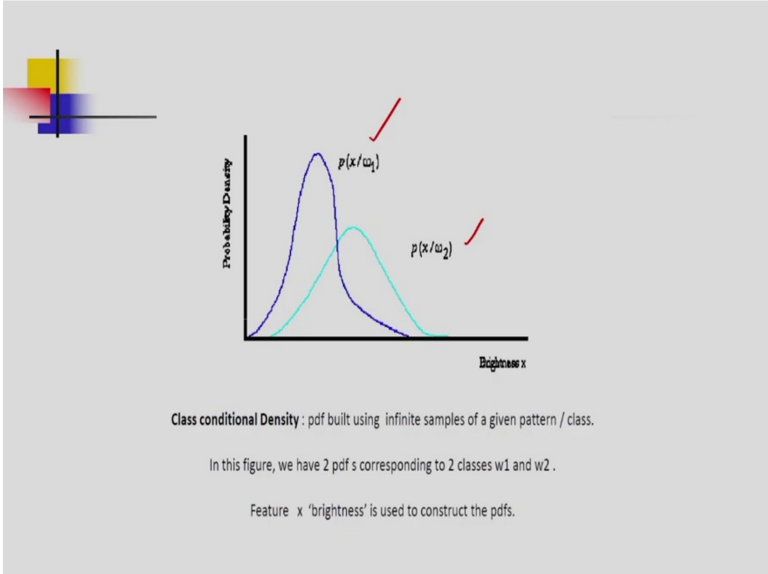
Supposing if I consider the Gaussian distribution in case of the Gaussian distribution, I have two parameters, one is the mean, another one is the variance. So, that I can determine if I know the density form. So, based on this class conditional density, because if I know the density form, if I know then based on this, I can take a classification decision, because if I want to determine this the probability of w_j given x I have to know the probability of x given w_j and also, I have to know the prior probability.

So, by using the Bayes law you get this one and if the density form is known that is the density form the likelihood. This is the likelihood is known, then it is called the generative classifiers and sometimes the density form may not be available. So, then in this case we have to estimate the density. So, there are two cases the first case is, the density forms its known and we have to estimate the parameters. So, this is the first case.

The second case is density form is not known, density form is not available or not known, then in this case we have to estimate the parameters. In the first case, number one the density form is known and we have to estimate the parameters, the parameters may be mean and the variance or if I consider high dimensional case, that will be the mean vector and also, I have to consider covariance and in the second case, the density problem is not available, so we have to estimate the density.

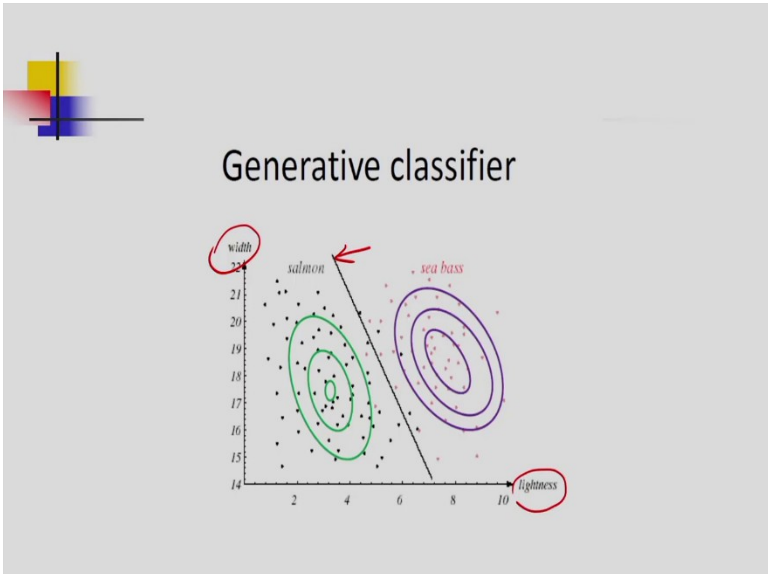
So, first I will consider the parametric classifier. So, what is the parametric classifier, the parametric classifier means the density form is known and we have to estimate the parameters. So, that is the parametric classifier, in case of the nonparametric classifier, density form is not known, then we have to estimate the density that is called nonparametric classifier.

(Refer Slide Time: 42:28)



In case of the parameter classifier, we know this pdf, the probability of x given w_1 that is the likelihood we know and similarly, we know the probability of x given w_2 that is also available. So, from this we can take the classification decision. So, this is about the parametric classifier that is the density form is known.

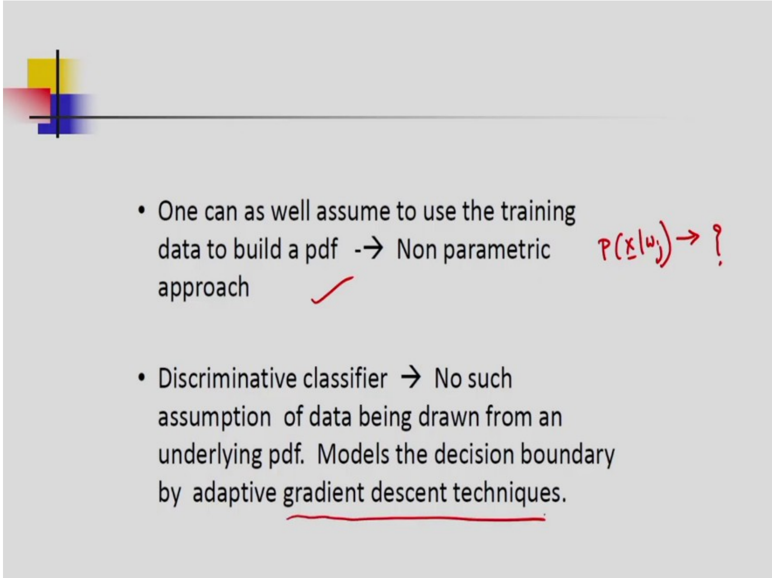
(Refer Slide Time: 42:57)



So, in this case you can see I am considering one generative classifier and you can see the two classes I am considering. One is the salmon, another one is a seabass two types of features I am considering and I am considering two features, one is the lightness of the fish, another one is the width of the fish and you can see the decision boundary between the classes.

So, this is my decision boundary, in case of generative classifier. So, already I have mentioned that one is the parameter approach, another one is the non-parameter. In parametric approach the density form is available but we have to estimate the parameters. In case of the non-parameter the density problem is not known, but we have to estimate the densities.

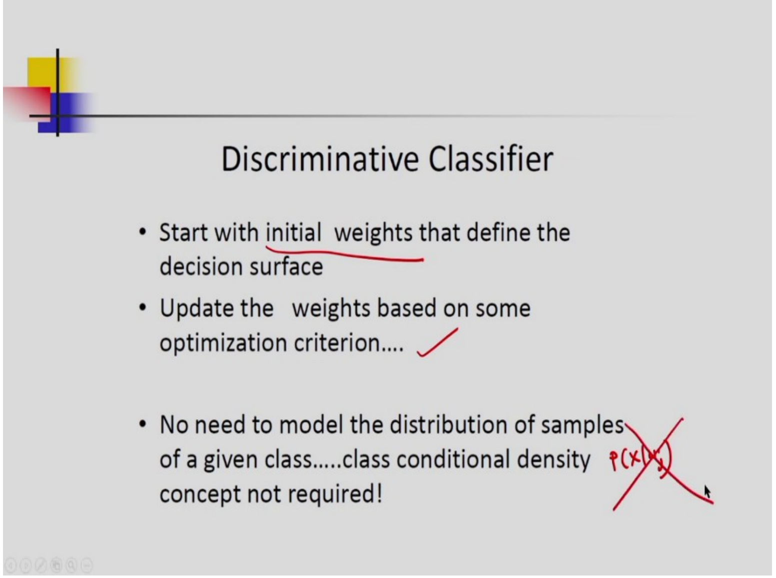
(Refer Slide Time: 43:43)



So, you can see here one is a nonparametric approach. So, that means, so from the training samples, we have to estimate the density of what, the density of this. So, we have to estimate this density that we have to determine, this is about the generative classifiers. In case of the discriminative classifier.

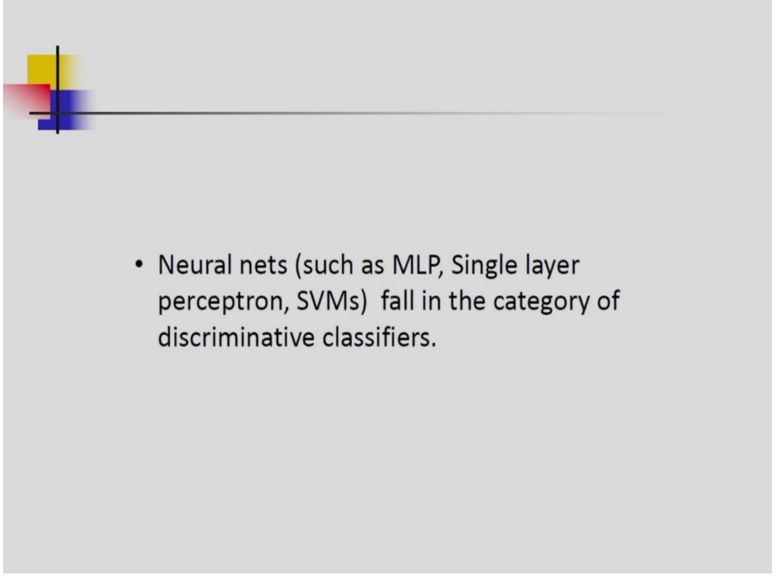
So, no such assumption of data being drawn from an underlying pdf that is not important, that is the class conditional density is not important, but what is important is that is that it models the decision boundary by adopting the gradient descent techniques. So, we can determine the decision boundary by considering one technique that is the gradient descent techniques. So, we can find a decision boundary.

(Refer Slide Time: 44:40)



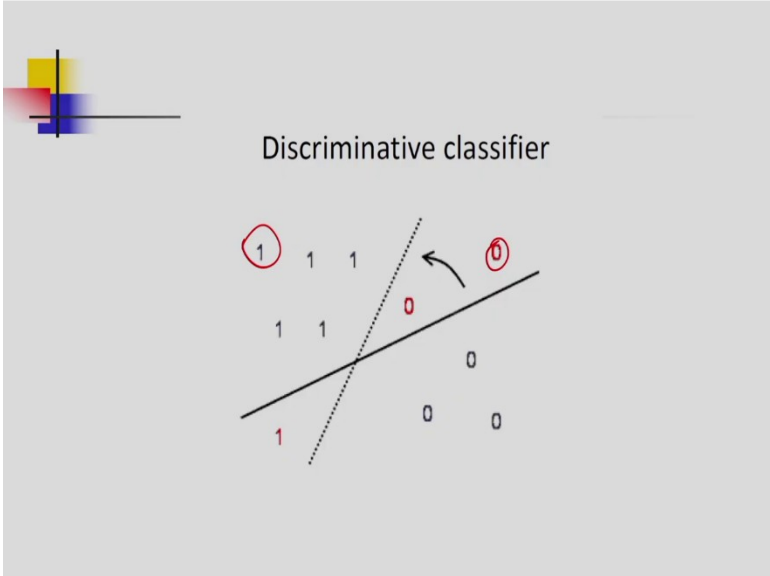
So, what is the discriminative classifier, start with initial weights that the define the decision surface. So, first I have to consider initial weights that define the decision surface. After this, I have to update the weights based on some optimization criterion that we can consider and in this case no need to model the distribution of samples of a given class, that is the class conditional density is not important, what is not important, so this is not important in this case a the discriminative classifier, so this is not important.

(Refer Slide Time: 45:18)



So, some examples of the discriminative classifiers are neural networks like multi-layer perception, single layer perceptions support vector machines are some examples of discriminative classifiers.

(Refer Slide Time: 45:34)

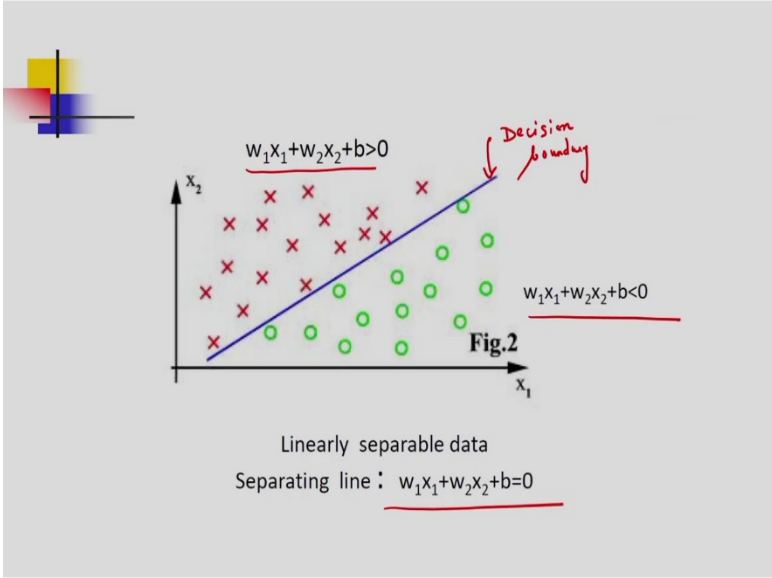


So, you can see first I am considering some initial weights and after this I am adjusting the weights, I am determining the weights to find the best possible decision boundary between the classes. So, in this example, I am showing two classes, one class is one and another class is 0. So, between these two classes, I want to find the decision boundary. So, first I am considering one initial decision boundary corresponding to this I have the width, but I have to adjust the width, so that I can find a best decision boundary between the classes.

(Refer Slide Time: 46:09)

And you can see in case of a discriminative classifier, we have linearly separable data and also, we have non-linearly separable data. So, in this example, you can see I can draw the decision boundary between the classes. So, I have two classes, one is the plus another one is the minus. So, between this I can easily draw the decision boundary and similarly, in the second figure also I have two classes and between this I can easily draw the decision boundary. So, that is why this is the linearly separable data.

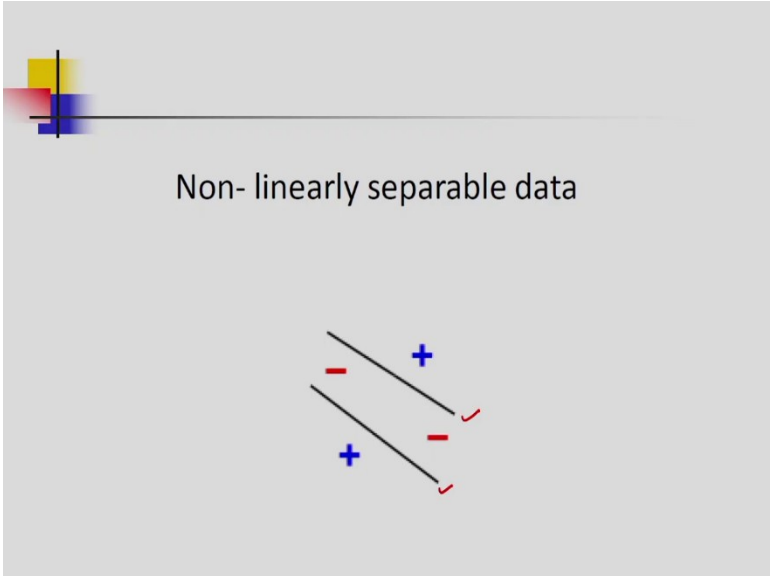
(Refer Slide Time: 46:45)



And corresponding to this linearly separable data, if I consider this one, this equation $w1x1+w2x2+b>0$, the b is the Bayes, w1, w2 are the weights and x1, x2 are the features, I am considering the two-dimensional feature space I am considering and you can see this is the decision boundary, decision boundary.

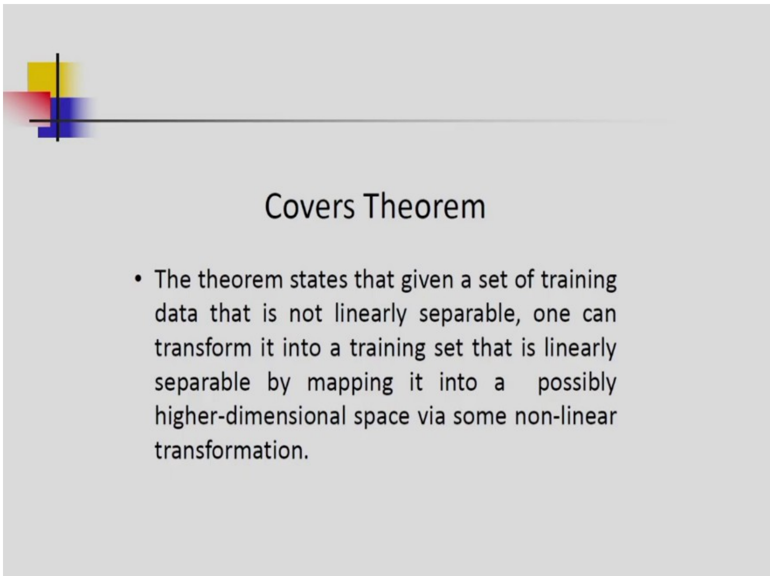
So, what is the equation of the separating line the equation of the separating line is $w1x1+w2x2+b=0$and if I consider this green class corresponding to this green class, the equation will be $w1x1+w2x2+b<0$that is the inequality and similarly, for the red class the equation is $w1x1+w2x2+b>0$. So, based on this formulation, I can distinguish these two classes, one is the red class and another one is the green class and also you can see the equation of the separating line between the classes.

(Refer Slide Time: 48:05)



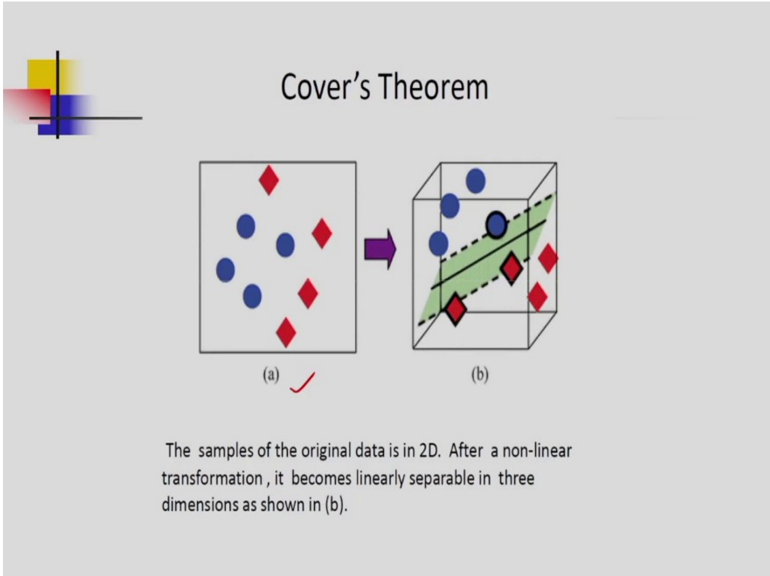
If you see this example, again I am considering two classes, one is the blue class and other one is the red class that is a plus and the minus and, in this case, it is very difficult to draw the decision boundary, the linear decision boundary between the classes. So, that is why I am considering one boundary like this, other boundaries like this. So, by using these two boundaries, I am separating these two classes. So, that is why I can say it is a non-linearly separable data. Previously I have shown the linearly separable data, but this is one example of non-linearly separable data.

(Refer Slide Time: 48:44)



And one important theorem is the covers theorem. So, what is the covers theorem? The theorem states that given a set of training data that is not linearly separable, one can transform it into a training set that is linearly separable by mapping it into a possible higher dimensional space via some nonlinear transformation. So, that the concept I am explaining in my next slide. So, what is the importance of the cover's theorem.

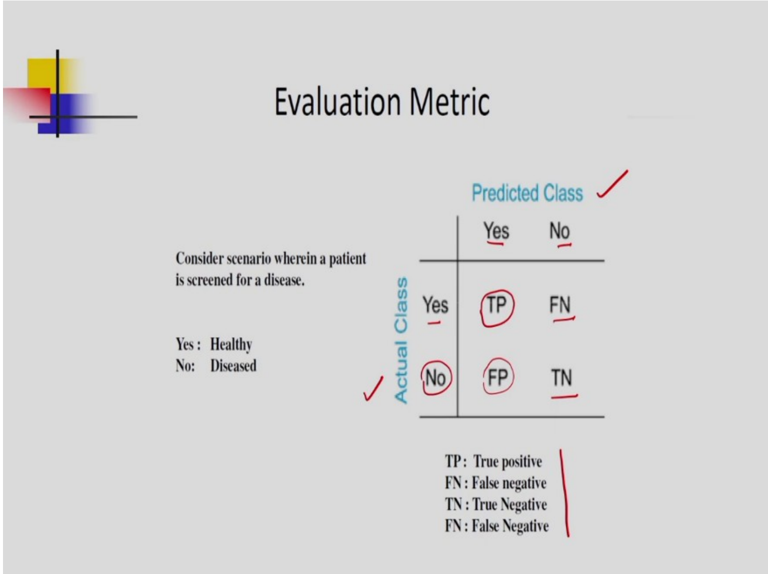
(Refer Slide Time: 49:16)



So, here you can see in the first figure, figure a I am considering some sample points of the original data that is in 2D in two dimensional which is not linearly separable. So, what we can do? I can apply some nonlinear transformation. So, nonlinear transformation is applied so that this original data can be mapped into high dimensional space.

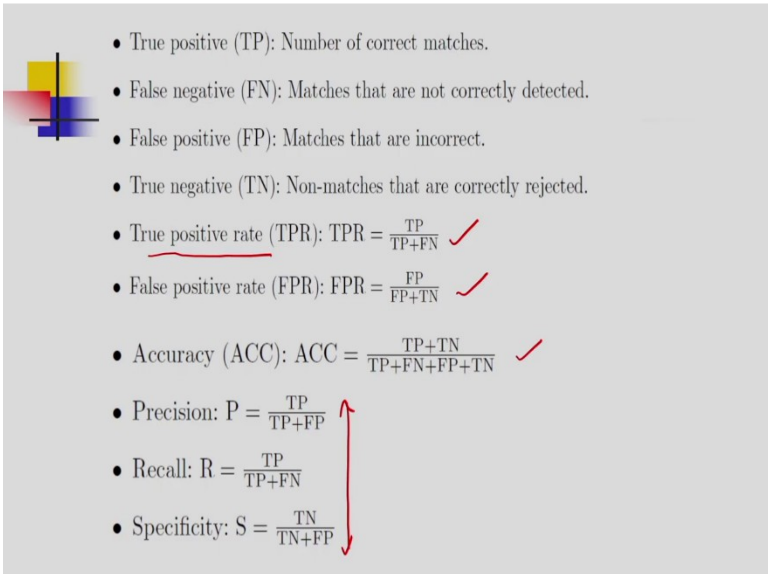
So, if you see the original data that is a two-dimensional space, but after this transformation, I am considering the three-dimensional space that is the high dimensional case I am considering. So, in high dimensional case, this samples will be linearly separable. So, that means in 2D it was not linearly separable, but in the 3D, it is linearly separable. So, that is the concept of the cover's theorem.

(Refer Slide Time: 50:17)



For pattern classification problem, I have to consider some evaluation metric. So, one is like this, suppose we have the predicted class and we have the actual class. So, actual classes as yes the predicted class is yes that means, it is that true positive. If the actual classes suppose no and the predicted classes yes, that is false positive. If the actual class is yes and the predicted class is no, then it is false negative and similarly, true negative I can determine. So, one is the actual class another one predicted class. So, these parameters one is the true positive and other one is the false negative, one is the true negative and the false negative I can determine from the actual class and the predicted class.

(Refer Slide Time: 51:11)



And these parameters are generally used in case of the pattern classification problem, one is the true positive, the false negative, false positive, true negative and the true positive rate also you can determine that is nothing but $\frac{TP}{TP+FN}$. So, that is the true positive rate, false positive rate also you can determine $\frac{FP}{FP+TN}$ that also you can determine.

Accuracy also you can determine $\frac{TP+TN}{TP+FN+FP+TN}$ that also you can determine and these are some important parameters, one is the precision, precision is nothing but $\frac{TP}{TP+FP}$ and also you can determine recall and the specificity also you can determine. So, these are the parameters.

(Refer Slide Time: 52:06)

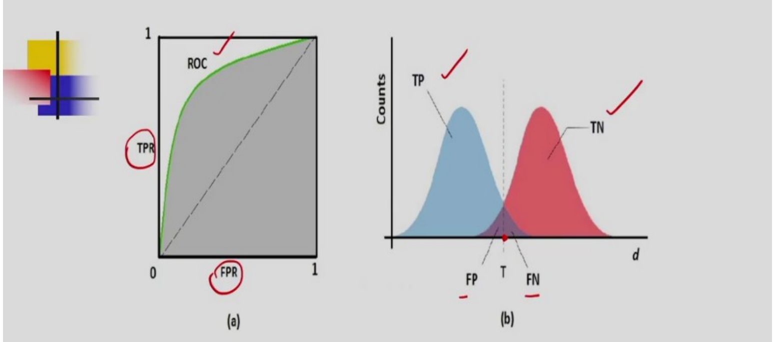
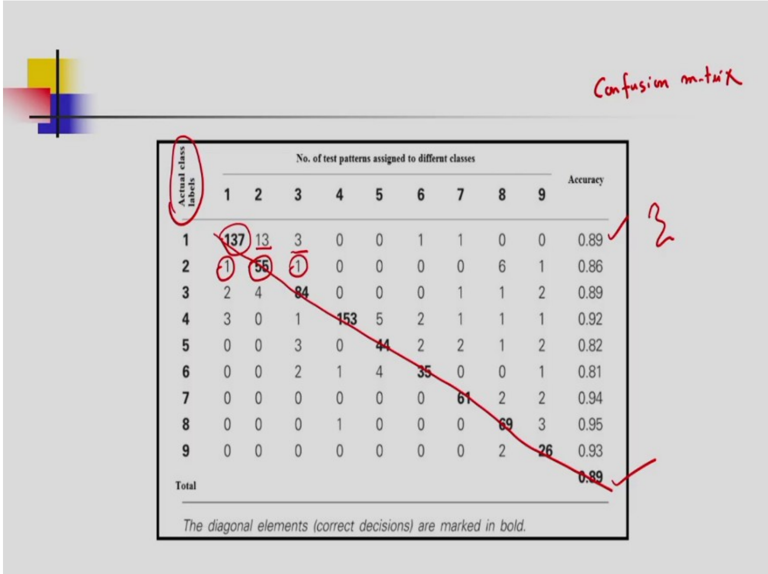


FIGURE : (a) The ROC curve showing TPR against FPR for a particular pattern matching algorithm (b) The distribution of positive and negative matches, which is shown as a function of inter-feature distance d.

M.K. Bhuyan, Computer Vision and Image Processing – Fundamentals and Applications, CRC press, USA, 2019.

And one thing is that you can determine, the ROC the region of convergence also you can determine, it is the you can see the ROC curve showing TPR against FPR for a particular pattern matching algorithm. So, I am showing the TPR that is a true positive rate and FPR I am showing and you can see the arrow see ROC, ROC you can see. So, you can see in the second figure what I am showing the distribution of positive and negative matches I am considering that means the true positives and true negatives I have shown which is shown as a function of inter feature distance the inter feature distance is d. So, you can see I am getting the FP and FN also there is a false positive and false negative and you can see threshold I am considering. So, based on this threshold, you can determine the true positive and the true negative.

(Refer Slide Time: 53:09)



Actual class labels	No. of test patterns assigned to differnt classes									Accuracy
	1	2	3	4	5	6	7	8	9	
1	**137**	13	3	0	0	1	1	0	0	0.89
2	1	**55**	1	0	0	0	0	6	1	0.86
3	2	4	**84**	0	0	0	1	1	2	0.89
4	3	0	1	**153**	5	2	1	1	1	0.92
5	0	0	3	0	**44**	2	2	1	2	0.82
6	0	0	2	1	4	**35**	0	0	1	0.81
7	0	0	0	0	0	0	**61**	2	2	0.94
8	0	0	0	1	0	0	0	**69**	3	0.95
9	0	0	0	0	0	0	0	2	**26**	0.93
Total										**0.89**

The diagonal elements (correct decisions) are marked in bold.

And finally, I want to show another important matrix that is called the confusion matrix confusion matrix. So, this is used to show the performance of a pattern classification algorithm. So, in this example, you can see, I am considering actual class labels and also you can see the number of test patterns assigned to different classes. So, what is the meaning of 137 that means, this suppose the class, suppose the pattern 1 is recognised as 1, how many times 137 times, the pattern is recognised, the pattern 1 is recognised as 2, 13 times the pattern 1 is recognised as 3 3 times.

So, like this, I can determine the confusion matrix and from this you can see from this you can determine the accuracy you can determine and also you can determine the misclassification rate also you can determine and also you can determine the rejection rate. Rejection rate means, how many times it is not recognised as 1, 2, 3, 4, 5, 6, 7, 8, 9. Suppose, I am drawing a character suppose, like this. So, it is not recognised as 1, 2, 3, 4, 5, 6, 7, 8, 9 So, that is why it is not recognised. So, from this I can determine their rejection rate.

Similarly, what is 55. That means two is recognised as two how many times 55 times, the 2 is recognised as 1, how many times 1 that is the misclassification, there is the misclassification and 2 is recognised as 3. So, how many times he does 1 time. So that is the misclassification and you can see from this information, I can determine the accuracy.

So, in the confusion matrix, I will be getting a diagonal matrix. So, I will be getting the diagonal matrix, if I get the high values in the diagonal matrix, that corresponds to the high accuracy of the pattern classification algorithm. In this class, I discussed the concept of pattern classification. I discussed the concept of supervised and unsupervised learning. After this briefly, I have highlighted the concept of semi supervised learning and the reinforcement learning. After this, I discussed the concept of generative classifier and the discriminative classifier. So, let me stop here today. Thank you.