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 30

Introduction to Machine Learning

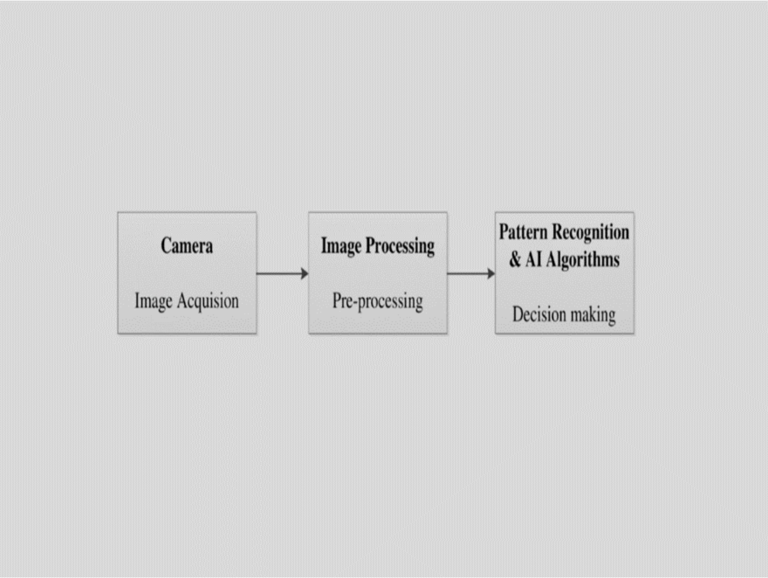
Welcome to NPTEL MOOCS course on Computer Vision and Image Processing- Fundamentals and Applications. I have been discussing about image features after image features, the next step is image classification, image recognition. So, for this I have to apply pattern classification or pattern recognition algorithms, the machine learning algorithms I have to apply for object recognition, object classification.

So, now I will discuss some fundamental concepts of Machine Learning, pattern classification and it is not possible to discuss all the machine learning concepts in this computer vision course. So, that is why I will discuss the briefly some important concepts of machine learning and after this I will discuss some of the applications of computer vision.

So, for pattern classification there are mainly three approaches one is a statistical approach, one is the structural approach and one is the sub computing base approach. In statistical approach, we will consider some statistical methods and based on this we can do pattern classification, pattern recognition. Another one is just structural method that is nothing but description of a particular pattern and finally, the soft computing-based pattern recognition in this case we can use fuzzy logic, artificial neural networks, genetic algorithms and genetic programming.

So, I will mainly discussed only the statistical pattern recognition techniques and also the soft computing base pattern recognition techniques, but all the concepts I will be discussing in brief. So, let us see what is pattern recognition.

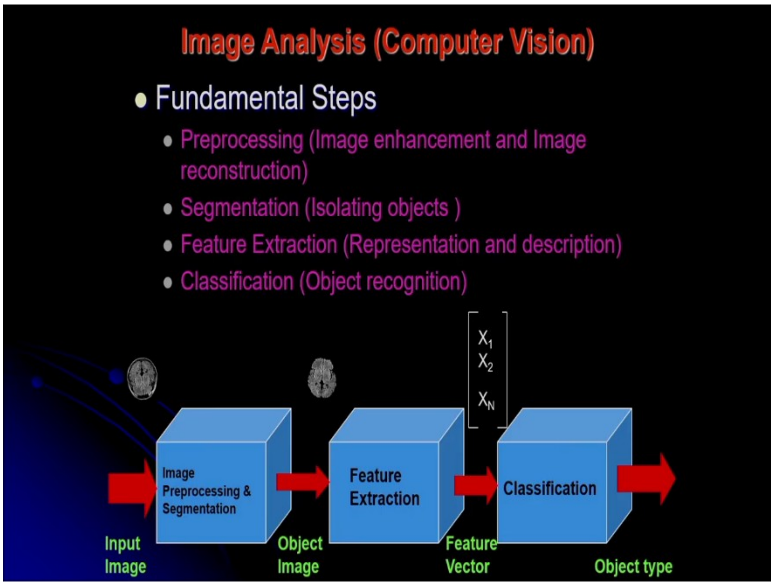
(Refer Slide Time: 02:22)



So, in my first class I discussed about this block diagram, this is a typical computer vision system. So, first one is the image equation by camera and after this we have to do some pre-processing and final step is the feature extraction and the pattern classification that is nothing but decision making.

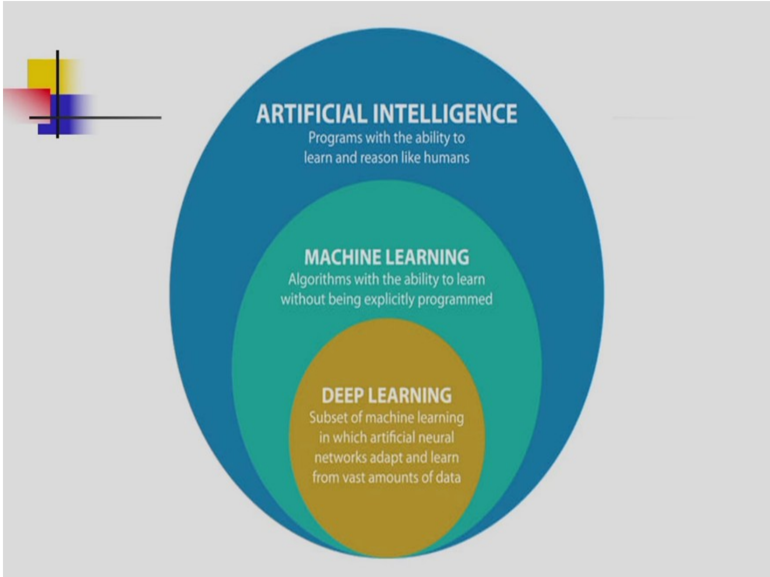
So, pattern recognition and artificial intelligence algorithms. So, already I have discussed about image features, so different feature like colour feature or texture features, a shape and a boundaries like this we have discussed. So, now I will discuss some pattern recognition concepts which are mainly used in computer vision applications.

(Refer Slide Time: 03:06)



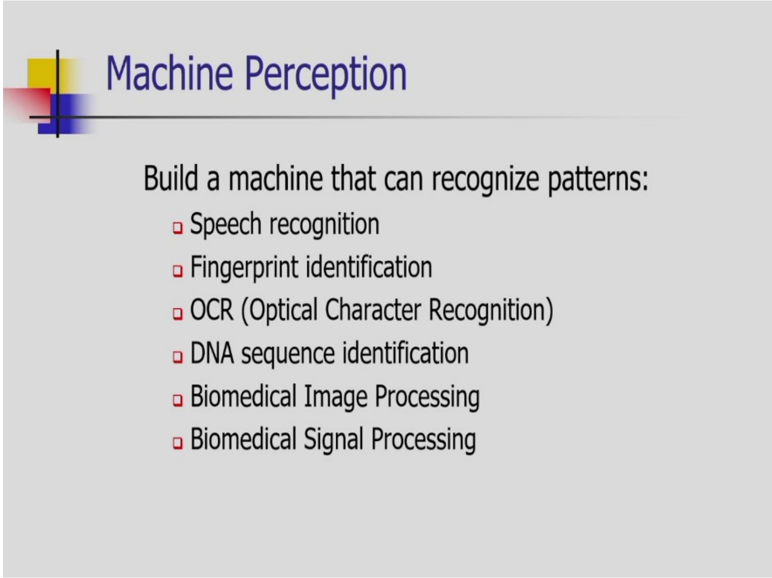
So, this block diagram also I have shown in my second class and that is the image analysis system you can see input images are available and after this we have to do some pre-processing, after pre-processing we have to do feature extractions. So, here you can see, I have shown the feature vector x_1, x_2,… x_n these are the elements of the feature vector and after this we have to go for classification that is the object classification, object recognition.

(Refer Slide Time: 03:39)



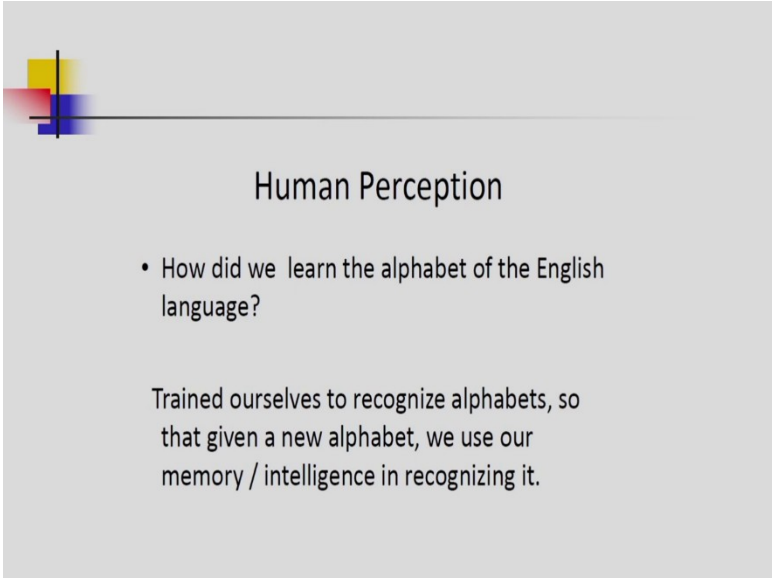
And you can see the domain of the machine learning. The first one is the artificial intelligence, the programs with the ability to learn and reason like humans and that is the artificial intelligence and you can see the machine learning is a subset of artificial intelligence. So, algorithms with the ability to learn without being explicitly programmed, that is machine learning and if you see the deep learning, the deep learning is a subset of machine learning and in this case, we use the artificial neural networks and also the deep networks. So, this is about machine learning and the deep learning.

(Refer Slide Time: 04:20)



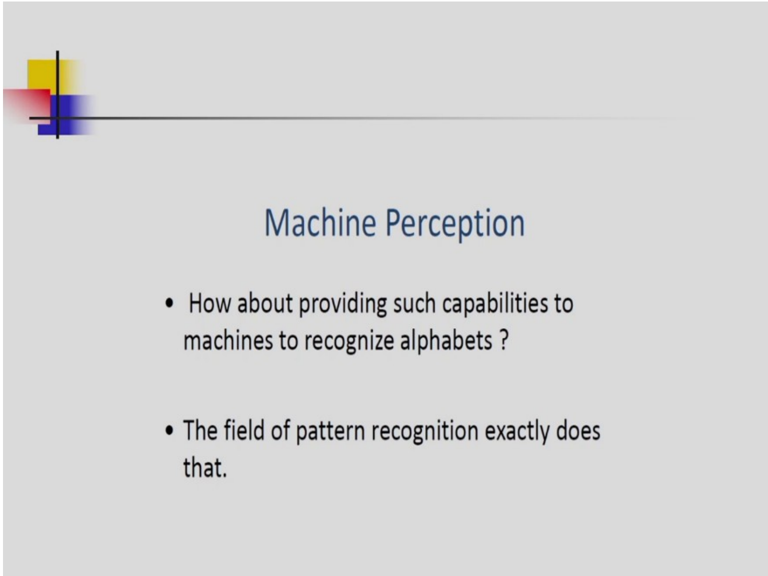
And in this case, there are some applications of pattern recognition, machine learning and in this case, I have shown that machine perception for speech recognition, fingerprint identification, optical character recognition, a DNA sequence identification, biomedical image processing and biomedical signal processing. So, build a machine that can recognise different patterns. So, these are the examples.

(Refer Slide Time: 04:48)



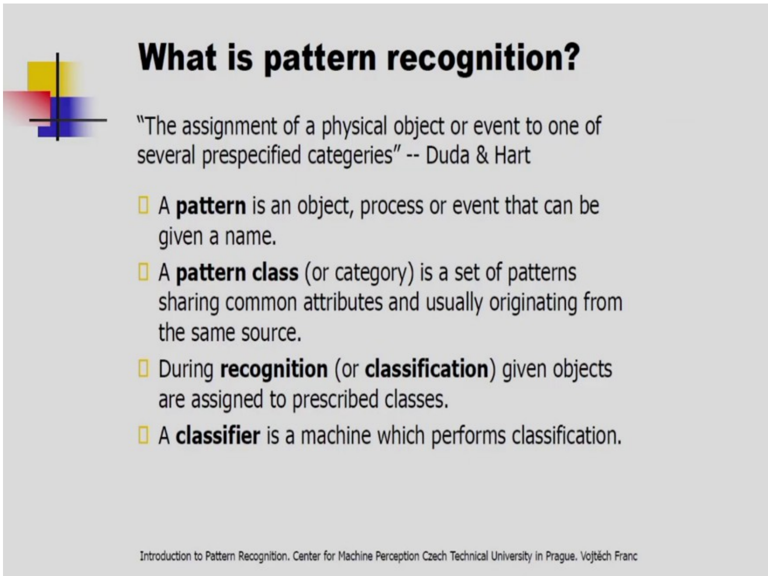
So, I can give one example of human perception. So, we can recognise all the alphabets of the English language. So, for this we train ourselves to recognise a different alphabet and whenever suppose the new alphabet is coming or it is available, we can recognise based on this learning and in this case, we have the memory and the intelligence in recognising and the alphabets. This is one example of human perception.

(Refer Slide Time: 05:19)



The next one is the machine perception. So, how about providing such capabilities to machines to recognise alphabets. So, in this case also we have to do training the machine and after this the machine can recognise different alphabets. So, first thing is the training and after this the testing.

(Refer Slide Time: 05:43)

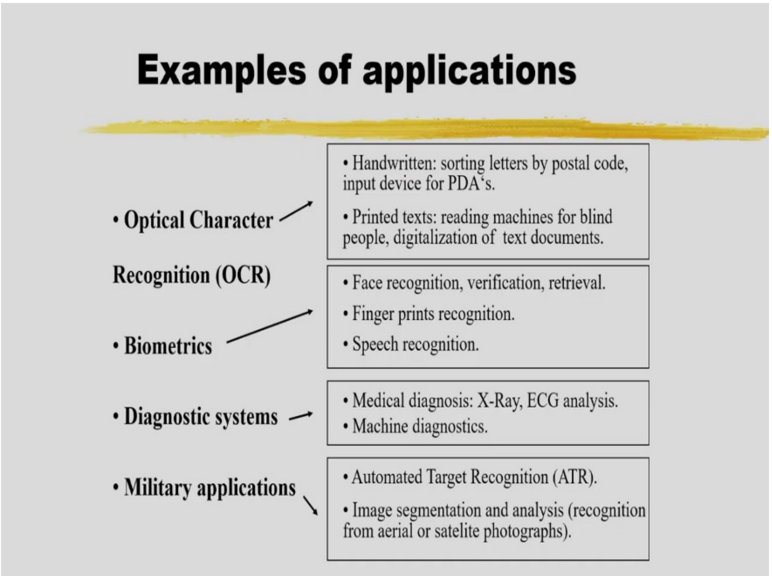


So, for pattern recognition, I have to define some terms the first one is the pattern. So, I can give one example of the pattern, the pattern maybe signal or maybe object. So, if you see the definition of the pattern, so a pattern is an object process or event that can be given a name and after this next one is a pattern class.

So, a pattern class is a set of patterns, sharing common attributes and usually originating from the same source. So, suppose the fingerprint is a pattern and suppose if I consider the fingerprint of school employees, then that is the class, the pattern class. So, my pattern will be the fingerprint and if I consider a pattern class that is the fingerprints of the school employee, particular school employee, if I consider that is the pattern class.

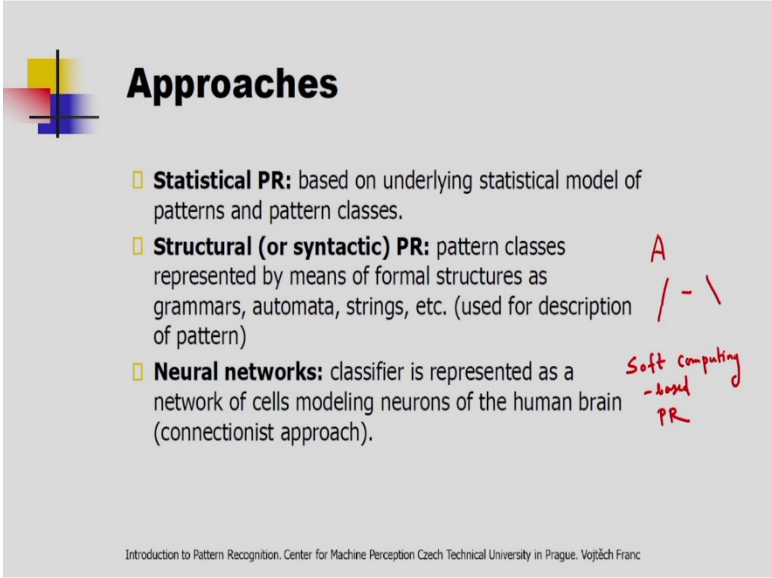
After it is the recognition. So, during recognition or classification, given objects are assigned to prescribe classes and for this we consider classifier. So, this is about the pattern classification. So, first I have to understand the pattern and after these the classes, the pattern classes, after this what is the recognition and the classification and what is classifier.

(Refer Slide Time: 07:05)



Some of the pattern classification examples are like optical character recognitions, even in the biometrics and the face recognition, fingerprint recognition, speech recognition, machine learning is used and for medical diagnoses, the X ray imaging, ECG analysis there are many applications of machine learning, pattern recognition and also applications in military.

(Refer Slide Time: 07:35)

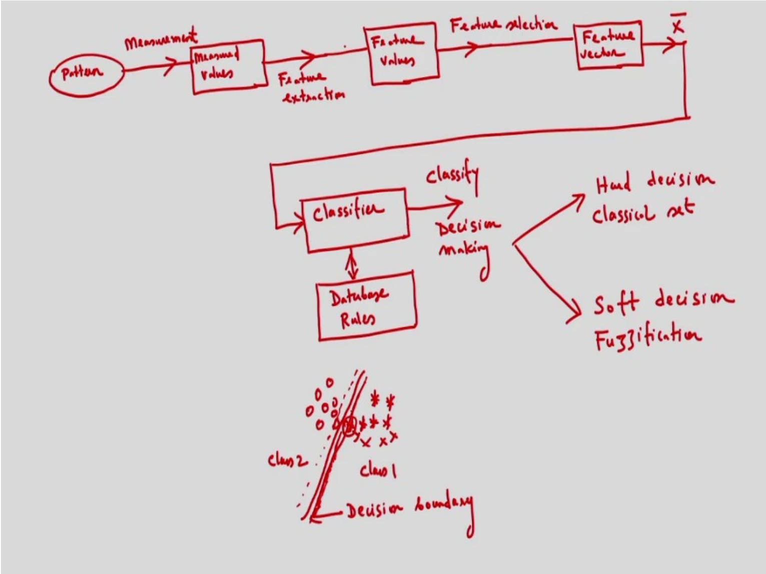


So, the approaches are statistical pattern recognition, so it is based on underlying statistical model of patterns and the pattern classes. So, statistical techniques will be used for statistical pattern recognition systems. So, mainly I will discuss about statistical pattern recognition systems, the next approach is the structural approach, it is nothing but the description of a pattern.

So, suppose if I considered a pattern A, so, A I can consider like this A has three primitives, if you see the A has three primitives. So, by using these structural components, I can represent the pattern A, this is about the structural pattern recognition system, nowadays it is not much of use. So, that is why we will discuss a statistical pattern recognition and also the sub computing base pattern recognition. The next approach is the sub computing base, here I have shown the neural networks, but it is the sub computing base pattern recognition, sub computing base pattern recognition.

So, for this we can fuzzy logic, artificial neural networks, genetic algorithms and the genetic programming. So, these are the approaches of pattern recognition. Now, I will show one typical block diagram of a pattern recognition system.

(Refer Slide Time: 09:10)



So, for this you can see I have suppose patterns, after this we have to do some measurement and we have the measured values. After this we have to extract features, so after feature extraction, we will be getting the feature values. So, all the features are not important. So, that is why we are considering feature selection and after feature selection, we will be getting, the feature vector we will be getting.

So, feature vector is suppose x, x is feature vector. So, after getting the feature vector what I can consider, I can consider a classifier. So, input to the classifier is the feature vector and suppose we have a database and suppose some rules are available and after this, what is the output? Output is the, that classify that is nothing but that decision making.

So, if you see this typical block diagram of the pattern recognition system. So, we have the input patterns and after this we have to do some measurements and I will be getting the measured below that is nothing but the feature extraction. So, I will be getting the feature values, after this all the features are not important.

So, what we are considering, we are selecting some of the important features that is the discriminative features we are selecting that is the feature selection and after this we are getting the feature vector, the feature vector is x that is input to the classifier. So, you can see the classifier and we have the database, suppose that is the training samples are available and some rules are available and based on this we can do pattern classification, we can do decision makings and these decision makings may be the hard decisions or maybe the soft decisions.

In case of the hard decision, we use classical set theory and in case of the soft decision we considered a fuzzification, that is a fuzzy set theory is used. So, we consider fuzzification. So, one may be hard decision, another one is the soft decision. In case of the hard decision, we use classical set theory and in case of soft decision we consider fuzzification, that is nothing but the fuzzy logic we can consider.

So, this is a typical block diagram of a pattern recognition system and in this case, what is the hard decision and what is the soft decision. Suppose if I consider two classes, these are the classes suppose, the samples of a particular class and this is another class, the samples of another class and suppose this is a decision boundary between the classes.

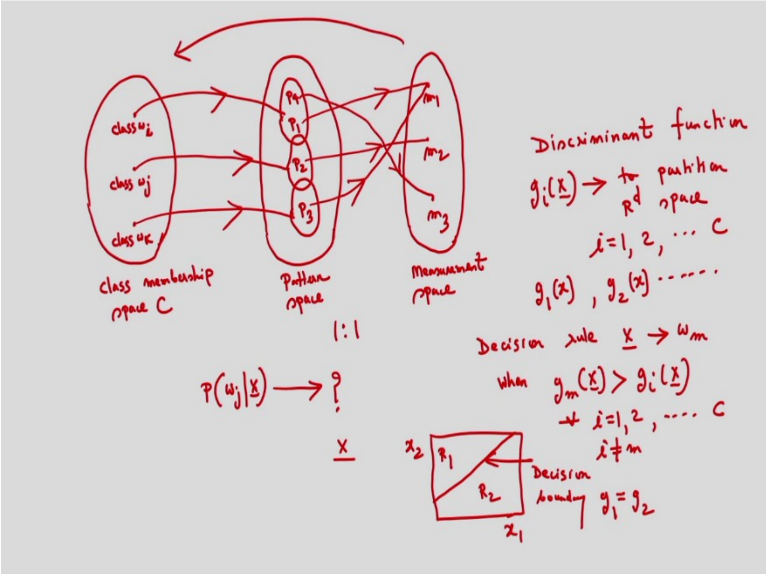
So, I am considering that class 1 and suppose with a class 2 and between class 1 and class 2, I am considering the decision boundary. So, this is the decision boundary. Now, in this case in case of the hard decision, you can see I have a clear separation between the classes, the class 1 and class 2, but if I considered a fuzzy decision boundary, the decision boundary will be something like this, this is not a rigid in case of the fuzzy decision boundary.

So, what happened, there is a possibility that a particular sample, suppose this sample belongs to another class. In case of the hard decision already I have shown the decision boundary, there is no possibility that a particular sample belonged to another class, but in case of the soft decision, there is a possibility that a particular sample may belongs to another class, this depends on the fuzzy membership grid.

So, if you read the fuzzy logic, then you can understand one concept that concept is that fuzzy membership grid. So, based on the fuzzy membership grid, there is a possibility that a particular sample may belong to another class. So, one is the hard decision. That is there is no possibility, because the decision boundary is very rigid and you will be getting the separation between two classes, the class 1 and the class 2.

But in case of the fuzzification, in case of the soft decision based on the membership grid, there is a possibility that a particular sample may belong to another class, it is based on the membership grid that is about the hard decision and the soft decision and how to represent the pattern classification problem. So, I can show you what is the pattern classification.

(Refer Slide Time: 15:36)



So, for this what I am considering, I am considering a space, at this space is the, the class membership, class membership space C I am considering and in this case I am considering the class suppose that class is w_i, another class is suppose class w_j and another class I am considering class w_k. So, three classes I am considering suppose. So, that is the class membership space and I have the pattern space. So, pattern space is something like this, suppose I have two patterns one is P_4 and another one is P, suppose 1 and another patterns species P suppose this is P_2 and another species is suppose P_3.

These I am considering as the pattern space, this is I am considering as patterned space and after this I am considering another space and that is the measurement space. So, corresponding to this I have three measurements suppose m_1, m_2 and suppose m_3. So, each class w_i generates a subset of pattern in the pattern space, if you see the pattern space P_1, P_2, P_3, P_4 they are overlapping that is the meaning is patterns of different classes may share some common attributes.

So, in this case if I considered a mapping suppose. So, corresponding to the class w_i, I have two patterns the P_1 and P_4 corresponding to the class w_j my pattern is P_2 corresponding to the class w_k, my pattern is P_3, this is a mapping I am considering. So, here you see the corresponding to the class w_i, I have two patterns, so one is p1 and P4 corresponding to the class w_j, I have the pattern, the pattern is P_2, corresponding to the class w_k I have the pattern P_3 and you can see here the patterns of different classes may overlap, why it is overlapping?

The, because the patterns of different classes may share common attributes. So, that is why it is overlapping and corresponding to these patterns if I considered a mapping from the pattern space to the measurement space. So, corresponding to the pattern p_1, a measurement is suppose m1 corresponding to the pattern P_2 the measurement is m_2, corresponding to the pattern P_3, the measurement is suppose m_1 and corresponding to the pattern P_4 the measurement is suppose m_3.

So, this is a mapping from the pattern space to the measurement space. Now, in this case the problem of the pattern classification is, I have the measurements from the measurements I have to determine the class, the corresponding class. So, suppose the measurement is m_1. So, what is the corresponding class I have to determine, but in this case, it is not one to one mapping, if you see, it is not one to one mapping that means from the measurement I have to determine the class, that is nothing but the inverse mapping.

So, I have done measurements and from the measurements I have to determine the corresponding class, but it is not one to one mapping. Had it been one to one mapping, then the pattern classification problem would have been very easy. Since it is not one to one mapping, then in this case we have to apply some techniques like machine learning techniques.

That means from the measurement, how to determine the particular class and this is nothing but the inverse mapping, the mapping from the measurement space to a class, a membership space. So, like this I can define the concept of pattern classification, statistically I can represent this problem like this, the probability of w_j given x. So, what is the meaning of this? So, I have to find a probability of obtaining a particular class given the feature vector, the feature vector is x, x is nothing but the measurement.

So, from the measurement, I have to determine the particular class. So, that is the concept of the pattern classification and I have to determine this. So, in pattern classification, I have to determine this probability and, in this case, there are two learning techniques one is the supervise, in case of a supervise learning technique, we know the class levels and the corresponding training samples are available.

So, suppose I have the class w_i and corresponding to w_i I have the training samples, that information is available in case of supervised learning and after learning after training, we have to do the classification based on the train model, that is the supervised classification. In case of the unsupervised classification, we do not know about the class levels.

So, what is available only we have the feature vector. So, we have to group the features of vectors based on some similarity. So, that is nothing but that clustering. So, what do we have to consider we have to group the feature vector based on some similarity that is the unsupervised. So, later on I will discuss about supervised and unsupervised, the main concept is that in case of the supervise, we know that class levels and also, we have the training samples corresponding to a particular class.

In case of the unsupervised learning, we do not know about the class levels, but the feature vectors are available. So, we can group the feature vectors based on some similarity that is about the unsupervised technique and for classification, one mathematical representation is that we can define one function that function is called discriminant function.

This function is used for taking classification decisions. So, this discriminant function is represented by g_i (x) that is to partition R to the power d space. So, g_i (x) is nothing but two partition R to the power d space and in this case, I am considering i is equal to 1, 2 up to C number of classes.

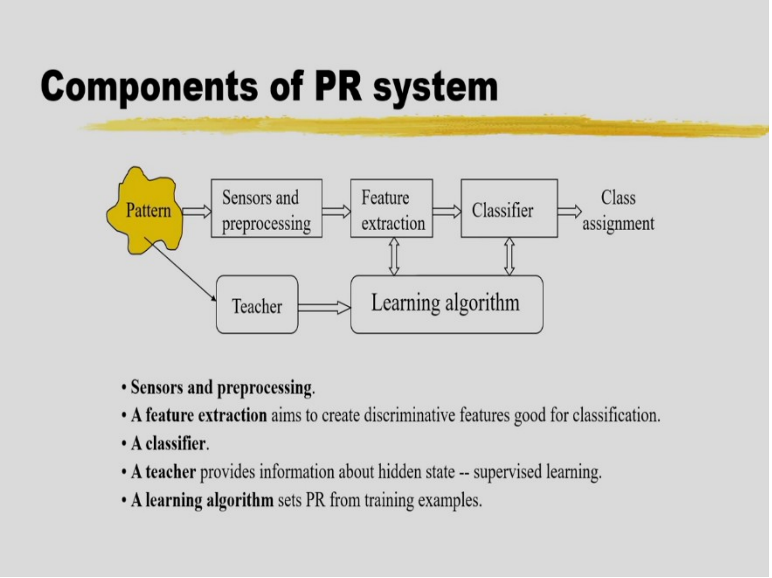
So, for class 1 suppose, I have the discriminant function $g_1(x)$, for class 2 I have the discriminant function $g_2(x)$ like this. So, I have to determine the discriminant function and after this I have to find the maximum discriminant function that corresponds to that particular class. So, I have to determine the discriminant function for the class 1, for class 2, class C I have to determine the discriminant function and, in this case, I have to determine the maximum discriminant function that corresponds to that particular class.

So, the decision rule will be something like this, what is the decision rule, the decision rule will be like this the feature vector x which is assigned to this particular class, the class is suppose w_m, when the g_m (x) is greater than g_i (x) for I is equal to 1, 2 up to C and i is not equal to m. So, based on this condition if gm x is greater than g_i (x), then in this case based on this I can assign the feature vector to the class that classes w_m and in this case you can see the discriminant function is used to partition R to the power d space that is the d dimensional space that is the feature space.

Suppose, if I consider the two-dimensional feature space, the 2D feature space, So, suppose x_1 and x_2, two-dimensional space or space and you can see this is the decision boundary I am considering and corresponding to the class w_1, their region is R_1 and corresponding to the class w_2, the region is R_2 and here you can see this is the decision boundary. This is the decision boundary.

It is nothing but g_1 is equal to g_2, g_1 will be equal to g_2. So, in this case what I have to consider I have to determine a discriminant function corresponding to the class 1 and corresponding to the class 1 region is R1 and corresponding to the class 2 the region is R2. So, for class 2 also we have to determine, the discriminant function and this maximum discriminant function, the maximum, which one is the maximum g1 will be maximum or maybe the z2 may be maximum based on this we can take a classification decision and in this case, what is the equation of the decision boundary the equation of the addition boundary is z1 is equal to z2. So, one important point is that decision boundary and the discriminant function.

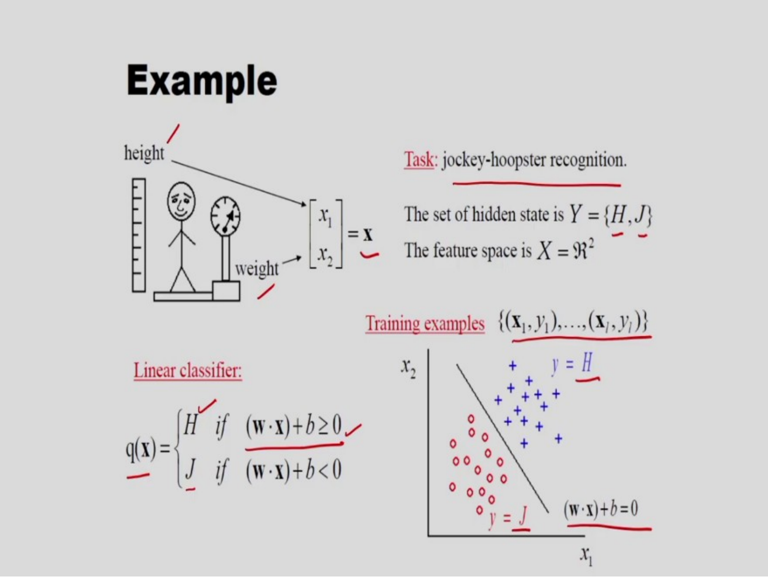
(Refer Slide Time: 25:54)



Here I have shown the components of pattern recognition system. So, already I have defined So, we have the patterns, after this we have some sensors for measurement, that is for the data equation and after this we have to do some pre-processing, after this we have to extract some features, after extracting the features, we can select some important features that is called the feature selection and after this we have to go for classification.

And in this case, we have shown here, learning algorithms we have shown that is nothing but if I consider the supervised training and the supervised learning, the training samples are available for all the classes and we know the class levels and based on this information, we can train the classifier we can train the system. So, this is about the components of the pattern recognition system.

(Refer Slide Time: 26:53)



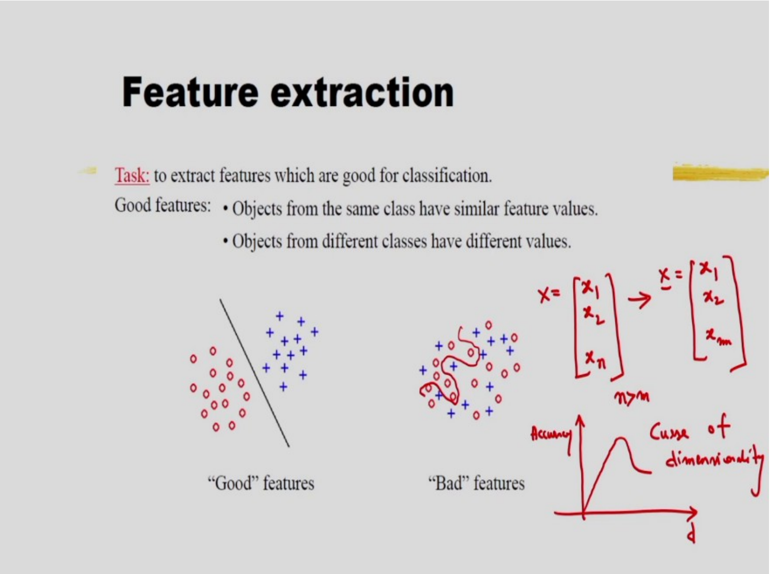
And in this case, I have shown one example of pattern classification here you see the problem is the jockey and the hoopster recognition and for this I am considering these two states one is the H and another is J that is the two classes I am considering and I am considering the two-dimensional feature space and in this case I am considering two features, one is the height, another one is the width.

So, x is the feature of vector and x_1 is the height and x_2 is the width. I am measuring by using the sensor. So, the feature of vector has two components, one is x_1 another one is x_2 that is these are elements of that feature vector. After this what I am considering, I am considering the function q (x) in this case what I am considering this equation I am considering $w.x$ that is a dot product between the width vector w and the input vector, input feature vector is x plus b is the offset I am considering.

If it is greater than 0, then the corresponding class will be H and if this $w.x+b$ is less than 0, then the corresponding class will be J. So, based on this condition, I am determining the class, particular class that is the decision making I am doing based on these equations and in this case you can see I am showing two classes in the figure, the red is the class J. So, this is the class J and the blues are the class H and in between you can see the decision boundary.

So, what will be the equation of the decision boundary, the equation of the decision boundary is $w.x+b$ is equal to 0 and for this I have the training examples, the training samples are available. So, x_1, y_1, x_2, y_2 like this, that training examples are available to train the system and after this we can find a decision boundary between these two classes and in this case, you can see based on this objective function that is $w.x+b$ is greater than or equal to 0, I can decide that class H and also if I consider $w.x+b$ less than 0, then the class will be J. So, this is the example, one example I am considering.

(Refer Slide Time: 29:11)



And for feature extraction, you can see here in these two figures, I am considering good features and the bad features. In the first figure you can see I am considering two classes; one is the red and other one is blue. In case of the good features I can draw, I can easily draw a decision boundary between these two classes and this is a linear decision boundary. In case of the bad features, it is very difficult to find a decision boundary between the classes.

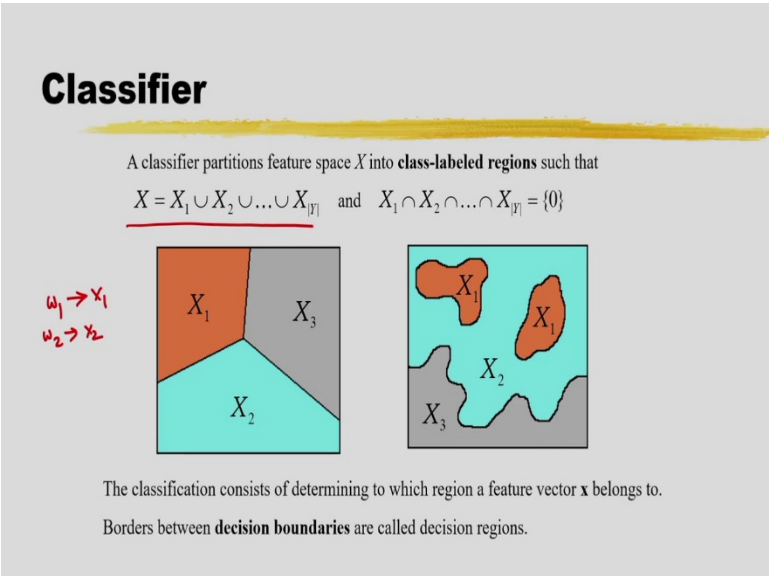
So, here you see in the second example in the second figure, it is very difficult to draw the decision boundary between the classes. So, it is very difficult to draw the decision boundary between the classes. So, these are the examples of the bad features. So, that is why the features selection is important.

So, suppose I have a feature vector, the dimension of the feature vector is suppose n dimension x_1, x_2… x_n. So, all the features may not be important. So, that is why I have to reduce the dimension of the feature vector. So, I can reduce the dimension of the feature vectors by some methods. So, one method already I have discussed that is the PCA, the Principal Component Analysis.

So, you can see n is greater than m. So, I can reduce the dimensionality of the feature vector. So, that is called a feature selection and one issue is like this suppose, if I consider, suppose D is the dimension of the feature vector and if I consider this accuracy this side, so if I increase the dimension of the feature vector initially the accuracy will increase and after if I increase the dimension again the accuracy will drop like this.

Because, if the number of training samples are limited, limited training samples are available and if I increase the dimension of the feature vector, that may not increase the accuracy, the accuracy may decrease. So, that is why feature selection is important, because of the high dimension the accuracy drops and this is called, this is called the curse of dimensionality. If I increase the dimension of the feature vector, the accuracy may not increase it may drop and this is called the curse of dimensionality. So, we have to reduce the dimension of the feature vector. So, that is why the feature selection is important.

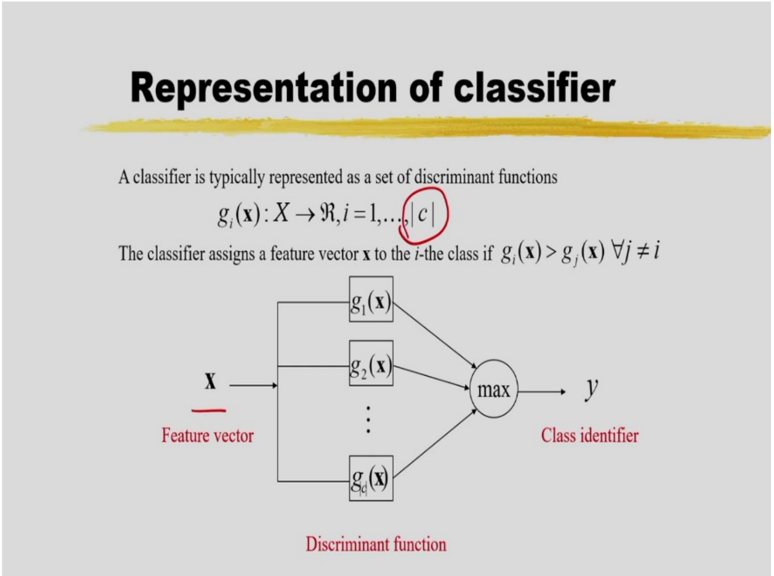
(Refer Slide Time: 31:44)



And in this case, I have shown, the classifiers and also you can see the regions X_1, X_2, X_3 corresponding to different classes. So, suppose the class w1 corresponding to X_1, the region is X_1, corresponding to w_2, the region is X_2 like this, I am considering and a first case if you see the first figure that is the linear decision boundary I am getting between different classes and if I consider the second figure.

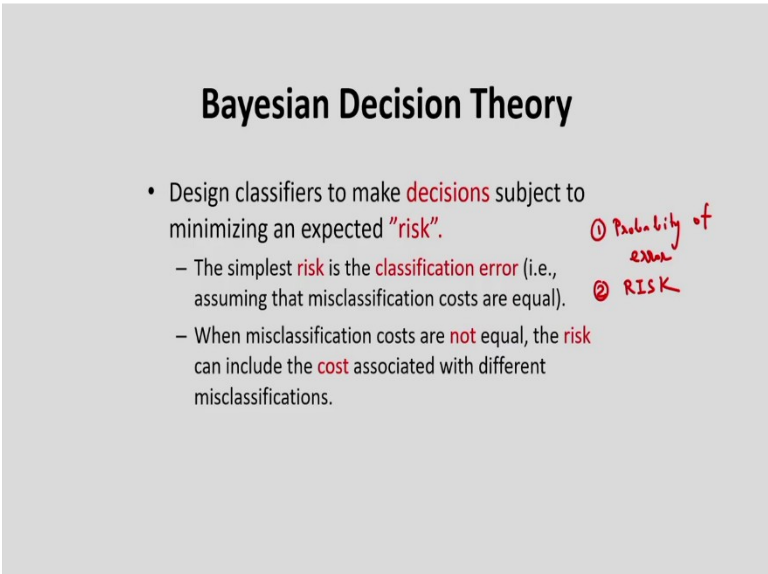
In the second figure, I am not getting the linear decision boundary between the classes. So, you can see X_1, X_2, X_3 like this I am considering. So, in the first case I have the linear decision boundary and in the second case I am not getting the linear decision boundary and here you can see x is nothing but, x1 union of x2, union of like this. So, this is the total space, that is a feature space.

(Refer Slide Time: 32:49)



And this concept already I have explained. So, here you can see the input is the feature vector here and corresponding to this feature vector I am determining the discriminant functions g_1 (x), g_2 (x), g_3 (x) like this for all the classes. I have C number of classes and I have to determine the maximum discriminant function. So, suppose g_2 is maximum that means, the corresponding class will be 2. So, based on this discriminant function, I can do the classification. So, based on the discriminant function, I can do the classification.

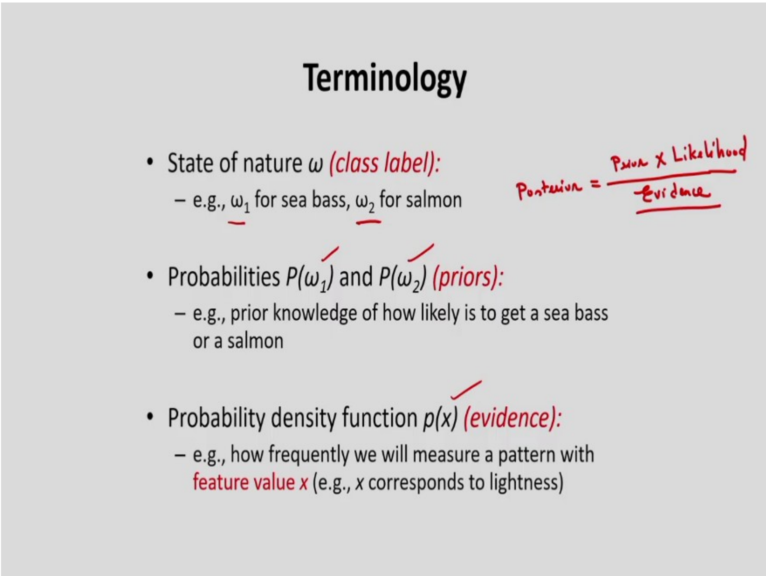
(Refer Slide Time: 33:25)



And later on, I will discuss about the Bayesian decision theory that is mainly on the Bayes law. And in this case, I will be considering two measures, one is the probability of the error. So, probability of error also I can determine and also the risk, risk also I can determine. In case of the base decision theory and based on these two measures, one is the probability of error and the another one is risk I can take a classification decision.

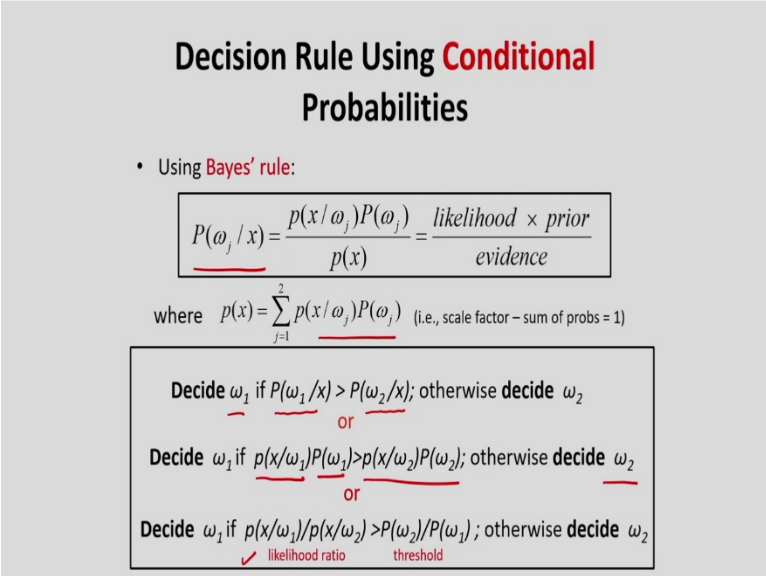
So, what is risk? Risk of taking a particular action. So, I have to take some actions and I have to determine the risk of taking these actions and I have to determine the conditional risk and based on the conditional risk I can decide a particular action and also the classification decision I can take. So, by using these two measures, one is the probability of error and other one is risk, I can do the pattern classification.

(Refer Slide Time: 34:37)



And in this case in case of the Bayes law, these are the terminologies one is the class, the class are represented like this w_1, w_2 like this. So, in this case, I am considering two features. One is the sea bass, other one is just salmon. And in this case, I am considering the prior probabilities. The prior probabilities are P (w_1) and P (w_2) these are the prior probabilities and we have the evidence and that is the P (x). So, if I considered a Bayes law, the posterior probability is equal to prior probability into likelihood divided by evidence. So, this is the base law, evidence has no significance in classification, it is nothing but the normalising factor. So, based on the information of the prior and the likelihood, we can take classification decision.

(Refer Slide Time: 35:43)



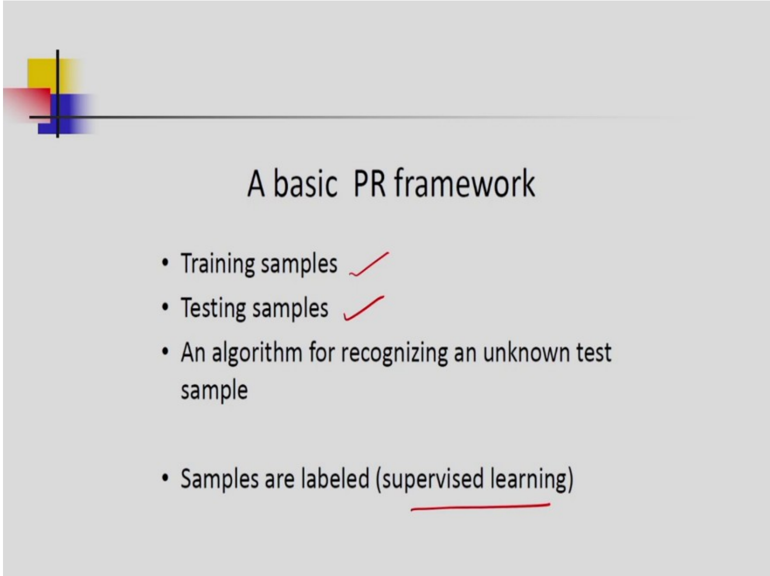
So, here you see I am showing the Bayes rule that is the posterior probability you can see, is equal to the likelihood into prayer divided by evidence, the evidence is nothing but the scale factor. So, if you see the evidence, the evidence is nothing but the scale factor that is used for normalisation. So, in classification, it has no significance.

So, based on the parameters, the parameter is likelihood and the prior we can take classification decision. So, let us consider these decisions. So, decide a particular class w1, if the probability of w1 given x is greater than probability of w2 given x, otherwise we can decide the class w2. So, this is based on the posterior probability, the posterior density.

Similarly, if I consider this likelihood and the prior, so this is the likelihood and the prior that based on this I can decide a particular class, if the probability of x given w1 into probability of w1 is catered in probability of x given w 2 into probability of w2, then I can decide the class w1 otherwise, I can decide the class w2.

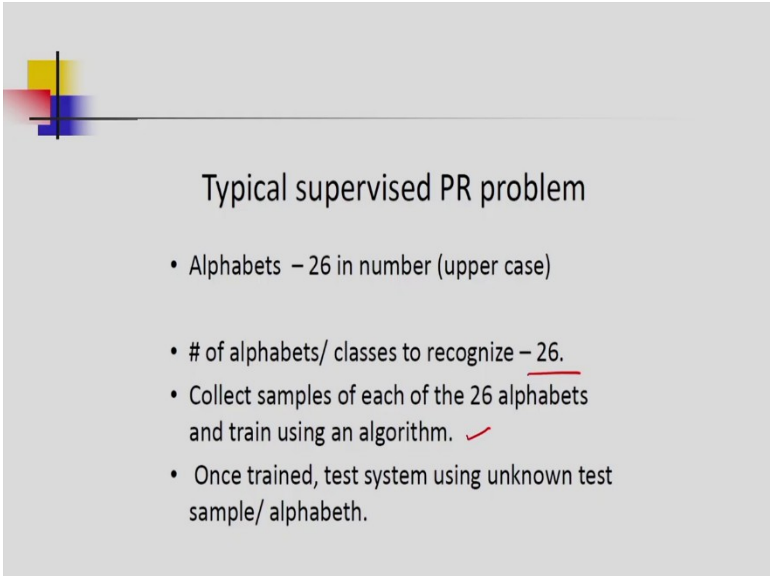
And in this case, we can define that term, the term is the likelihood ratio. So, from this I can determine the term the likelihood ratio, if the likelihood ratio is greater than a particular threshold, then in this case, the corresponding class will be w1 otherwise, we have to decide the class w2. So, based on the Bayes law, I can do the classifications you can see.

(Refer Slide Time: 37:28)



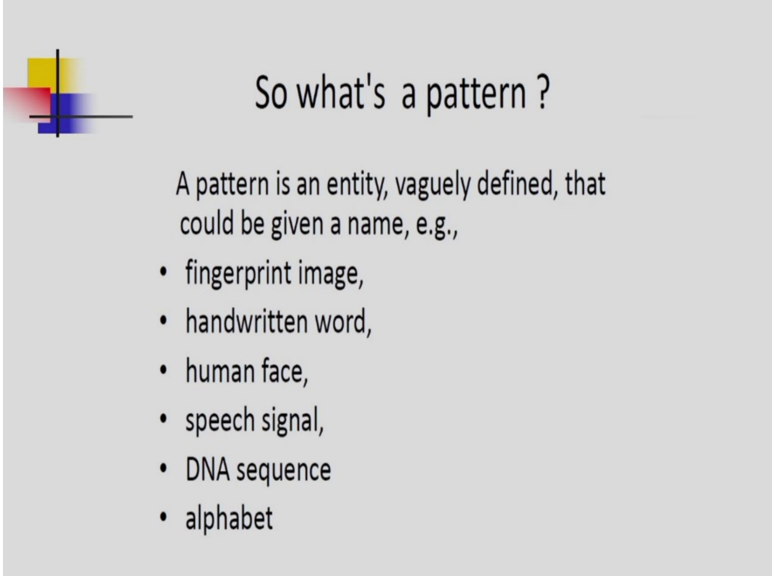
In case of the basic pattern recognition framework, we need the training samples, we need the testing samples and also the algorithm for recognising unknown test samples and we are considering the supervised learning that means, we know the class levels and corresponding to all the classes we have the training samples. So, we are considering the supervised learning.

(Refer Slide Time: 37:55)



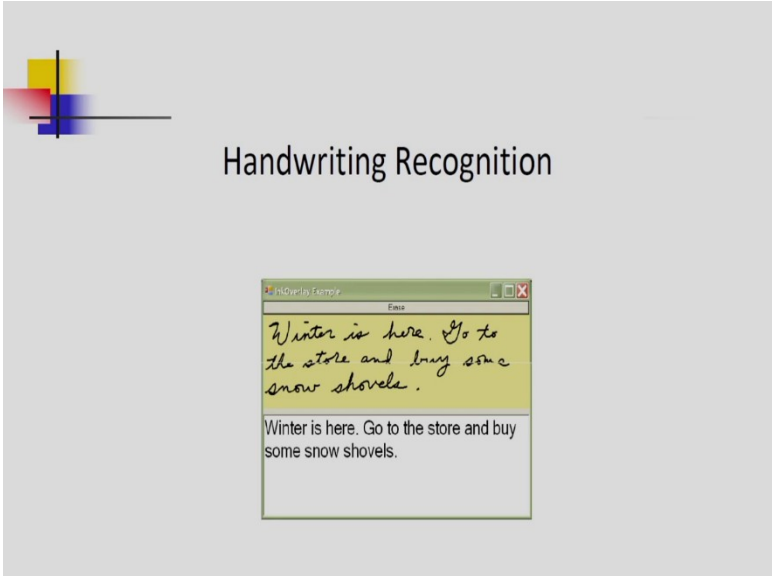
And in this case, one problem is suppose alphabet recognition. So, 26 alphabets are available, uppercase. So, for this what we have to consider collect the samples of each of the 26 alphabet and train using an algorithm. So, first we have to collect all the training samples of the alphabet and how many classes, we have 26 classes because we are considering 26 alphabets, uppercase alphabet. After training what we have to do we have to do the testing using unknown samples, that is the unknown sample means unknown alphabets we have to consider and after this we can determine the accuracies. So, this is the typical supervised pattern recognition problem.

(Refer Slide Time: 38:42)



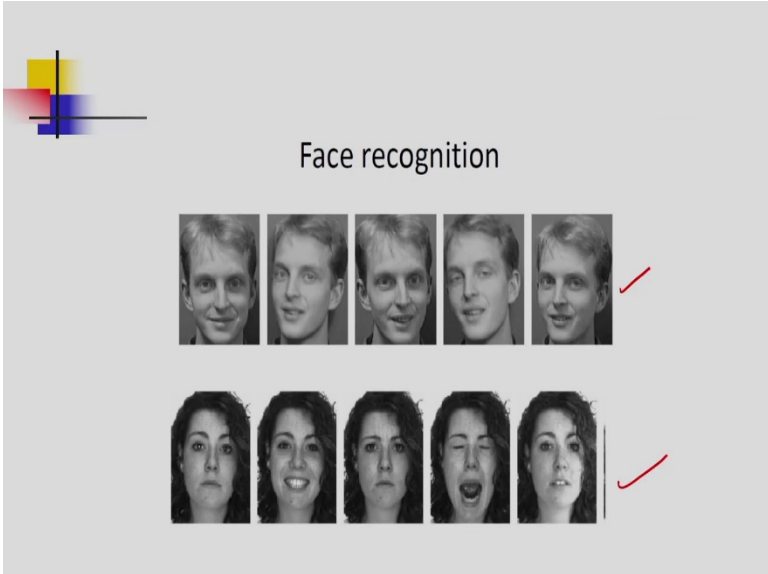
And what are the patterns, already I have defined what are the patterns, the patterns maybe a signal or maybe the fingerprint image, handwritten word, human face, speech signal, DNA sequence alphabets like this, these are the examples of the patterns.

(Refer Slide Time: 38:57)



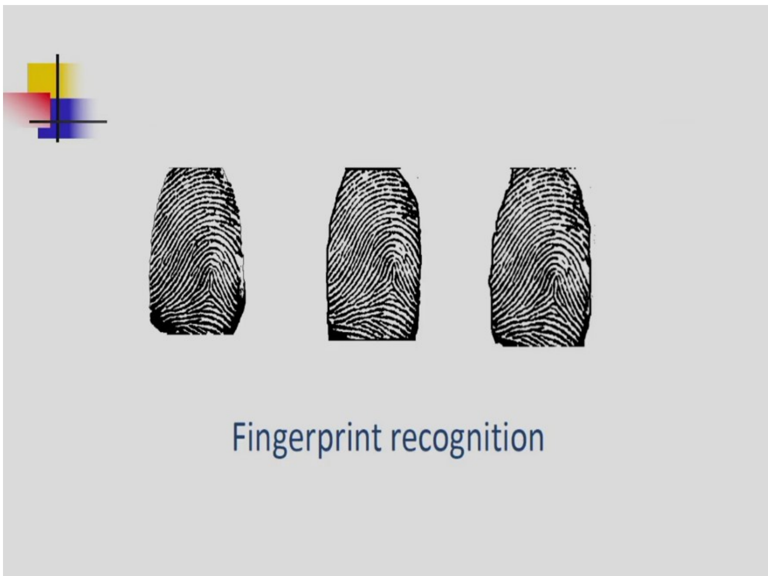
And in this case, I have shown one example of the handwriting recognition. So, for this also we have to extract the features and based on these features we can recognise and the handwritings.

(Refer Slide Time: 39:08)



And this is another example, the face recognition here if you can see, if I consider the first row, the same face, but different lighting conditions and different poses. In this case also we have to recognise the face. So, for this also we have to extract the features and based on these features, we can recognise that particular face. In the second row also, I have shown the face ability and, in this case, you can see different poses and different facial expressions and also different lighting conditions. So, for this also we have to recognise the face.

(Refer Slide Time: 39:48)



And this is another example that is the fingerprint recognition. So, in this class I discussed the concept of pattern classification and also, I have shown one mapping diagram. So, in the mapping diagram you can see from the measurement, I have to determine the classes, the corresponding class I have to determine.

So, the feature of vector is available and I have to determine the corresponding class. Statistically that is represented by probability of w_j given x. That is the definition of pattern recognition, pattern classification. In my next class, I will continue the same discussion. So, let me stop here today. Thank you.