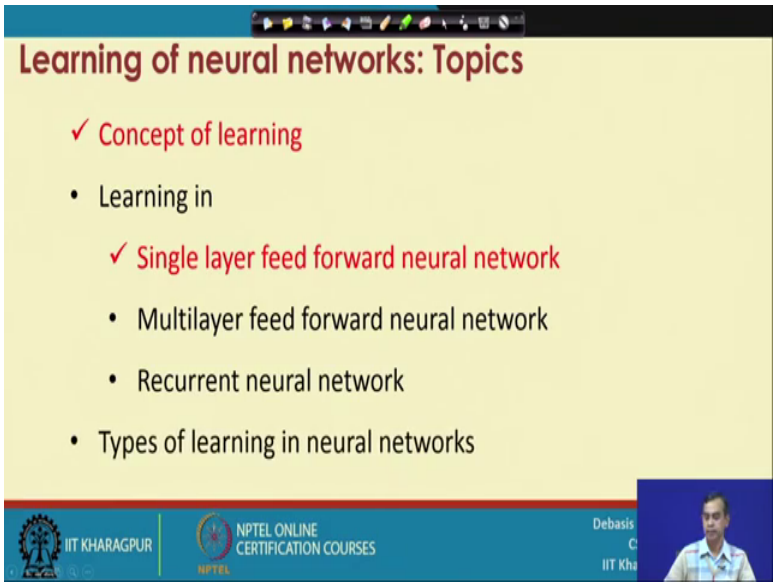
Introduction to Soft Computing
Prof. Debasis Samanta
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Lecture - 36
Training ANNs

We are discussing about solving problem in Soft Computing using artificial neural network. In the last 2 lectures, we have learned about how the basic unit which is there in our human body. So, for the nervous system is concerned that is the neuron can be mimicked to a perceptron. And then we have discussed about how such a perceptron can be modeled to solve problem. We have also discussed the different architectures that can be used to solve different problems or varieties, varieties in case of varieties of complexities and the different architecture we have discussed. And so for the neural network is concerned, the architecture is basically the first thing that we have to decide to solve our own problem.

Now, in order to discuss this architecture, we have decided that how a architect how an architecture can be modeled. Now modeling and architecture is basically by means of weight matrices and then the transfer functions and a threshold value in it is perceptron. In the next, in this lectures we will learn about the idea about how the learning can be accomplished and then what are the different learning techniques are there.

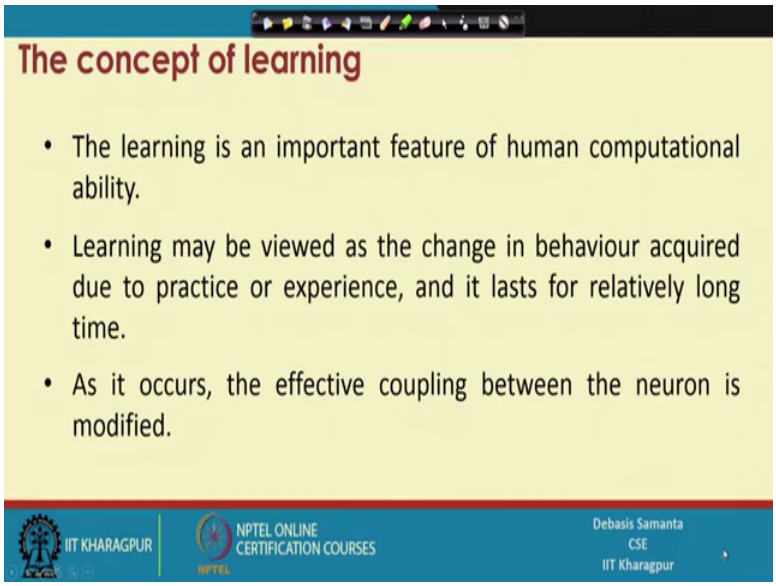
(Refer Slide Time: 01:32)



So, basically first we will discuss about the concept of learning and then as the learning as the architecture is different than definitely, the learning concept also will be different. So, in this lecture, we will discuss about after concept of learning how to learn a single layer feed forward neural network and then multilayer and then other neural network learning will be discussed in the next class, next lecture next lectures.

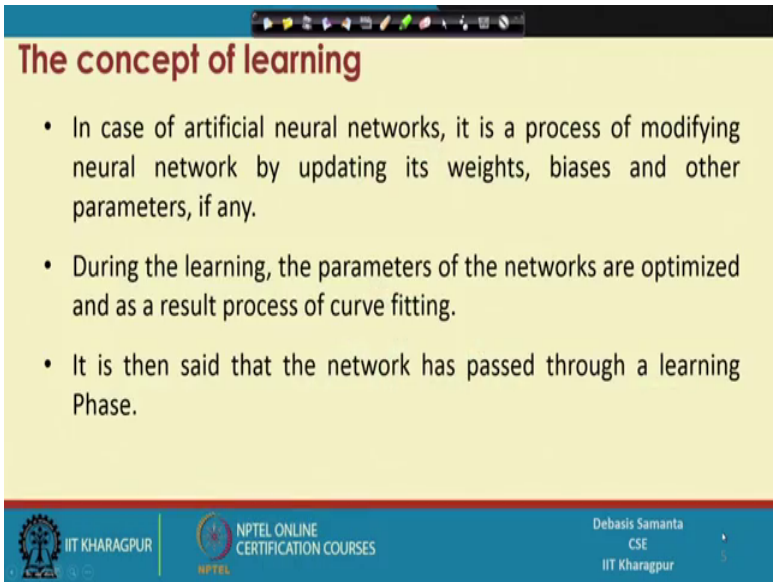
Now, so first let us understand about what is the concept of learning. Now concept of learning here in our every daily life, whenever we are learning every day in fact, whenever we see something, from there we can learn many things. So, basically we see, we here, we can sense and therefore, we can learn. So, the learning it is inherent in our body, in other human or in any other living things; even trees, the animals everybody is in fact learning from the environment, from the atmosphere where they belongs.

(Refer Slide Time: 02:40)



So, it is also the same thing is applicable in case of our neural network. The learning is also there and the learning means basically in case of our neural network, learning means how the different values of each perceptron can be learned. So, this learning is basically the idea it is there.

(Refer Slide Time: 02:58)



And in case of artificial neural network, it is basically the idea is that it is a process of modifying a neural network by means of the different weights, the threshold values and other parameters namely the number of what is called the perceptrons, number of layers and then the threshold values and may, so many things are there.

So, basically when we can train one network artificial neural network it is also called when artificial networks learn. This basically they try to find optimum values of all these, I mean how we can find an optimum values in terms of number of perceptrons to be included in the network, how the network connections can be optimized, how the different threshold values can be decided and all these things are there. So, if the network can learn it by means of some methodology then we can say that the training of the network is over. So, the network can learn itself from the training data.

Now, here the idea about that how network can learn it in order to learn basically how we the human people can learn it. So, for human learning, we need a large number of inputs or in other word, we can say a teacher is there; teacher can say many thing and this is the input. So, if the input is known to us then we can learn from the input. So, here also in the neural network, the input is needs to be fed to the network and then from the input, neural network automatically learn it.

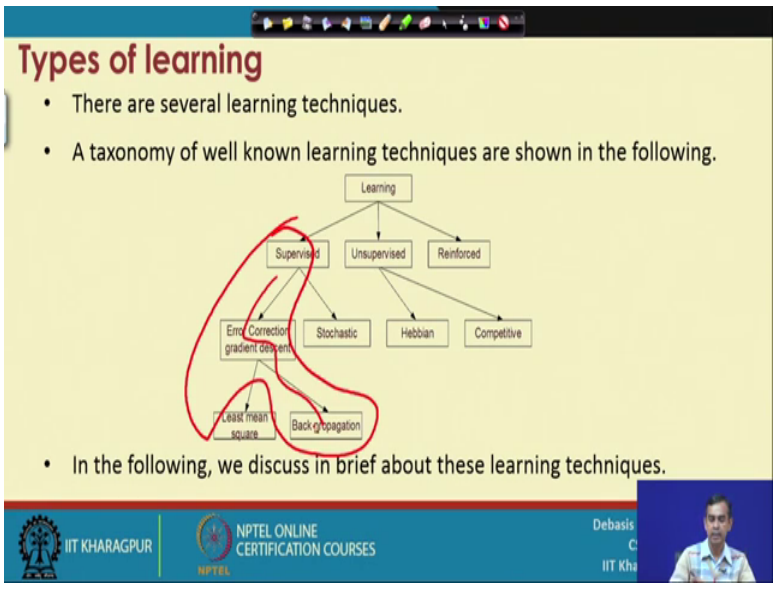
So, this input is called in neural network learning is called the training data. So, once the training data is known then we will be able to learn the network effectively. As an

example, I can tell this say suppose we have to recognize handwritten character now; so for the handwritten character and we want to use this neural network to solve this problem.

So, what is the training data in this case. I can consider 500 subjects. So, that they can write the different characters in the English alphabet, suppose a the different people allowed to write the A. So, 500 people in this case. So, 500 A's is are there. Now this 500 as can be the training data. So, if we give this training data to this neural net, then it will learn automatically to recognize the different the characters. So, for example, A character first then B character the C characters and so on, so on. So, batch wise the different characters or in a single batch, all the characters with different singles can be given and then the neural network can recognize this pattern as either it is A, B, C or this one.

So, this is the idea it is there, so training data maters. So, training data from 500 or training data from 50 people or 50 samples that can be fed to the neural network and if we give it to the neural network, it automatically decides it is how many perceptron it is required to solve this problem and how many weights and weight matrices are there, how many layers are there and what are the transfer function that can be considered and what are the threshold values that can be considered. So, this is the concept of learning there in the neural network.

(Refer Slide Time: 06:25)

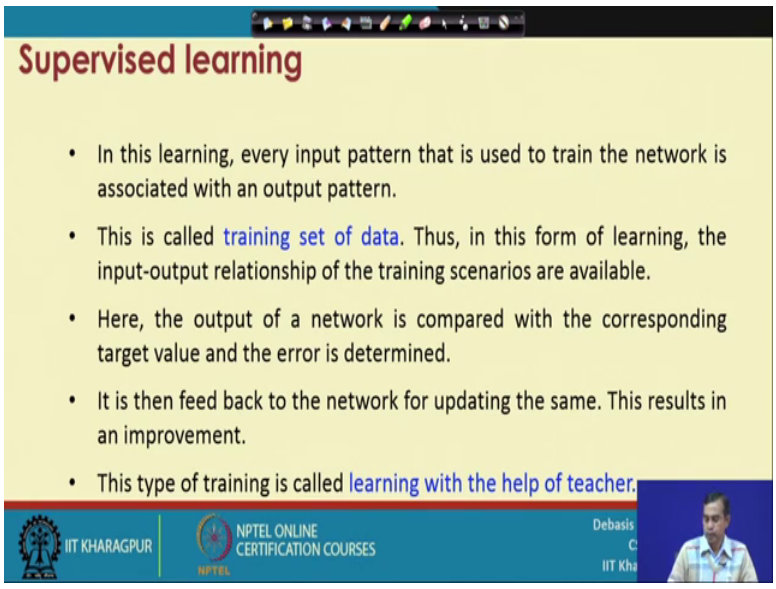


Now, so far the concept of learning is concerned, the theory of learning is very exhaustive there. In fact, there are many learning principles, many techniques are there. Now in this slide, we see some learning techniques. All the learning techniques can be classified into broad 3 categories called the Supervised learning, Unsupervised learning or Reinforced learning.

Now, supervised learning can be again decided as a stochastic learning or error gradient based learning, error correction gradient descent learning method. And then error correction gradient descent method again either least mean square, least mean square method or back propagation method. On the other hand, Supervised, Unsupervised learning; there are 2 types, one is called the Hebbian learning technique, another is competitive learning technique.

Let us quickly have some brief idea about all these different learning techniques and then as it is the constant, the timing constant is there will not be able to cover all the learning techniques but only few learning techniques will be covered; more precisely, we will discuss about Supervised learning and then the propagation learning this one. So, the we will limit our discussion to these kind of things are there. All other techniques, we should have a little bit brief idea about the overview, not the details concept of the learning it is there.

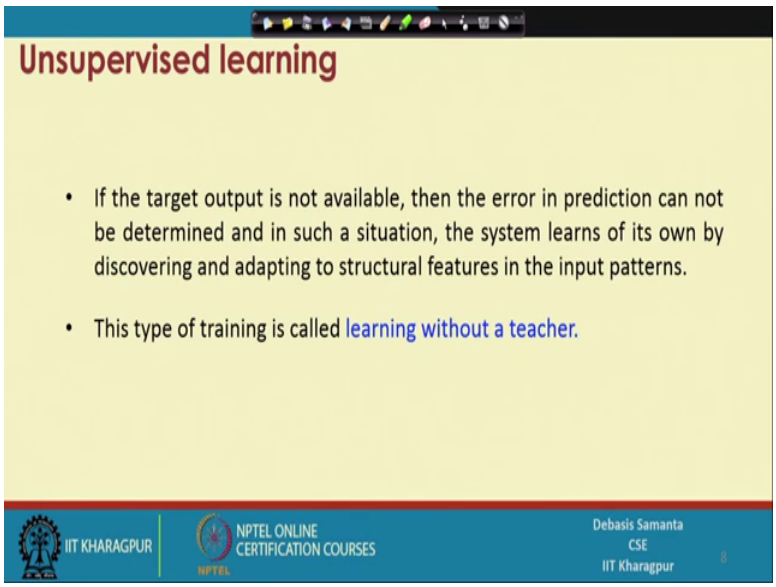
(Refer Slide Time: 07:50)



So, in the few next slides, let us quickly learn about what us the different learning techniques are there. I will first discuss about Supervised learning. Now in each learning, one thing is common that we need training data or a set of data that can be used to learn the network. So, training data is there. So,, so in this basically, in case of supervised learning, so for input what will be the output it is already known to us. So, that is why this is called the supervised training data; that means, for every input what is the output we should say it.

So, one hand written character, the input is there, then we say that it is a like this one this. So, this is why this kind of learning is also called learning with a help of teacher. So, basically teacher said the question or tell the, I mean ask the question and then also if you are not able to give the answer, is also tell the answer also. So, this is why it is called learning with the help of teacher. So, Supervised learning is very simple and state forward impact and yet is very effective also.

(Refer Slide Time: 08:55)

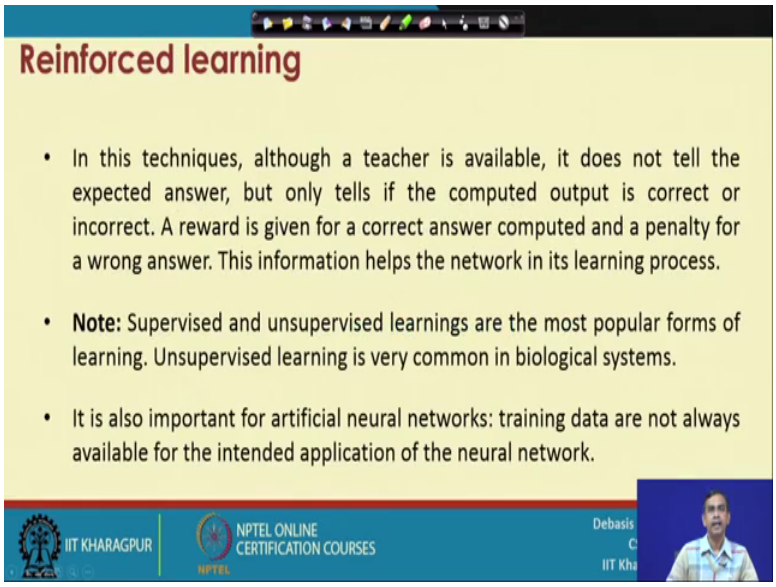


Now, the next is Unsupervised learning. If the supervised learning is learning with the teacher, unsupervised learning also can be termed learning without a teacher. So, here in this learning concept, if in case of supervised learning as I told you input is given then, output is there, but in case of unsupervised learning, input is only given; no output is told.

So, in this case for example, in case of character recognition, we can give 26 different patterns that I write and 26 different patterns related to the different input characters and then the 10 if it is learned by means of unsupervised, then they can automatically decide about whether it is which character it is.

So, here basically output is not mentioned here, in the line of input ; only the input is there. So, this is the Unsupervised learning concept is there.

(Refer Slide Time: 09:52)



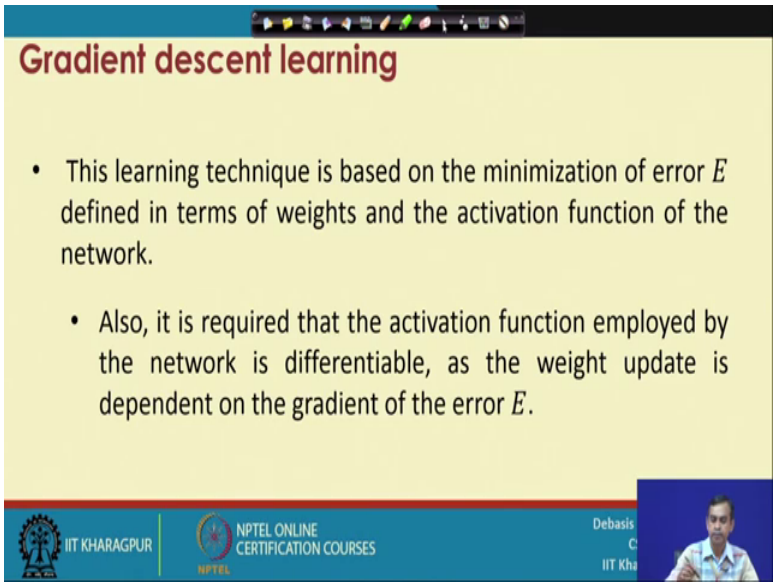
And the next is called the Reinforce learning it is the ok; supervised learning, unsupervised learning and the reinforce learning.

Now, in case of reinforce learning, a teacher is available like a supervised learning, but teacher does not tell the expected answer, but only tells that if the output is give answer is given, whether the answer is correct or incorrect. For a given correct answer, it basically reverse something and for the incorrect answer, it gives a penalty. Now knowing the questions and therefore, the reward or penalty they the network can learn it and this is the idea about Reinforced learning.

So, in the theory of learning, mostly the Supervised learning and Unsupervised learning are the most popular form of learning and Unsupervised learning is basically the common learning technique in our biological process. We people easily follow unsupervised learning otherwise supervised learning is more things are there.

Now, so, so here actually um, so for the neural network learning or training is concerned, we heavily depends on the input data or the training data it is there now. So, these are the learning techniques the 3 different learning techniques are there.

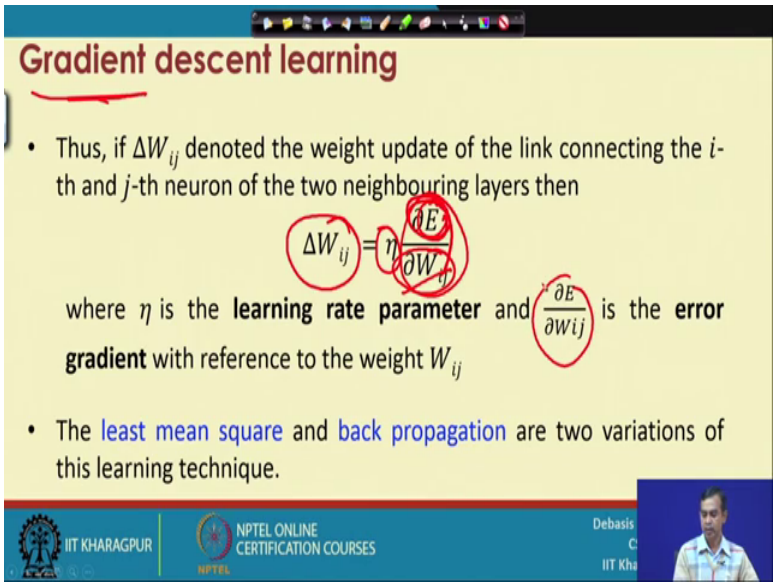
(Refer Slide Time: 11:16)



Now again, so for the supervised learning technique is concerned, there are again 2 type the gradient descent learning that can be my by means of least mean square or it is basically the error descent gradient, descent calculation formula is there or back proper propagation is there. We will discuss these things in details so that then it is basically, it basically calculates the error it is the input and output is known. So, it basically calculates the error and it basically try to optimize the parameters in the network so that this error values can be minimized.

So, this is the main idea about this gradient descent learning techniques. It is basically supervised learning techniques in general.

(Refer Slide Time: 12:02)

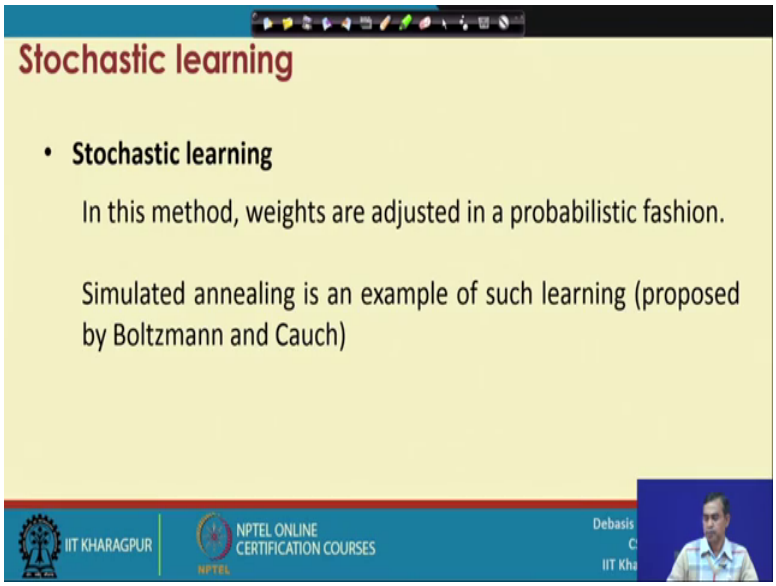


And, and the gradient descent techniques is basically, it is called the gradient descent as a gradient basically the gradient concept is there. So, gradient is because it basically, if E denotes the error at any instant or and then if we change the weight matrix for example, then the what will be the change of errors.

So, it will change this basically, it will calculate if the weight matrix E changes then it will see that how the error will be there and then accordingly the change to n matrix will be calculated. And here it is the one constant, it is called the learning parameter. And this del E by del W ij is called the error gradient.

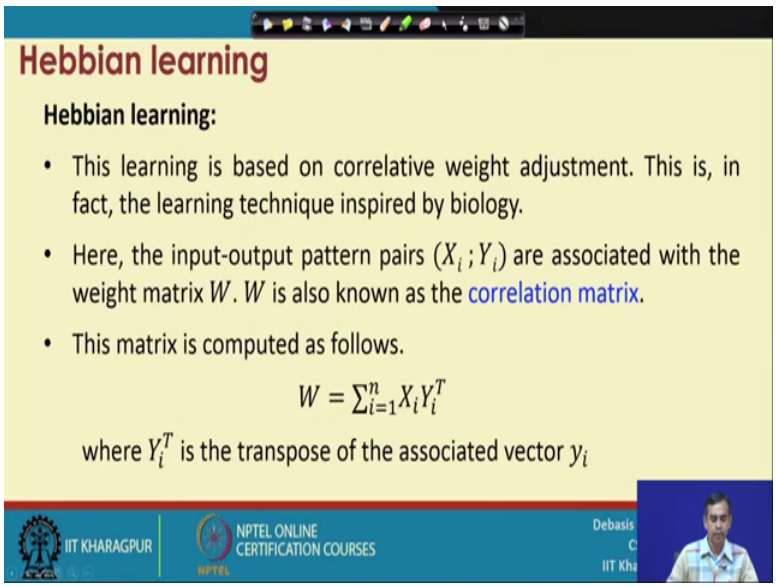
So, basically it will see exactly, if we change this weight then what will be it is error and then accordingly for the minimum errors, it will decide the weight changes and that will be the learning parameters actually. So, this is the idea about it. Now whenever this error needs to be calculated. So, for this error calculation in this gradient decent method 2 technique; either least mean square error or back proportion error is there. So, we will discuss about back proportion algorithm and least mean square error calculation in the training methods.

(Refer Slide Time: 13:17)



Now, so the supervise learning also can be a Stochastic learning. In this method, the different network parameters like weights and others are decided or adjusted in a probabilistic fashion. So, with certain probabilistic or with certain uncertainty, it can be decided. An example for this kind of is simulated annealing. We could not discuss the simulated annealing because the timing constant. So, it is not covered in this course. So, the simulated annealing is one kind of stochastic learning which is used. It basically simulated annealing it used to solve optimization problem solving.

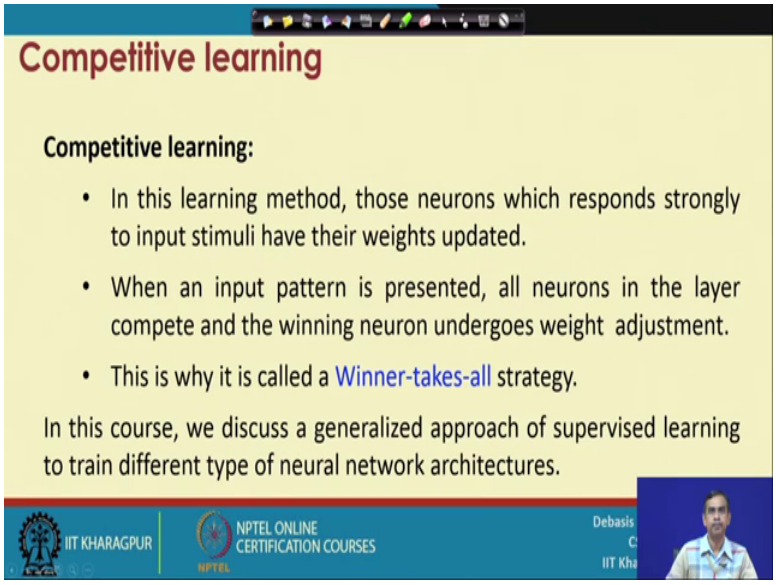
(Refer Slide Time: 13:56)



Now, next is so for the Unsupervised learning is concerned, there are 2 types of learning; one is Hebbian learning, another is Competitive learning. So, Hebbian learning is based on correlation analysis, is basically correlate weight adjustment. So, correlation on a statistical method which basically followed to decide if we decide this values, then how it is correlated with the actual output like this one. So, the correlation analysis and statistical things are involved here. So, a little bit mathematically complex, but it is also useful. So, for the heavier, so far the Unsupervised training if we want to adopt it.

Then it is called the Competitive learning.

(Refer Slide Time: 14:43)

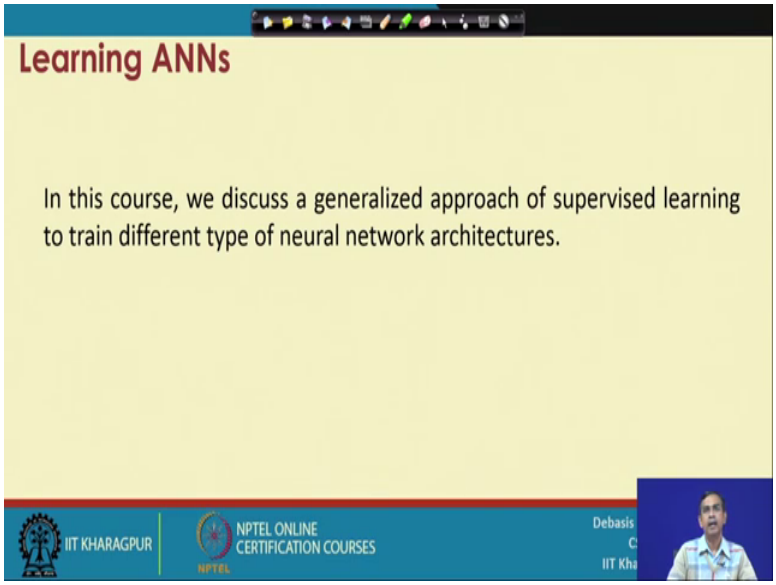


So, in case of competitive learning, so for certain input, if we see the neurons which responds very strongly; that means, it gives a lot of I mean different I mean input signals that it can passed or summation of the values are very strong, then it will decide these values as the learning parameter or this need neural network parameter for that particular neuron which basically responded to a particular input strongly is the parameter that is taken there.

So, it is called the competitive because in this case, the neuron which will basically responds in a strong manner is the winner. So, that is why this kind of learning, this kind of unsupervised learning technique is called Winner-takes-all strategy anyway. So, we have learned about briefly the different learning techniques and in this particular course, it is not possible to cover all the learning techniques. We will cover only the generalized

approach of the supervised learning techniques and then we will also discuss about the different architecture to learn it.

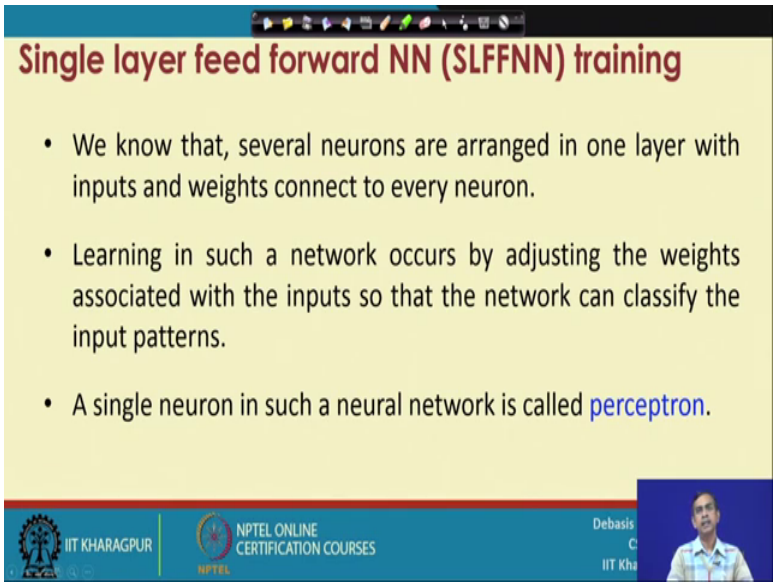
(Refer Slide Time: 15:55)



Ah, so now, so that is why we have limited our approach to generalize approach of supervised learning to different type of neural nat network architecture. Now let us start first training concept with the help of the simple one, training methods to train the Single Layer Feed Forward Neural Network.

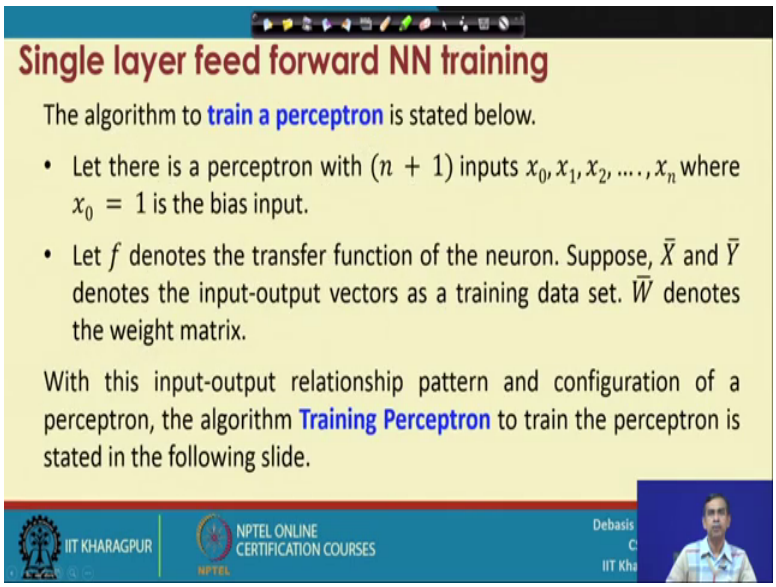
So, we will learning, we will see exactly how a neural network can be trained and we will consider the approach, the supervised training approach.

(Refer Slide Time: 16:25)



So, let us see how the single layer feed forward neural network can be trained and then the thing is using Supervised learning approach.

(Refer Slide Time: 16:35)



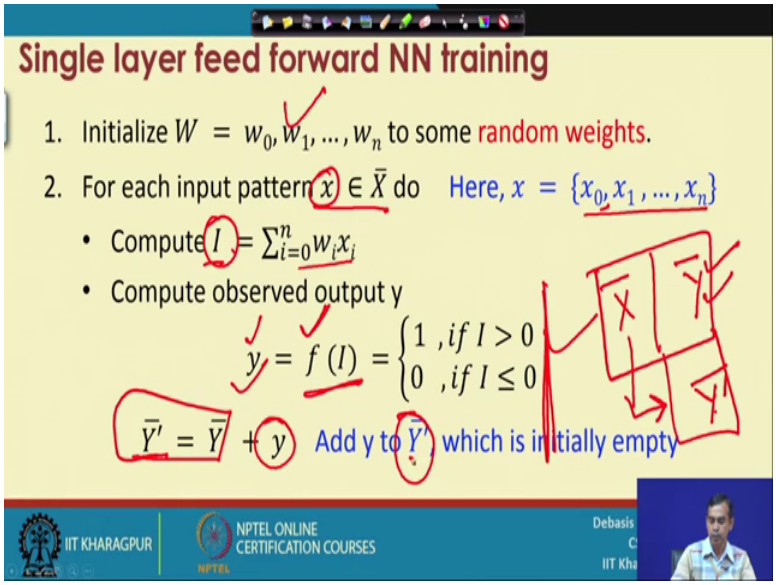
Now,, so first we will discuss about training one unit, the basic unit; that means, training a perceptron. Now in order to training a perceptron, we will consider the inputs. Let us consider these are the input and one input is basically the threshold values input is a bias input also called. So, these are the actual input n number of input erudition the bias input. So, they are n plus 1 inputs are there.

So, we have to basically trained a perceptron and to train this perceptron, we consider that n inputs are there. So, this is basically the inputs to the to the perceptron. And for this perceptron, let us assume this f, f denotes the transfer function. Now we have discussed there are many transfer function like say step of the sigma transfer function, log sigmoid, 10 sigmoid and all these things are there.

So, it can be anyone. So, as we can decide any one transfer function for the perceptron; let it be the step up function. Then we will consider the supervised training as I told you. So, this X bar and Y bar denotes the input and output vectors as a training data set because it is the input data and this is the output data that is there and W bar denotes the weight matrix.

Now, with this input-output relationship pattern and configuration of a perceptron, we are going to discuss about one algorithm it is called the training perceptron to train a perceptron which is ok. So, we will see that how it can be trained or it or a perceptron can learn from this combination ok.

(Refer Slide Time: 18:36)



So, let us see the learning perceptron algorithm. I am just going to define the different steps in it; it is very simple.

So, let us consider W the weight matrix in this case. So, this weight matrix is basically has the n plus 1 number weight value because n plus 1 inputs are there. Now initially all

the values in this weight matrix are chosen with a random values. So, random weights, so weight matrix can be normalized in the scale of 0 to 1 and these are the values are there.

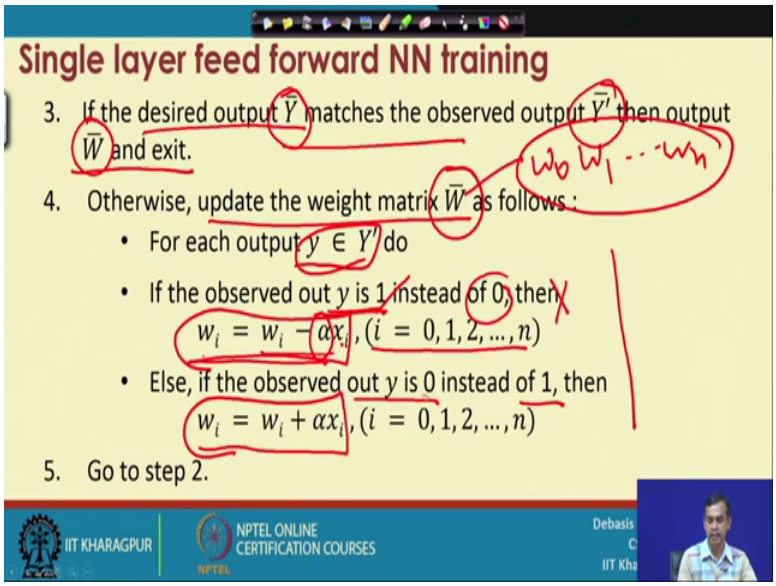
So, initially, this weight matrix is initialized with some random values. So, this is the initialization of the weight matrix and for each input pattern x belongs to this X bar, X bar is a input data set where x is like this one. We have to compute I, I is basically using this w and this I we can calculate I. Then for this perceptron and with respect to a particular input x, we will be able to calculate y that is f I where f is a another function that is assumed ok. As I told you it is a suppose a heavy side transfer function is assumed or sigmoid transfer function is assumed anyway.

So, take let us take with a. And then, so for this input I, we have to see if I greater than 0, it is a 1 and if I less than 0, it is 0 this is basically the simple transfer function or it is called the step transfer function that we have considered f I are in this case.

Now, it will result 1 y. Now in the decide or observe output, we will store y dash. We add this output into this y dash. So, we have decided earlier X bar and Y bar are the input and output as the training data and here for each input whenever we given it to the perceptron, we see for that input what is the observed output. So, it is basically actual output and this is basically observed output. So, for each input here, we will be able to obtain the observed output using this method, this concept. So obviously, this observe output is a set of all outputs, initially it is empty.

Now, this is the concept here; now next let us proceed then.

(Refer Slide Time: 21:08)



Having these are the for each input, we will be able to calculate the observed output. Once the observe output is known to us, we will be able to the matching. So, there is a matching. So, this is basically the actual or true output. If the true output matches the observed output Y bar, then we can say that the perceptron is learned correctly. And the output in that case whatever the output W bar is basically the output of the model. So, model is ready.

On the other hand, if it does not matches correctly, then there will be the few conditions are there. We have to update the weight matrix W bar because then the network is not learn properly. So, in this case, what is the learning procedure is that. So, basically we have to change the weight matrix w 0, w 1 and w n, weight matrix are to be to be changed or updated.

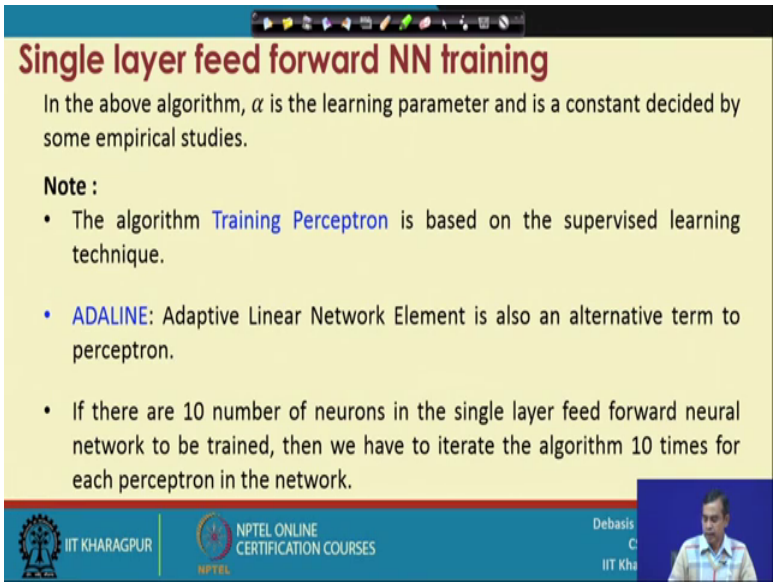
So, that updation can takes place in this algorithm like this for each output y in this object output set ; if the observe output y is 1 instead of 0 actually, then the we can adjust this w 1 output with this formula where alpha is a constant will be decided by the programmer. And for each I, 0 to 1, we have to calculate these values for each x i. So, each weight will be adjusted if we see that output is 1, but it should be 0; instead of 0 it is coming 1. If the output is 1 and it is also coming 1, so no need to do these things. On the other hand, if some other output is 0 instead of 1, then we have to adjust a weight matrix

in this formula. So, it is basically this formula when output is 1 instead of 0 and in this case output is 0 instead of 1.

So, or the weight value will be adjusted and for each input pattern that this is a new weight is there. Now with this new weight, again we can repeat the same thing what we have learned earlier; that means, this step is to be repeated. So, go to step earlier and then repeat again the same thing until we can find the full matches of this one. So, it is a repeative process, it is a iterative process. So, that the same input data can be used with the refine revised weight each time and then we can proceed until we can get it this one.

So, this way the, in this case the w parameters can be checked with one function say output function, it is the transfer function is called the step up function. The same algorithm again can be repeated using the different transfer functions and then again the same thing can be checked whether the output is coming properly or not. And the same the technique can be repeated to decide what will be the threshold function and so on. So, this way for each network parameter we can repeat it and then a perceptron can be trained. So, this is the idea about how a perceptron can be trained or how a perceptron can learn from the supervised data.

(Refer Slide Time: 24:24)



Now, this algorithm, the training perceptron based on the concept of Supervised learning technique and this is also called ADALINE because it is Adaptive Linear Network Element and basically we are in this case we have considered training or learning a

single perceptron. Now our objective is basically to learn all the learn the entire what is called the network basically SLFF SF NN network the Single Layer Feed Forward Neural Network. So, in that case, if the single layer feed forward neural networks counter n number of inputs and n number of layer suppose, we can n number of neurons in the layer for example, say 10 neurons in the layer then we have to learn each neuron at a time and then that learning process is to be repeated 10 times for every for all the perceptrons.

So, learning is basically same approach, but needs to be processed, needs to be computed or each one, one by one. So, this is the idea it is followed there in case of Single Layer Feed Forward Neural Network architecture training ok.

So, this is the concept of Single Layer Feed Forward Neural Network architecture learning we have learnt. In the next class, we will learn about learning the multilayer feed forward neural networks and then we will consider about one technique that is required to learn a neural network technique is called the error gradient descent method using back propagation algorithm which will be discussed in the next few lecture classes.

Thank you very much.