Course Name: Machine Learning and Deep learning - Fundamentals and Applications

Professor Name: Prof. M. K. Bhuyan

Department Name: Electronics and Electrical Engineering

Institute Name: Indian Institute of Technology, Guwahati

Week-1

Lecture-1

Welcome to NPTEL MOOCs course on machine learning and deep learning fundamentals and applications. This is my first lecture that is the introductory lecture of the course on machine learning and deep learning fundamentals and applications. In this lecture mainly I will consider the introduction of machine learning. So, what is artificial intelligence, what is machine learning and what is deep learning and also I will be explaining the pattern classification, pattern recognition process. And also I will be explaining some learning techniques like supervised learning, unsupervised learning, semi-supervised learning. So, this is the introductory lecture to explain all these fundamental concepts. So, let me start this lecture. The lecture is the introduction to machine learning. So, here in this figure you can see I am showing that one is the artificial intelligence, one is machine learning and one is deep learning. So, what is the definition of artificial intelligence?

The programs with the ability to learn and regions like humans. So, this is the definition of AI artificial intelligence. And you can see here the machine learning is a subset of artificial intelligence. The algorithms with the ability to learn without being explicitly programmed.

So, that is the definition of machine learning. And mainly I will be considering statistical machine learning techniques. The deep learning is a subset of machine learning and it is nothing but the advanced version of artificial neural networks. So, there are some problems of the conventional artificial neural networks and these problems are addressed in deep learning techniques. So, deep learning techniques I will be explaining in my last classes, last modules of the course and mainly I will be explaining the machine learning concepts. So, this is the definition of artificial intelligence, machine learning and deep learning. And what is pattern recognition?

So, pattern recognition is a process of recognizing patterns by using machine learning algorithms.

And it is actually it is a data analysis system or I can say it is a data analysis method. And this pattern recognition is a derivative of machine learning that uses data analysis to recognize patterns. So, what is pattern and what is actually pattern classification? What is pattern recognition? I will be explaining in my next slide. So, regarding this machine perception.

So, here you can see build a machine that can recognize patterns. So, the patterns may be speech signal that is the speech recognition, may be the fingerprints that is the fingerprint identification or recognition, the optical character recognition, DNA sequence identification, biomedical images that is the biomedical image processing, biomedical signal processing. So, these are some examples of pattern recognition. So, for pattern recognition, we are employing the machine learning algorithms. So, considering this case, that is, I want to explain what is actually the pattern recognition.

I am giving one example, the example is computer vision. So, what is the definition of computer vision? Computer vision is a field of computer science that works on enabling computers to see, identify and process images in the same way that human vision does and then provide appropriate output. So, this is the standard definition of computer vision. And actually it is a complement of biological vision. So, for computer vision, you can see my input is images, or maybe the videos. So, to capture the images or to capture the videos, I need image acquisition devices like camera, I have to consider there may be a single camera or multiple cameras for image or video acquisition.

And after this, I am doing some pre processing that is nothing but the image processing. And finally, I can apply machine learning algorithms. So, that means the pattern recognition and artificial intelligence algorithms for decision making, that is the image recognition, the video recognition, object recognition. So, for this, I have to consider machine learning algorithms, that is the decision making. And if you see the second figure, that is the figure below the first figure.

So, that is actually the human visual system. So, if you see this structure, the human visual system, the system is very similar to computer vision. So, the input is again the images or maybe photos, because human has two eyes.

So, we can see objects, we can see images, we can see videos. And that is nothing but that image acquisition.

And after this, we do some processing in our brain. And finally, the intelligent decisions. So, that is by brain, the intelligent decisions. So, if you compare these two structures, one is the computer vision, another one is the human visual system, they are very similar. So, you can see the structure of the computer vision system is very similar to human visual system.

So, now, this can be shown like this, what is the computer vision, what is the image analysis. So, in the figure you can see, I am considering one input image. So, first I have to do image pre processing to improve the visual quality of the image. And maybe we can do segmentation. Segmentation means the separation of the foreground and the background.

And suppose the problem is detection of the tumors in the brain. So, this is the CT scan image. And so from this input image, I want to detect whether the tumor is available in the image or not. So, this is the brain image.

So, for this what we are doing, I am first doing the image pre processing, doing the segmentation.

After this, I am extracting some features representing this image. And I am extracting the feature vector. And finally, with the help of this feature vector, I am doing the classification. So, we can recognize these images, whether the tumor is available, or the tumor is not available, whether it is a malignant tumor, or it is a benign tumor. So, based on that these features, I can do all this classification, I can do the recognition.

So, this is a typical pattern classification system. So, human perception is similar to the machine perception. So, suppose the problem is, how did we learn the alphabet of the English language. So, because we can recognize all the English alphabets. So, for this, we are training ourselves to recognize alphabets, that means the training is going on.

So, because of this training, I can recognize alphabet. So, suppose the new alphabet is coming, and based on our intelligence, we can recognize it. So, that is about the human perception. Now, the machine perception is similar. So, how about providing such capability to machines to recognize alphabets.

So, the concept is very similar. So, first I have to go for training, training of the system, that is the machine learning system or the pattern classification system. And after the training, with the help of the train model, I can recognize alphabets. So, if you consider this machine perception and the human perception, they are very similar. So, the one essential step is that learning is the essential step.

And after the training or the learning, we have to go for testing. So, what is the definition of pattern. So, already I told you the pattern recognition is a process of recognizing patterns by using machine learning algorithms. So, the patterns, the definition is maybe an object, a pattern is an object, or maybe a process or event that can be given a name. So, maybe I can consider one electrical signal, the speech signal, fingerprint image.

So, these are the examples of patterns. And after this, what is a pattern class. So, a pattern class is a set of patterns sharing some common attributes, and usually originating from the same source. And that is the definition of the pattern class, the patterns sharing common attributes, that is the pattern class.

And during recognition or classification, given objects are assigned to the prescribed classes.

So, the this the pattern classes are represented like this omega 1. So, this is one class, suppose this is omega 2. So, these are the classes, the pattern classes omega 1, omega 2. So, I can represent like this. And during the recognition, given objects are assigned to the prescribed classes.

And for this, we can consider a classifier. The classifier is a machine, which performs recognition or classification. So, this is the definition of the pattern, the pattern class and recognition. So, these are the examples of applications, this optical character recognition, biometrics, like face recognition, fingerprint recognition, speech recognition.

And in case of the diagnostic systems that medical diagnostics, x ray imaging, ECG analysis, that is the biomedical signal processing, machine diagnostics.

So, these are some applications. And even in the military also, we have many applications like automated target recognition, image segmentation and analysis, recognition from aerial and satellite photographs. So, there are numerous applications of pattern classification. And these are some examples. So, what are the main approaches of pattern classification, the pattern recognition.

So, one is the statistical pattern recognition, and it is based on the statistical models.

So, we can employ some statistical models like the Bayes law we can employ that is the Bayesian decision theory. So, mainly the statistical models we will be considering for statistical pattern recognition. The structural pattern recognition means the pattern classes are represented by means of formal structures, such as grammars, automata, strings, etc. So, suppose, if I consider suppose the pattern is A is the pattern A. So, this pattern A, it is composed of A this, this structure, this structure and this structure.

So, the A can be represented like this. So, this is one representation. And that is nothing but the structural representation for pattern classification the pattern recognition. So, nowadays, this is not much of use. Mainly we will consider the statistical pattern recognition. And another one is the neural network, actually it is the soft computing based pattern recognition.

So, for the soft computing based pattern recognition, mainly we will consider the fuzzy logic, the artificial neural networks, maybe the genetic algorithms, genetic programming. So, there are many standard methods. And in my course, mainly I will be considering artificial neural networks and the fuzzy logic. So, these are the main approaches of pattern recognition. So, now, what is actually the pattern recognition I want to show you.

So, already I told you, actually it is a process of recognizing patterns by using machine learning algorithms. So, I can show a block diagram. So, suppose I have some patterns. So, these patterns, the information from the pattern that is acquired by sensor. So, I have some sensors and from the patterns, I am doing some measurements.

That means I am getting some information. And based on this, I am getting features, some features I am extracting that is nothing but feature generation. After this, another step is very important. So, I am extracting all the features, but all the features may not be useful for a particular pattern classification, pattern recognition problem. So, I have to select the most discriminative features that means most important features for a particular pattern recognition problem and that is called feature selection.

So, this step is called feature selection. So, I am selecting the important features. So, after feature selection, we are considering a classifier. So, we have to design a classifier and with the help of this classifier, I can recognize or I can classify patterns. So, that is the classifier design.

And after this, the system evaluation. So, for system evaluation, I am just doing the feedback with all the blocks. So, this is a feedback connection. Because of the feedback, I am evaluating the system and I am actually so this is a feedback and because of this feedback, I am actually changing the parameters of the system. So, this is the feedback I am considering to improve the performance of the overall system. So, we can improve the performance of the overall system because of this feedback, because we can evaluate the performance of the system and based on this, I can give the feedback to all the blocks of the pattern recognition system.

So, this is a typical pattern recognition system. So, the same thing I can show another way. So, suppose I have the patterns. So, I am doing some measurements by using the numbers. So, this is the measurement. So, after the measurement, I am getting the measured values.

So, after getting the measured value, that is actually that is the measured value means the feature extraction, I am doing the feature extraction. So, this is nothing but feature extraction. And after this, I am considering the feature values. So, that means it is nothing but the feature selection.

Because for a particular pattern recognition system, I have to select the most discriminative piezores.

So, all the piezores may not be important for a particular pattern recognition problem. So, that is why the feature selection is important. And finally, I am getting a feature vector, the feature vector. So, the feature vector is represented like this, x is a feature vector, I have this feature vector. And what is this feature vector x? So, I have this feature $x\ 1, x\ \ 2, \dots, x\ d$.

So, it is a d dimensional feature vector. So, I am considering all the feature $x\ 1, x\ 2\ up\ to\ x\ d$. So, it is a feature vector, x is a feature vector. And based on this feature vector, I can do the classification. So, suppose I have this feature vector, the feature vector is x and we are considering a classifier. And suppose we have a database and maybe some information and rules.

So, based on this information and the rules, I can do the classification. So, the information is taken from this database. So, I have some rules for the classification and we can do the classification. So, classify or maybe I can say decision making. So, there may be two types of decision makings, one is the hard decision making and another one is the soft decision making.

So, one is the hard decision making another one is a soft decision making. So, in the hard decision, we are considering discrete boundary, boundary means the decision boundary between the classes and we employ classical set theory. So, you can see in the hard decision, we consider discrete boundary between the classes and we employ the classical set theory. And in the soft decision, we consider the fuzzy logic.

So, these two decision making processes, I will be explaining later on.

But you can see, in one case, I am considering the classical set theory, that is nothing but the hard decision and what is the soft decision. In the soft decision, I am considering the fuzzy logic. In my next slide, briefly, I will be explaining what is hard decision and what is soft decision.

So, suppose I have two classes.

So, these are the samples corresponding to one class,

suppose the class is omega 1 and another class suppose.

So, these are samples belonging to another class. So, actually these are the feature vectors corresponding to another class, omega 2 and between these two classes, I have the decision boundary. So, this is the decision boundary. So, in case of the hard decision, you can see that this boundary is fixed. And you can see that all the samples belonging to the class omega 1 and all the samples belonging to the class omega 2.

There is no possibility that a particular may belong to another class.

Suppose in case of the fuzzy logic. Here, you can see some of the samples are near to the decision boundary. There may be some possibility that a particular sample may belong to another class. Suppose if I consider this sample, there is a possibility that this sample may belong to another class.

And similarly, if I consider this sample,

there is a possibility that this sample may belong to another class.

So that is actually the soft decision. So, for the soft decision, the decision boundary I can draw considering the same case. So, these are the samples corresponding to the first class and we have the samples corresponding to the second class. So, this is omega 2 and this is omega 1, two classes. And already I have shown the decision boundary between the classes.

So, this is the decision boundary. In case of this soft decision, the decision boundary is something like this. This is not rigid. So, this is the soft decision, soft decision boundary. So, in this case, there is a possibility that a particular sample, suppose this sample may belong to this class or there may be some possibility that this particular sample may belong to this class.

So, this possibility is determined by the membership grade. So, mu is the membership grade in fuzzy logic. So, the membership grade lies between 0 and 1. So, suppose the membership grade is suppose 0.9 for a particular sample.

So, that means, it is a high possibility that this particular sample may belong to another class.

So, based on this membership grade, we can take a decision. That means, there is a possibility that a particular sample may belong to another class also. So, that we are considering in case of the soft decision, decision making. In the hard decision, this case we are not considering. So, the first case is the hard decision and the second case is the soft decision.

That is the fuzzy logic we are considering. And one important point is, so how to consider feature selection. So, suppose we are extracting some features corresponding to two classes, suppose here the same example I am giving. So, two classes we are considering.

So, these are some features corresponding to two classes.

So, in this case, I can draw the decision boundary between the classes. So, easily I can draw the decision boundary between the classes. And suppose I am considering another example. So, these are the samples, samples means the feature vector corresponding to some classes. And these are also some samples belonging to another class.

So, in this case, in the second case, it is very difficult to draw the decision boundary.

So, it is very difficult to draw the decision boundary. So, maybe I can consider a decision boundary, maybe something like this. This type of decision boundary I have to consider. So, this is very difficult to draw the decision boundary.

So, that means the first one is the first is the example of good features.

So, the first is the example of the good features. And second is the example of the bad features. So, that is why we have to select the most discriminative features. And one feature should not affect another features.

And each feature should convey some information about the pattern.

So, that means x is a feature vector. So, this x 1 is a feature x 2 is a feature and x d is a feature. So, it is a d dimensional feature vector. So, this feature conveys some information about the pattern. Similarly, the x 2 also conveys some information about the pattern.

And this feature should not affect each other. So, that is why the feature selection is quite important in a pattern recognition problem, the feature selection is quite important. Now, I will explain the concept of pattern classification in more detail. So, move to the next slide. So, what is pattern classification? So, the pattern classification I can show the pattern classification is nothing but the information reduction or information mapping process.

So, for this what I am showing I am showing the classes suppose this is space we are considering this space is called the class membership space.

So, class membership space. So, we are considering suppose the classes are like this class omega i plus omega j and class omega k. After this I am showing another space that is the pattern space. Suppose this is the pattern space.

So, I am showing the patterns like this P 4, P 1, P 2 and suppose P 3.

So, all these patterns I am showing P 1, P 2, P 3, P 4. So, this is the pattern space. After this I am showing the measurement space. So, I am showing the measurements like M 1, M 2, M 3. So, these are the measurements and after this I am showing the mapping from the class membership space to the pattern space. So, suppose corresponding to the class omega i I have two patterns P 1 and P 4. Corresponding to the class omega j I have the pattern the pattern is P 2 and corresponding to the class omega k I have the pattern the pattern is P 3.

So, this is the mapping from the class membership space to the pattern space. So, I am repeating corresponding to the class omega i I have two patterns P 1 and P 4 corresponding to the class omega j I have the pattern P 2, corresponding to the class omega k I have the pattern the pattern is P 3. And corresponding to the pattern P 1, I have the measurement, the measurement is suppose M 1. Corresponding to the pattern P 2, the measurement is suppose M 2. And corresponding to the pattern P 3, suppose the measurement is M 1.

And corresponding to the pattern P 4, the measurement is M 3 suppose. So, you can see the mappings. So, I am first doing the mapping from the class membership space to the pattern space and from the pattern space I am mapping to the measurement space. So, that is why this pattern classification system or the pattern recognition system is nothing but the information reduction or the information mapping process.

And here you can see these patterns are overlapping,

you can see I am showing the overlapping you can see the overlapping here, overlapping here.

That means the patterns of different classes may share some common attributes. So, that is why it is overlapping. And here you can see what is the problem of pattern classification. The problem of pattern classification is it is actually the invert mapping from the measurement I have to determine the corresponding class. So, that means it is the invert mapping, invert mapping is from the measurement I have to determine the class. So, I have the measurements, measurements are M 1, M 2, M 3 these are the measurements

and from the measurement I have to determine the corresponding class.

And this is not one to one mapping. So, it is this is not one to one mapping. This is not one to one mapping. So, had it been one to one mapping, the pattern classification problem would have been very very easy. But it is not one to one mapping. So, what is the pattern classification from the measurement,

I have to determine the corresponding class and that is the invert mapping.

So, invert mapping from the measurement space to the class membership space. So, this is the definition of pattern classification, or I can say this is the definition of pattern recognition. So, statistically, this can be written like this. So, I have to determine the probability of omega I omega is a class X is the Feature vector. So, that is the objective of the statistical machine learning.

So, we have to determine the probability of obtaining a particular class given the Feature vector,

X is the Feature vector.

So, we have to determine this probability and this is the posterior probability. So, I will be explaining this one later on. But what is the statistical pattern classification or pattern recognition. So, I have to determine the probability of obtaining a particular class given the Feature vector that is the probability of omega I given X.

So, that is the definition of the statistical machine learning.

And we may consider the supervised learning and unsupervised. So, supervised means for each and every classes I have the training data samples. So, that I can show like this. Suppose I have the class omega I for omega I have a training data set D I for the class omega j I have the training data set D j for the class omega k I have the training data set the training data set is D k.

So, here you can see I have independent I have independent training data set for each and every classes.

And that is nothing but the supervised learning supervised learning. So, with the help of this training data set data set D i D j D k I have to train the classifier and after the training with the help of the train classifier I have to do the classification. That means I have to consider the testing data and this testing data can be recognized or testing data can be classified with the help of the supervised learned classifier. So, that means first I have to do the training with the help of this training data set and these are the independent training data set for each and every classes. And that is nothing but the supervised learning. In case of the unsupervised learning we have the Feature vector the Feature vector is x and we have to group the Feature vector based on some similarity.

So, the after the grouping I will be getting the clusters like this is one clusters and this is maybe another clusters the cluster of the Feature vectors the cluster of the samples. So, this cluster may belong to one class and this cluster may belong to another class omega 1 and omega 2. So, in the unsupervised learning we have the Feature vector the Feature vectors is x and we have to group the Feature vectors based on some similarity. So, maybe we can consider some distance measure and with the help of the distance measure I can determine the similarity between the Feature vectors. And after this I can do the grouping of the Feature vectors and after the grouping I will be getting the clusters the clusters corresponding to a particular class omega 1 and cluster corresponding to the another class like this I can do the grouping.

So, you can see the distinction between the supervised learning and unsupervised learning and I want to again explain that supervised learning. So, you can see here I have the independent training data set that means the training data set D I it is not for that class omega J that is not for that class omega J. The training data set D I is only for the class omega I the training data set D I is not for the class omega J and this training data set D I actually this is a training data set D I. So in the training data set I have all the samples x 1 x 2 these are the samples up to x suppose up to suppose x n.

So n number of samples are available. So these are the samples of the data set and with the help of this training data set I can train the classifier and after the training with the help of the train classifier I can do the recognition I can do the classification. So this is the fundamental concept of the supervised learning and unsupervised learning. Okay now I will explain the concept of discriminate function. So briefly I will explain what is discriminate function and based on the discriminate function I can take a classification decision.

So move to the next slide. So what is actually the discriminate function. Discriminant function so that is represented by g I x and this discriminate function is used to partition R to the power d space that is the d dimensional space. So it is used to partition R to the power d space the d dimensional space and that is the feature space and we are considering c number of classes y is equal to 1 2 up to c number of classes we are considering. So for c number of classes I have c number of discriminate functions.

So now what is the decision rule. The decision rule I can consider like this decision rule. So I have to assign I have to assign the feature vector x to the class the class is suppose omega m this is a class if some condition the condition is if the discriminate function g m x is greater than g I x. So based on the discriminate function I can take a classification decision and this is for all the classes I is equal to 1 2 so c and I is not equal to m. So based on this I can assign the feature vector to the class the class is omega m.

So this is the discriminate function. So I will be explaining later on in my next classes. So what is the discriminate function. So first I have to explain the Bayesian decision theory and after explaining the Bayesian decision theory I can define the discriminate function. So for the time being just you can understand that this is the discriminate function g I x is the discriminate function and I have c number of discriminate function and based on this discriminate function I can take a classification decision. So you can see here suppose if I consider this is a feature space.

So it is a two dimensional feature space and I am showing a decision boundary between the classes. So this is a decision boundary between the classes the region is R1 another region is R2 and this is nothing but this is a decision boundary. So this region R1 corresponds to the class the class is omega 1 and region R2 corresponds to the class the class is suppose omega 2. So in this case what is the equation of the decision boundary the equation of the decision boundary is g1 is equal to g2 that is g1 x is equal to g2 x. So the equation of the decision boundary is gk x, x is a feature vector and suppose g m x.

So this is the equation of the decision boundary. So in my next classes I will be discussing about the nature of the decision boundary it may be a linear decision boundary or maybe if I consider something like the nonlinear decision boundary also we can consider and suppose if I consider high dimensional feature space then I can consider something like that hyper plane I can consider or maybe the circle I can consider the ellipse I can consider. So there are many types of decision boundaries. So I will be explaining all these concepts when I will discuss the concept of the Bayesian decision theory.

So this is the equation of the decision boundary. So up to this point now I am explaining I am not going into detail. So this is the concept of the discriminant function and based on the discriminant function I can take a classification decision and the discriminant function is something like this. This is a linear discriminant function gi x is equal to this is W i t transpose x plus W o i. So this is the form of the discriminant function this is a linear discriminant function linear discriminant linear discriminant function. So this expression also I will get later on. So this is the weight vector so this is the d cross 1 vector that is the weight vector actually that is used for the class I can say it is a weight vector weight vector for the class for the class omega i and this W o i that is actually it is the bias.

So this is the bias. So this is the one example so this is the this is one example of linear discriminant function this expression I will get later on but for the time being you can see I have the weight vector the weight vector is W i t that is for the class omega i and x is the feature vector and I have a bias the bias is W naught i. So this is the concept of the discriminant function. Now move to the next slide. So what are the components of a pattern recognition system? So already I have explained in my previous slide. So you can see here I have the patterns and for taking the measurement I am considering sensors and I am also doing the pre-processing and based on this step actually I am just doing the measurements and I am extracting the features that is nothing but the feature extraction and after feature extraction I have to consider feature selection because already I told you so all the features may not be useful for a particular pattern classification problem I have to select the most discriminative features and finally I am considering the classifier based on the feature vector.

So it is the learning algorithm and we can consider the supervised learning techniques. So these are the components of a typical pattern recognition system. So move to the next slide. So here I am showing one example how to do the classification and the problem is jockey and the Hoopster recognition. So I have two states one is h, h means the Hoopster and j means jockey. So that means I have these two states y is equal to h or maybe j and in this case I am considering a two dimensional feature space because I am considering two features x1 and x2, x1 corresponds to height and x2 corresponds to width.

So x1 corresponds height, height of the person and x2 corresponds to the weight of the person. So it is the problem is the person recognition based on these two features one is the height another one is the width. So the feature vector x has two components one is x1 another one is x2 and based on this feature vector I can do the classification. So you can see I have the training samples like this x1 so corresponding to x1 the output is y1 corresponding to x2 output is y2. So I have the training samples and based on this training samples I can do the training and you can see here in the figure this is the decision boundary I have shown the decision boundary between the classes. So some of the red points I have shown that belongs to one class so that class is y is equal to j and the blue points if you see the blue points these are the sample points corresponding to the class y is equal to h.

And you can see the equation of the decision boundary is W dot x that is the dot product plus B, B is the bias is equal to 0. So that is the equation of the decision boundary and based on this linear equation I can take a classification decision.

So $W.x + B \geq 0$ then the class will be h and if the $W.x + B \leq 0$ then the class is j. So this is a linear classifier and based on this condition the condition is W dot x plus B greater than equal to 0 my output is h that is the output means the class is h and if W dot x plus B less than 0 the class is j. So this is one example of a linear classifier and

I have shown the decision boundary between the classes.

And already I have explained that what is what do you mean by good features and what is the meaning of the bad features. So in case of the good features it is easy to draw the decision boundary between the classes and in case of the bad features it is very difficult to draw the decision boundary between the classes. So that is why I have to consider the good features so that is why the feature selection is quite important. And in this case I am showing the concept of the classifier. So a classifier partitions feature space x into class level regions.

So here in the figure you can see I am considering the region x and sub regions are like this sub regions are x1, x2, x3 like this. So if I take the union of this x1 union of x2 union of x3 so

I will be getting the total space feature space the feature space is x. So here you can see I am taking the union of this and if I consider the intersection of this all the subsections that will be the null set it will be equal to 0. So in this case corresponding to x1 the region x1 I have the class the class is omega 1. Similarly corresponding to the class x2 corresponding to the region x2 my class is omega 2 and corresponding to the region x3 my class is omega 3.

And you can see the decision boundary between these classes these are linear decision boundaries. In the second figure also I am considering these classes x1, x2, x3 these are the regions corresponding to the class omega 1, omega 2 and omega 3. And here you can see the decision boundary is not linear it is a non-linear decision boundary. So this is one example of the partitioning of the feature space the feature space is x. So move to the next slide and already I have explained that is based on the discriminate function I can take a classification decision.

So gi x is the discriminate function and for c number of classes I have c number of discriminate function. So this classifier assigns a feature vector x to a particular class if the discriminate function gi x is greater than gj x. So that means based on the value of the discriminate function I can take a classification decision. So here in this figure you can see my input is the feature vector and I am determining the discriminate function. So I am determining the discriminate function g1 x, g2 x, gc x for c number of classes

I have c number of discriminate function.

And out of all this c number of discriminate function I am selecting the maximum one which one is the maximum I am selecting. And based on this I am taking the classification decision. Suppose g2 is maximum that means the corresponding class is omega 2. So you can see how we can do the recognition with the help of the discriminate function. And for taking classification decision in case of the Bayesian decision theory we can consider the Risks.

So we have to minimize the Risk while taking a classification decision. So we have to minimize the Risk for taking a particular decision. And for this we have to determine the classification error. The classification error we have to determine and that is actually the Risk. So the Risk can be calculated based on the classification error or the probability of error. So I will be explaining this concept what is Risks and what is the probability of error and how you can take a classification decision with the help of this parameter the parameter is the Risks.

Now some of the terminologies like class levels omega we are considering. So omega 1 is a class omega 2 is a class suppose I am considering two classes omega 1 is for sea bass fish that is a kind of fish and omega 2 and that is also another class for the fish salmon. And we are considering the prior probability the probability of omega 1 and probability of omega 2 these are the prior probabilities. So prior information is available what is the probability of obtaining sea bass that is the probability of omega 1 what is the probability of obtaining the salmon so that is probability of omega 2. And also we have the evidence so that point I will be explaining later on what is the evidence.

So these are some important terminologies these are some important terminologies. So this is the Bayes theorem so here you can see I can determine the probability of omega j given x and that is the posterior probability that is equal to probability of x given omega j that is called the likelihood or I can say it is a class conditional density into probability of omega j that is the prior probability divided by evidence. So evidence I can represent like this. So this is a simple Bayes rule and with this rule I can determine the probability of omega j given x that means what is the probability of obtaining a particular class given the Feature vector the Feature vector is x. So I have to determine this probability and based on this probability I can take a classification decision. So for example I can consider or I can decide the class omega 1 if probability of omega 1 given x is greater than probability of omega 2 given x otherwise I have to decide omega 2.

And similarly you can see if I want to decide omega 1 so based on the likelihood that probability of x given omega 1 into probability of omega 1 is greater than probability of x given omega 2 into probability of omega 2 then I have to select the class omega 1 otherwise I have to decide omega 2. And based on the likelihood ratio also you can see this is the ratio that is probability of x given omega 1 divided by probability of x given omega 2 that is the likelihood ratio. So if the likelihood ratio is greater than a particular threshold then I can select a particular class the class is omega 1 otherwise if it is less than the threshold I have to select the class omega 2. So all these concepts the concept of the Bayesian decision theory I will be explaining in my next classes. So in a pattern recognition framework so already I told you if I consider the supervised learning so I have to consider training samples testing samples also I have to consider and after this first what I have to do I have to do the training with the help of the training samples and after the training we can do the classification we can do the recognition with the help of the train model and that is nothing but the supervised learning.

So first one is the learning the training with the help of the training samples and after this I have to do the testing with the help of the train model. So typical supervised pattern recognition problem suppose the problem is the recognition of 26 alphabets uppercase. So that means I have to collect samples corresponding to 26 alphabets and we have to employ these training samples for training the algorithm. Algorithm means the machine learning algorithm.

So once it is trained the machine learning algorithm is trained so we have to test the system with the help of the testing samples. This is the procedure for supervised classification. So already I have defined the patterns so maybe the pattern maybe the fingerprint image and written word human face speech signal DNA sequence alphabet.

So these are some examples of patterns.

This is one example handwriting recognition and this is one example of face recognition. So face recognition is a very popular problem in computer vision. So you can see I can recognize the different faces with the help of pattern recognition or the machine learning algorithms. And another application is the fingerprint recognition. So for all these cases I first I have to extract features and based on these features I have to do the classification I have to do the recognition.

The space of machine learning methods. So what are the popular machine learning methods the supervised and unsupervised. So this I am giving the highlight of these methods. So here you can see here I am showing the machine learning method that is the supervised method. So I have labeled data and you can see the machine learning algorithm is there.

So with the help of the label data that is the supervised learning. So for each and every classes I have the label data. So I have to train the algorithm the machine learning algorithms and after this I will be getting the learned model and after this the testing that is available. So with the help of the testing data I can do the prediction. So one is the training and

after this the prediction also I can do.

This is one example of a supervised learning system. And here I am showing the types of learning the first one is the supervised. So already I have explained what is supervised. So you can see in the example the first example that is the text recognition and I can consider two classes the class A and class B. This is a classification problem. So I have to select I have to identify which one is the class the class A or class B.

Unsupervised is nothing but the clustering that I have shown here is a clustering the grouping of the feature vectors based on some similarity. And I am showing the concept of the regression. So what is the regression the fitting of a line or fitting of a curve between the sample points. Here you can see I am showing some of the sample points and I am fitting the best fit line I am considering.

So considering all the sample points I am finding the best fit line. And another learning is the reinforcement learning. So deeply I can explain what is the reinforcement learning. So in this case actually one particular action is not important.

The group of actions are more important. Suppose I can give one example the playing of chess game. So one particular move is not so important. The group of the group of moves are more important. And what is the goal. The goal is the winning of the game.

So corresponding to one good action I will be getting a reward and my action is not good then I will be getting the penalty. So it will be penalized. So main concept is one particular action is not so important. The group of actions are more important and we have to we have to fulfill the goal. So in this example of the chess the goal is winning the game.

So one bad action may affect the performance but that is not measured. The group of actions are more measured and corresponding to a good action I will be getting a reward. So briefly I can explain like this. This is the reinforcement learning. So these are some machine learning popular methods supervised and unsupervised and reinforcement.

So I will not be discussing about the reinforcement learning. So mainly I will be discussing the supervised learning and unsupervised learning. So in the supervised I will be discussing the concept of regression, decision tree and random forest. These are very popular algorithms and for classification we can discuss KNN, K nearest neighbor, trees, logistic regression and Naive-Bayes classifier, support vector machine. So this concept I will be going I am I will be going to discuss and regarding the clustering mainly we will consider PCA, K means clustering.

So all these algorithms I will be discussing. So these are some popular machine learning techniques. So for a pattern recognition algorithms so that is the bag of algorithms that can be used to provide some intelligence to a machine. So that is actually the machine learning. So I am giving some intelligence to a machine and these algorithms have a solid probabilistic framework. So already I told you so mainly in case of the machine learning we will be considering statistical machine learning techniques.

So all the statistical mathematical tools or statistical mathematical concepts we will be using in developing the machine learning algorithms. And algorithms work on certain characteristics defining a class referred as features. So already I have explained so corresponding to the classes that classes can be identified or classes can be recognized based on the features. So in this case what is feature suppose if I want to recognize this letter the letter is I one is the capital I another one is the small i.

So what will be the feature the base feature this is a pattern recognition problem. So I have to identify which one is the capital I and which one is the small i. So the feature I can consider here move to the next slide the presence of a dot in I can distinguish the small i from the capital I. So presence of a dot in small i can distinguish small i from capital I. And this feature value may be discrete or continuous in nature and in practice a single feature may not be sufficient for discrimination.

So that means I have to consider more and more features for a particular pattern classification problem. So one feature may not give accuracy the correct accuracy. So that is why I have to consider more number of features. So now problem is this recognition of two types of pieces one is Sea bass another one is salmon. So suppose I am considering this problem.

So now for this I have to do some measurements and in this case we are considering some sensing device. So to see these pieces one is Sea bass another one is salmon. So it is a two class problem. So Sea bass means I am considering it is omega 1 and salmon is the omega 2. So two class problem and I have to consider how to do the classification.

So take simple images of the pieces.

Now we can consider some features the length of the piece the lightness of the piece the width of the piece number and shape of the pins position of the mouth. So these are the some features and with the help of these features I have to design or I have to decide and based on these features I have to take a classification decision. So first I have to do some pre processing. So maybe first in case of the image processing I have to do the segmentation. So that is isolating piece from one another and also from the background in case of the background I have to do some image processing operations like the segmentation also we can do that is segmentation is nothing but partitioning of an image into connected to homogeneous region.

So I can separate program from the background also. So this is the pre-processing is important and after this after pre-processing we have to extract features. So here you can see this is the image of the piece and I am doing the pre-processing I can improve the visual quality of the image and also we can do the segmentation we can separate program from the background the background is not important piece is more pieces important after this we have to extract features feature extraction and with the help of these features we can do the classification and the classification output is two types of pieces one is salmon another one is sea bus. So now I am considering only one piece so the piece is the length of the piece. So select the length of the piece as a piece and here you can see I am showing the counting the that how many times the salmon is counted and how many times the sea bus is counted. So here you can see if I consider only this piece at the length of the feature there are some misclassifications you can see the salmon is recognized as sea bass and sea bus is recognized as salmon.

So that is shown by the red one that sea bass is recognized as salmon and salmon is recognized as sea bus. So this is because of considering only one feature it has no discriminatory power so that is why I have to consider maybe another feature maybe two features we can consider so let us see if I consider two features whether we can reduce the classification error. Now in this case another feature we are considering now not two features we are considering another feature and that feature is the lightness of the fish. So if I consider this as a possible feature then what you can see in this case also the misclassification is there sea bass is recognizes salmon and salmon is recognized as sea bass. So the previously we considered the length as a feature the length is not giving good performance misclassification is taking place.

So we are considering the second feature the second feature is the lightness. So with the help of this lightness feature alone the performance is slightly improved but still the misclassification is taking place. So the threshold decision boundary and cost relationship. So you can see I have to reduce the misclassification.

So reduce number of sea bass classified as salmon. So how to do this how to reduce the classification error that is the task of decision theory. So now what we are considering now I am considering a feature vector the feature vector is X it has two features X1 and X2. So X1 is the lightness of the fish and X2 is the width of the fish. So previously we only considered only one feature but now we are considering two features the X1 and X2.

So let us see the performance. So you can see here I am considering the two dimensional feature space. So one feature is the lightness another one is the width of the fish. And you can see the decision boundary between these two classes one is the sea bass and another one is the salmon. So this is the decision boundary. So you can see some of the samples are misclassified.

So like this suppose this sample that is misclassified this sample it is misclassified. And similarly this sample this black sample this is misclassified this sample is misclassified. That means sometimes the sea bass is recognized as salmon and sometimes the salmon is recognized as sea bass. But with the help of these two features you can see the misclassification reduces the misclassification rate reduces. Previously we only considered only one feature but this time we are considering two features and because of this the misclassification rate the misclassification error it is reduced.

So that means we can consider another features to reduce the error. And one important thing is if I consider noisy features that may not improve the performance of the classifier. So it may reduce the performance of the classifier. So that is why I have to select the most discriminative features. That means I have to select the good features and that is the importance of the feature selection. So that if I consider good features and based on these features I can determine the decision boundary and I can get the best decision boundary between the classes.

So that provides an optimal performance. And here you can see this is the decision boundary I am getting and that is the optimal performance I am getting and you can see the misclassification in this case it is almost zero no misclassification. So I am getting a perfect model during the training. Previously if you see in the previous case I am getting this decision boundary and in this case of this decision boundary that I have the misclassifications. This is a simple model we are considering and with the help of this simple model I am getting this decision boundary and this is a linear decision boundary and I have some misclassifications. But in this case I am considering a complex model and with the help of the complex model and with the help of the training samples I am getting a nonlinear decision boundary and this is one of the best decision boundary I am obtaining during the training.

So that means objective the dealing with the novel data what is the objective dealing with novel data novel data means the new data. So first we are doing the training with the help of the training data and after this we have to see the performance on the new data the novel data whether our the train model is good or not that we have to observe we have to see. So in this case I am getting the decision boundary like this. So I am considering a very simple model and corresponding to this model I am getting the decision boundary like this and with the help of this decision boundary you can see I can do the classification and there may be some misclassification the misclassifications are like this these are the misclassified samples. So these are the misclassified sample. So I can consider this type of models so I can consider that these type of models the first case you can see I am considering a very simple model and in this case if I consider a very simple model it is not sufficient to represent the training samples because I am considering the very simple model this is a linear model.

So this is the first case and that corresponds to the under fitting the model is very simple and that corresponds to high bias. So during the training the error is very high you can see the error is very high and during the testing also the error is very high. So that means this model that this model is very simple and it cannot perfectly represent the training data or the testing data. So move to the next case that means the third case. So I am getting the best decision boundary misclassification is zero with the help of the training data set.

So with the help of the training data set I am considering a very complex model and with this with this complex model I am getting this decision boundary. So this is a very complex decision boundary nonlinear decision boundary and you can see during the training the misclassification is almost zero but during the testing that is the unseen data. So that means during the testing with the help of the unseen data the error will be high because it cannot handle unseen data because we are considering a very complex model and with the help of this complex model I can represent perfectly the training samples but in the testing case it may not be true and that corresponds to the case of overfitting. So that is nothing but the high variance. So the case number one that is the case number one is the first case one and case number two second this is not good.

So case number one corresponds to here that is high bias and case number two corresponds to this case that is the overfitting. So now I am considering a relatively simple model not a very simple model in case of the third one this is a third one and with the help of this relatively simple model I am getting this decision boundary and it is obtained with the help of the training samples. So during the training I have some classification error so that you can see this is a classification error and during the testing also I have classification error but that is the good compromise. So during the testing and during the training and this error is reasonable. So that means the third model I should consider that is the good compromise. So in the first case it corresponds to the high bias condition for the training and for the testing it is not good because this model the simple model cannot represent all the testing samples and all that all the training samples.

In the second case if I consider a very complex model it can perfectly represent the training samples but that may not be good for the unseen test samples. So that case number two is also not good but case number three that is the good compromise that is good because the error is reasonable during the training and error is reasonable during the testing also. So that this concept of the bias and the variance I will be discussing in my next classes this concept is very important in this class actually I am just introducing the concept of overfitting. So move to the next slide. So in a pattern recognition system so these are the components first I have to do the measurement so maybe we can consider some sensing devices using some transducers like camera microphone so these sensors we can consider and after this we have to do some pre-processings segmentation also we have to do we have to do the grouping.

So the pattern should be well separated so this is about the segmentation in the grouping. So in this block diagram you can see first the input is there input patterns and I am doing the measurement with the help of these sensing devices and we are doing the segmentation and after the segmentation we are doing the Feature extraction after Feature extraction we can consider Feature selection because all the features may not be useful for a particular pattern classification problem after this we are considering the classification system that is the classifier and after this the post-processing is nothing but system evaluation and you can see the feedback here in with all the systems I have the feedback that is for the system evaluation and based on this feedback I can improve the performance of all the steps. So this Feature extraction is quite important because we have to select the discriminative features and the features should be invariant with respect to affine transformations like translation, rotation, scaling even it should be invariant to automatic variations. So we have to select the invariant features and after this with the help of this feature vector we have to do the classification. So the feature extraction and classification we have to consider and finally the post-processing we have to see that means we have to see the performance of the system as a whole. So that means whether the Feature extraction is good whether the Feature selection step is we are considering there is the best feature selection method and whether the classifier is a good classifier for a particular pattern classification problem.

So all these things we have to we have to observe. So the design cycle is we have to collect collect data after this the Feature selection is important and after this how to select a particular model for a particular application. So it depends on the application so we have to select the model the classifier we have to select and after this we have to train the model training and we have to go for the evaluation and one important point is the computational complexity of the system that also we have to see. So for many real-time pattern recognition or machine learning systems we have to consider that issue that is the computational complexity issue we have to consider. So the system should be computationally simple same thing here I am also showing the start so collection of data selection of the features selection of the appropriate model train the classifier and evaluate the performance of the classifier. So this is the design cycle so data collection data collection already I have explained so data I need for training and data I need for testing and Feature selection is important already I have explained that concept and the Feature should be invariant to different transformations like affine transformation translation rotation or the photometric variations and it should be insensitive to noises.

Model selection so we have to select the best model for a particular application so we have to select the model based on the performance. So based on the performance we can select the model and after selecting the model we have to go for the training because if I consider the supervised learning system so first we have to go for the training and after the training we have to go for evaluation system evaluation and for system evaluation we have to consider the testing samples and we have to see the error rate we can measure the measurement of the error rate and based on this we can see the performance of the classifier or maybe the pattern recognition system as a whole and already I told you that computational complexity so the system should be computationally simple and the learning and the adaptation supervised learning already I have explained and unsupervised learning is nothing but natural groupings so that is nothing but the clustering. So unsupervised learning the concept already I have explained so that is nothing but the grouping and we have the Feature vectors and we can find a similarity between these Feature vectors based on some distance measured. So maybe we can consider some Euclidean distance measured and we can see the similarity between the Features and based on this similarity we can do the grouping. So the clustering is mainly the concept is we have to see the high interclass similarity and low interclass similarity.

So these two points are important based on this consideration I can do the grouping so one is the high interclass similarity and another one is the low interclass similarity. So like this here you can see I am finding the similarity between these images between these letters between the fingerprints and the similarity may be some distance measure we can consider and we can find a similarity between these patterns. So just I am showing the distance this is not normalized 0.

23, 3, 3, 42, 0.7 so these are not normalized just I am explaining that means I am finding the similarity between these patterns. So here I am showing briefly what is the supervised and what is the unsupervised. So in the first figure you can see I am considering the supervised learning and these are my classes that that not that not that and first I have to do the training. So I am getting the predictive model and with the help of this predictive model I can do the classification. So whether it is that or not that that we can consider so that is about the supervised learning. In the unsupervised you can see this the grouping I can do so this is one group this is another group and this is another group the grouping can be done with the help of the measure the measure is the similarity measure.

So based on the similarity we can do that grouping. So this is another example of the grouping so we have the training data set and you can see I am just doing the grouping based on similarity. So this point just I want to explain the clustering is subjective. So suppose we are considering these are the objects so suppose we are considering these this is the problem so if I do the grouping one is the group of the family one is the group of the school employees one is a group of females one is the group of males. So that means I can say the clustering is subjective. So this figure I am showing to explain that concept that the concept is the clustering is subjective and the concept of the reinforcement learning already I have introduced and that concept is nothing but corresponding to a particular good action I will be getting a reward and one particular action is not so important the group of actions are more important. So briefly the concept is like this so in this course I am not explaining this concept reinforcement learning and semi-supervised learning is quite important in many cases suppose I do not have level data so that means I have small amount of level data and the large amount of unlabeled data.

So then we can consider the semi-supervised learning. So one example is suppose in case of the medical imaging or medical image classification the problem is it is very difficult to get the level data. So maybe we have the small amount of level data and the large amount of unlabeled data. So then what we can consider this approach we can consider that is the semi-supervised learning approach we can consider. So actually it falls between the unsupervised learning and the supervised learning.

So this example already I have given that is the example of the medical image classification. In many cases it is very difficult to find the large amount of level data. So if I have small number of or small amount of level data and with large amount of unlabeled data then we have to consider this approach the approach is the semi-supervised learning and the concept of the regression already I have explained. So it is nothing but the fitting of a line or fitting of a curve between the sample points. So this concept also I will be explaining in my next classes the concept of regression. So for the time being just you should understand it is a problem of line fitting it is a problem of curve fitting and is a best fit curve we have to determine the best fit line we have to determine and all the mathematical things I will be explaining in my next classes that is the concept of the regression and this is the classifier.

So you can see here I am showing two classes one class corresponding to that is red samples and another one is the green samples and you can see the decision boundary between the classes. So empirical risk minimization so in this case what we can consider we have to minimize risk for taking a particular decision. So that is I have to minimize the risk or maybe the loss function we can define and we have to minimize the risk for taking a particular classification decision. So that is actually the risk minimization algorithm. So that concept also I will be explaining in this course. The no free lunch theorem it is impossible to get nothing for something that means the meaning is that one cannot hope for a classifier that would perform best for all the possible problems that one can imagine that means you cannot see that that one particular the classifier you cannot say this is the best classifier.

So it depends on the application. So for a particular application one classifier may be very very good but for another application that may not be good. So you cannot get one universal classifier which can do the classification the best classification for all these applications and that is the concept of the no free lunch theorem. So classifier taxonomy so we have generative classifiers another one is the discriminative classifiers. In case of the generative classifiers we can consider the parametric and the non parametric. So what is actually the this generative classifiers. So if you see the Bayes theorem so we have to determine the probability of omega J given X that is the posterior probability and that is equal to this likelihood the likelihood is also called a class conditional density and P W J that is the prior probability and this is the evidence the evidence has no role in classification.

So we are not considering the evidence because it is a simply the normalizing factor. So this is the Bayes theorem that we have to determine that this probability of omega J given X and based on this we can take a classification decision. So for determining this I need this information that is the class conditional density class conditional density. So this information I need so if I have this information then I can easily do the classification. So because I can determine probability of omega J given X if I know probability of X given omega J and probability of omega J.

So in this case if the density form is known because it is a class conditional density the density form is known but I do not know the values of the parameters then it is called the parametric estimation. So suppose if I consider is a Gaussian density or Gaussian distribution in case of the Gaussian density I have two parameters one is the mean vector another one is the covariance matrix or maybe in one dimension is a mean and another one is the variance. So two parameters so density form is known but I do not know the values of the parameters. So we can determine these values of the parameters with the help of some techniques. So in my next classes I will be discussing the parametric methods and mainly I will discuss the method likes the maximum likelihood ML estimation the maximum likelihood estimation and also another very popular technique that is called the Bayesian estimation.

So this is called the parameter estimation the parametric method density from is known but I do not know what are the values of the parameters. So already I have given the example in case of the Gaussian density the density is Gaussian but the what are the parameters the mean vector and the covariance matrix. So I have to determine the mean vector and I have to determine the covariance matrix. So I can consider these two popular algorithms one is the maximum likelihood estimation and another one is a Bayesian estimation. So in case of the non-parametric estimation this density from is not known the density is not known that is the this class conditional density that is not information is not available.

So we have to determine the density so directly we can determine the density that is the density of probability of omega j given x directly we can estimate the density. That means directly estimate the density probability of omega j given x directly you can estimate this. So for this I will be explaining two very good methods in my next classes one is the Parzen window technique and another one is the k nearest neighbor KNN method. So these two popular algorithms I will be explaining so how to estimate the density. This is about the generative classifier so in the generative classifier we need the information of the class conditional density.

So this concept already I have explained that is the samples of the training data of a class assumed to come from a probability density function. So this is the concept already I have explained and based on this we can consider a parametric classifier and you can see I have shown the plot of the class conditional density the probability of x given omega 1 and probability of x given omega 2 these are the class conditional density and based on this we can take a classification decision. And in case of the generative classifier this information is not required we have to find the best decision boundary between the class. So here you can see I am showing two classes and I am showing the decision boundary between the classes.

The objective is to find the best decision boundary between the classes. So how to get the decision boundary the best decision boundary that is the concept of the generative classifier. So in case of the discriminative classifier so we need not consider the class conditional density that is the PDF that is the PDF from that is the class conditional density that information is not important but we can determine the best decision boundary by considering some algorithms. So one algorithm may be something like the gradient descent algorithms or maybe in this case we can consider the artificial neural networks these are some example of the discriminative classifier. So examples like support vector machine the artificial neural networks so the objective is we have to find the best decision boundary between the classes.

In this case the concept of the class conditional density that is not important. So this is the discriminative classifier start with the initial weights that define the decision surface. So we are initially consider one decision surface and after this we are updating the weights based on some optimization criterion that means we are finding the best decision boundary based on this optimization method and we do not need the information the information is the class conditional density that is not required. So we have to update the weights and based on this we have to select the best decision boundary between the classes. So for example like the neural networks MLP means the multi-layer perceptron single layer perceptron support vector machine. So these are some examples of discriminative classifiers.

So you can see I am considering two classes the samples are shown like this 1 1 1 0 0 0 and you can see so initially my decision boundary is like this I am just moving the decision boundary so that the misclassification will be less. So that means just I am moving the decision boundary so corresponding to the first position you can see I have misclassifications. So these two samples are misclassified these two red samples are misclassified they are classified as one but in the second position this dotted position the classification it is less the classification error is less. So that is why we have to find the best decision boundary between the class and that is the objective of the discriminative classifier. Here you can see I am showing one example of linearly separable data.

So you can see two classes we are showing and these are linearly separable data and I am showing the decision boundary between the classes and you can see the equation of the decision boundary is $W1\,X1 + W2\,X2 + B$. So W1 W2 these are the weights we are considering and B is the bias. So we have to change the value of the weight during the training so that I will be getting the best decision boundary between the classes and based on the first condition if the $W1\,X1 + W2\,X2 + B > 0$ then I will be getting or I will be considering the one class that is the rate class and if I consider the second condition $W1\,X1 + W2\,X2 + B <$

0 I will be getting the second class. By adjusting the weights of this equation the equation of the separating line I can determine the best decision boundary between the classes. So here I am showing the example of nonlinearly separable data.

So you can see I am showing two lines that is I can consider as decision boundary and these are nonlinearly separable data and corresponding to this I am discussing one theorem that is the Covers theorem. So what is the Covers theorem? Suppose that given a set of training data that is not linearly separable one can transform it to a training set that is linearly separable by mapping it to a possibly higher dimensional space by considering some nonlinear transformation. So that means initially the samples are not linearly separable. So what I have to do I have to map it into the high dimensional space.

So I am mapping it to high dimensional space and because of this mapping data will be linearly separable and this mapping I can consider some nonlinear transformations. So initially the training samples are not linearly separable. So what we are doing we are doing some mapping that is the mapping from low dimensional space to high dimensional space by considering nonlinear transformations and because of this transformation into the high dimensional space the training samples will be linearly separable. So that I can so in the here so in the first figure you can see that training samples are not linearly separable and I'm doing the mapping into the high dimensional space. So initially 2 dimensional space and I am just doing the mapping into the 3 dimensional space. So in the 3 dimensional space they are linearly separable in the 2 dimensional space it is they are not they are not linearly separable but in the high dimensional space that is in the 3 dimensional space they are linearly separable.

So this is the concept of the cover's theorem and to evaluate the performance of a classifier we can consider some parameters or the metric like we can consider true positive false negative true negative false negative. So these are the parameters we can consider so suppose we have actual classes then the predicted classes is there actual classes yes the predicted class is yes that is nothing but true positive if the actual class is yes the predicted class is no that is nothing but false negative. If the actual class is no and the predicted class is yes that is the false positive if the actual class is no and the predicted class is no that is the true negative. So you can consider these parameters the true positive false negative true negative and the false negative we can determine. And based on this we can define some of the parameters like accuracy precision recall specificity so all these things I will be explaining in my next class.

So how to evaluate the performance of a classifier in this class I briefly explain the concept of machine learning and the pattern classification and I have shown a typical pattern recognition system. So what are the important steps like measurement, feature extraction, feature selection, classification, system evaluation. So all the steps I have explained and also I have explained the concept of the supervised learning unsupervised learning and also the semi-supervised learning. This is the only introductory class to explain the concept of machine learning. So let me stop here today. Thank you.