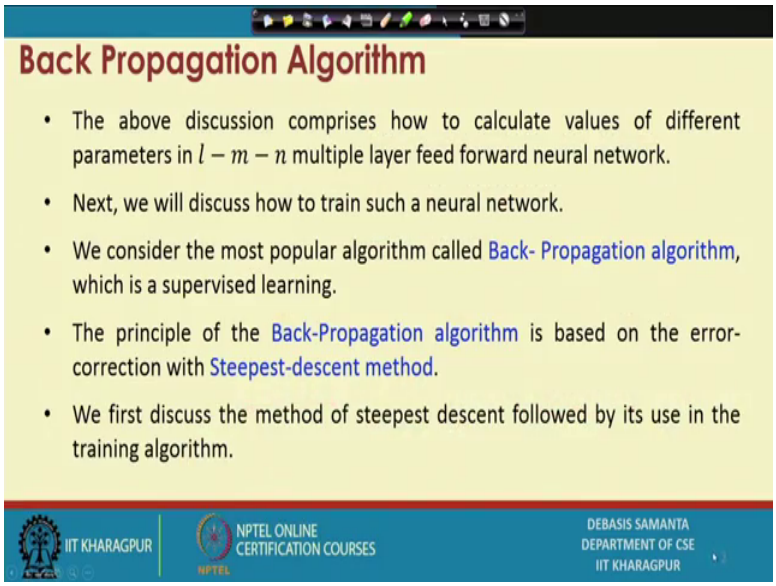
Introduction to Soft Computing
Prof. Debasis Samanta
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Lecture – 38
Training ANNs (Contd.)

So, we have model neural network, more precisely multilayer feed forward neural network and is a simplistic version of the multilayer feed forward network is element, where l is the number of neutron neurons in the input layer m is a number of neuron in the hidden layer and n is the number of neuron in the output layer. Now, after modeling the element network and we have learned about that how such a network can be model and that model can be represented in the form of a matrix.

Now, we will discuss about once this network is model then how it can be learn the different values that is there in the model. So, in particular we are going to learn about V and W matrix that is there in the model.

(Refer Slide Time: 01:13)



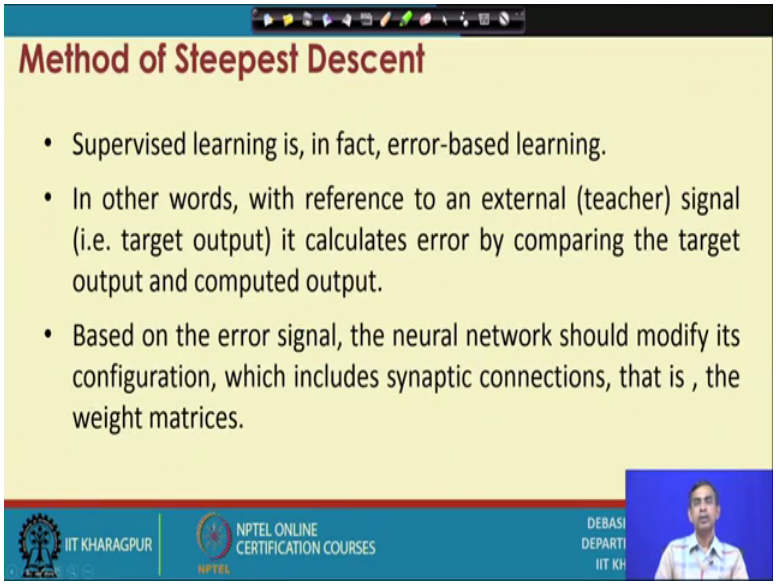
Now, so, the training algorithm that we are going to discuss is called the back propagation algorithm. So, it is a most popular neural network training algorithm and this algorithm is based on the concept of supervised learning. In this algorithm the basic concept that it is followed is basically error correction that means, whenever some input is given to the network it will produce an output. So, it is called the observed output and

as a supervised learning we know for each input what will be the true output. So, error is basically the difference between the true output and then observed output.

So, this propagation algorithm, back propagation algorithm try to correct the errors; that means, it will train the network in such a way that the error that it will be obtained for a given input is as minimum as possible. Now, so, it is basically error minimization technique and error minimization technique which is followed here in back propagation algorithm or there are many error minimization technique of course, like say a least square mean least square error method, but here we will discuss about one error minimization method it is called the steepest-descent method.

Now, here I one thing you can notice that this back propagation algorithm is nothing, but finding the values of the different neural network parameters which basically minimize the error. So, it is basically an optimization problem; that means, we have to the objective function is to minimize the error value; that means, for a given set of input it will find the neural network parameters like V, W, l, m, n, theta, transfer function and everything so that the error is minimum. So, this is the optimization problem in fact, so, back propagation is although also that is why it is called an, optimization problem it, right.

(Refer Slide Time: 03:28)

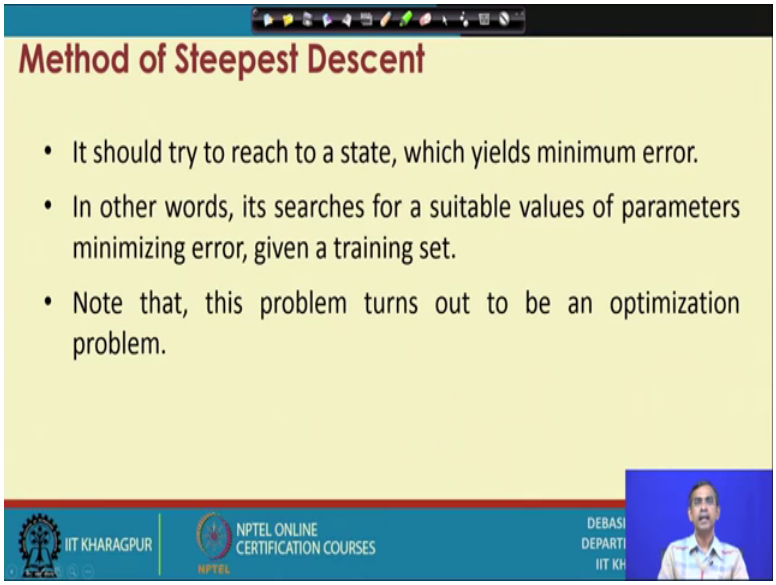


Now, first before going to discuss about this back propagation algorithm we have to learn about the what is the steepest descent method and see the this supervised learning is

always error based learning and so, there is an error I already told you the error is basically difference between observed output and then the true output that is also called a target output and then computed output. Target output means it is a true output and then the resultant output the output which obtained from a given network is called the computed output.

Now, it will sense what is the error magnitude. So, based on the error magnitude the neural network should modify its parameters, its configurations. So, that is the concept that is followed. Now, again for the simplicity of the discussion we will not consider all the neural network parameters to be calculated we will only consider the calculation of V and W matrix as the neural network parameter.

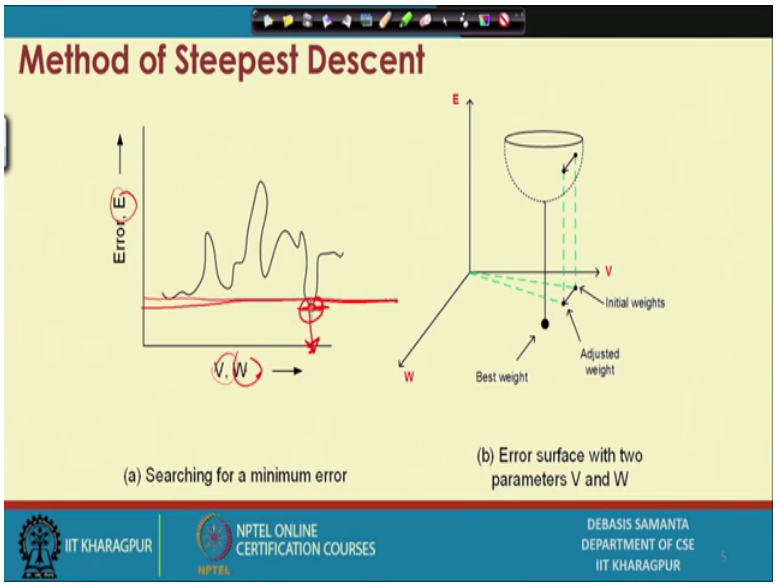
(Refer Slide Time: 04:42)



So, now let us see what is the steepest descent method, that is there. So, basically idea it is that for the input it will produce the output and then thereby the error will be computed and then this error is basically is a function of the neural network parameter values.

So, we have to set the values or we have to search for the right value of the neural network parameter so that the error that can be obtained is minimum. So, this is the concept that is there.

(Refer Slide Time: 05:12)



Now, let us see what is the steepest descent method; it is basically here. So, as it is an optimization problem and if we consider V and W are the input parameter. So, for the different values of V and W we can have the different errors. So, this can be represented by this kind of variation that how the error matrix is varies with the different values of V and W.

Now, so far the minimum error is concerned so, this is basically the global minima. So, we have to find this is the V, W value so that this gives the minimum error. So, basically we have to search the V and W value over the entire search space and you have to find this point. So, that this V, W, it is just like a genetic algorithm concept also. So, that means, for the different values of V and W we have to find one V and W value. So, that the objective function here e is minimum now this concept is followed here, but it is a little bit in a vector representation that steepest descent is basically considered the vector representation and a concept is like this.

So, suppose at present this is the neural network parameter then from these values we can go anywhere in this any direction, but if we come to this value then this is basically is called a error. So, it is basically how the error varies with the different V and W suppose, it is represented by this formula. So, so it is basically error versus V, W. So, it is basically the V, W space and then error is there and suppose the error is represented by this. Now, so, at any instant if the neural network parameters are presented by this point

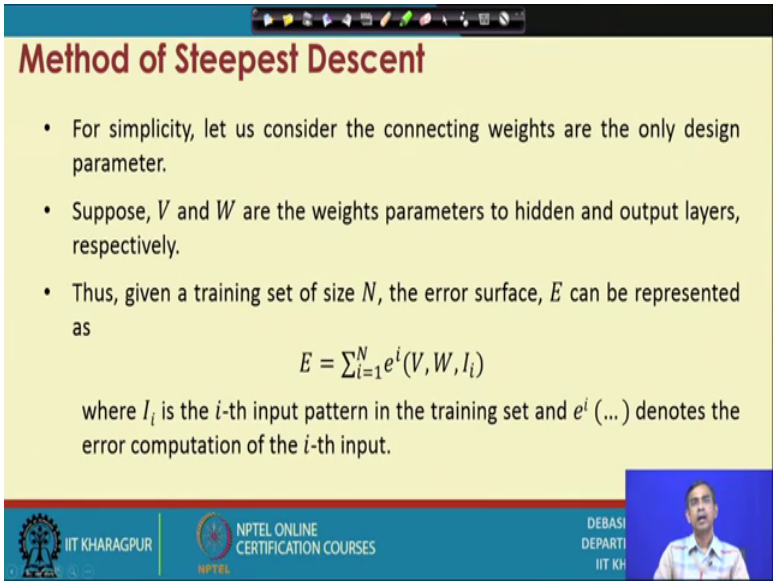
then we have to find the next what is called a modification so that this modification will go towards this minimum value.

But, it cannot go at one step from this to this one rather it is an is small incremental steps. So, from this V and W value it will come to this V dash and W dash values and then from another V dash W dash value then ultimately it will V. So, this is basically incremental step and in each increments it will basically leads towards the minimum plato or it is called a minimum values of the objective function. Now, so, this way so this is basically the value of correct V, W value if at any present this one we have to search these values V, W out of this entire search space.

So, this is the problem and steepest descent method directs the network that if at any instant this is the V, W value then what will be the input what will be the values V dash W dash at the next incremental level. So, steepest descent method tell about this idea and it is followed here. So, if this is the current weights then it is the adjusted weight at the next step. Now, we will see that how from the current weight and adjusted weight can be obtained; that means, from the current weights V and W, how the adjusted weight V dash W dash can be obtained so that it will towards the minimum error.

So, this is the concept that is there in the steepest descent method and we will discussed about the method the steepest how this steepest descent method the concept can be implemented and then how this back propagation algorithm it is.

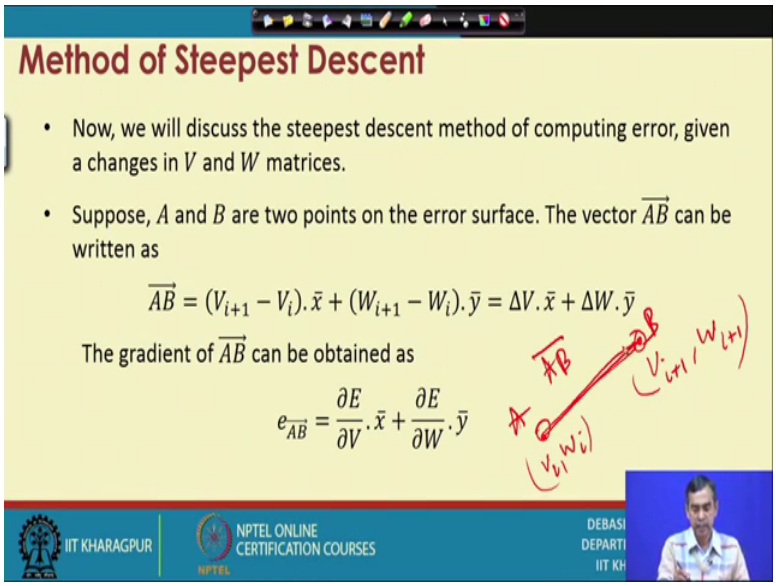
(Refer Slide Time: 08:35)



So, V and W are the network parameters therefore, the network parameters here V is basically the weight values to from the input layer to the hidden layer and W is the from hidden layer to output layer, right. So, in our in our context in this context we can represent the error function E. So, error function E is basically is a function of a three parameters the first parameter V and W because this error depends on the values of V and W and also the error depends on the input the input to the network and if e i denotes the input due to the i-th if e i denotes the error due to the i-th input to the system then the total error E can be expressed summation of all the errors due to the all inputs.

So, here N is the size of the training data if it is there then it is basically summation of all the N number of inputs that is there in the training set. So, this way the error function E can be described. So, error function E is therefore is a function of V, W and the input. So, this is the idea about representing error and then once the error is represented in terms of V, W and I i we will be able to calculate error function and this error function E are to be minimized for a certain values of V and W.

(Refer Slide Time: 10:33)

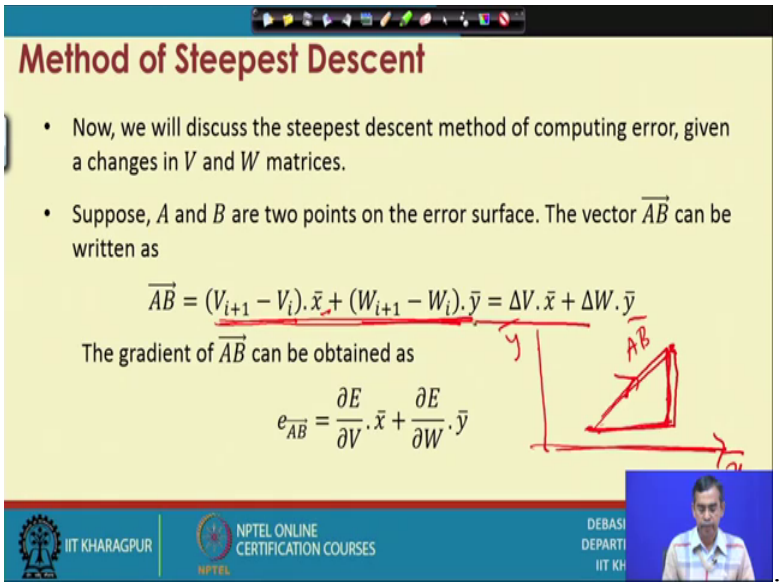


Now, we will discuss about the steepest descent method how it basically search the right values of the neural parameters in the than nn parameters so for the minimum error is concerned.

Now, it required little bit the vector concept. Now, here suppose A and B are the 2 points. So, A is suppose V i and W i it is in the 2 dimensional space of the V and W and this is V

i plus 1 and W i plus 1 are the next point, so, two points. Now, if the two points are given then we can represent this vector. So, here suppose it is A and it is B then the vector AB can be represented like this one. So, so this vector representation also can be represented in a vector form like this one.

(Refer Slide Time: 11:33)

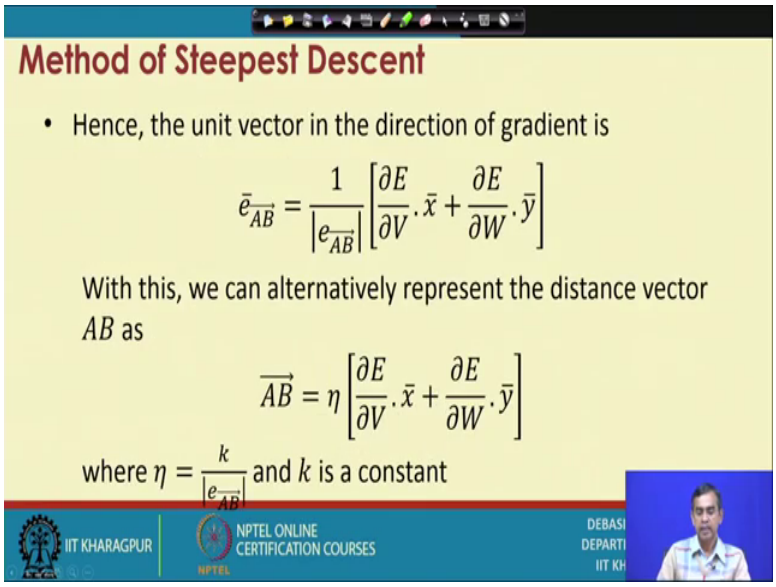


So, it is basically V i plus 1 it is basically this point and this is basically this is AB AB and this is V i, W i and V i plus W 1 plus. So, this plus this is basically represent this. So, here V i plus minus V 1 minus this one into x it is if it is x direction x and then it is W i plus 1 minus W 1 this one so, it is this one and if it is a y direction. So, it is basically x and y if the two dimensional space then any vector in the two dimensional space having the coordinate axis x and y can be represented this form. So, is basically the representation of a vector with in terms of true points V i and V i W i and V i plus W 1 in the search space and this can also be in a compact way can be represented this form where del V is same as this one and del W is same as this one.

Now, So, this is a vector representation if the vector representation is AB, then we will be able to find the gradient; gradient is basically called the slope, the slope of a vector AB we can represented by this e AB it is basically the gradient formula del e by del V into x plus del E by del W into y, it is basically the error gradient.

So, it is basically if AB is the any vector then it is gradients will be represented by this form. Now, let us see how this concept is basically useful in the gradient descent method here.

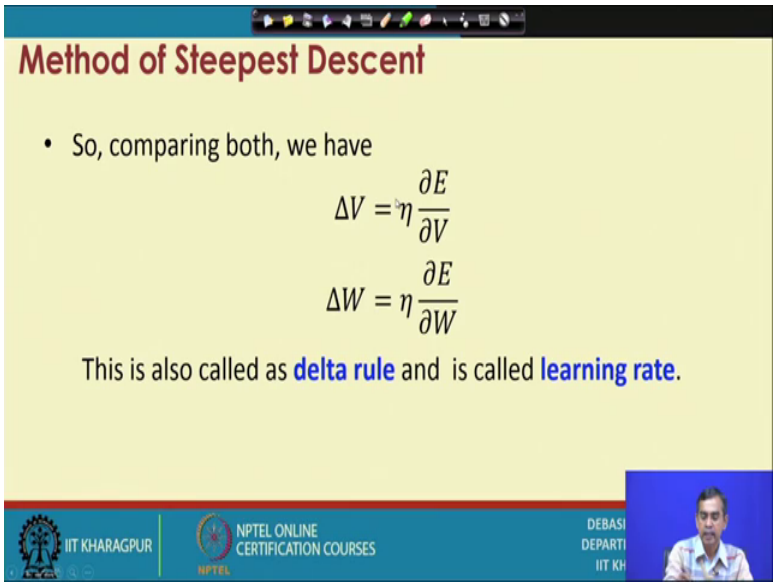
(Refer Slide Time: 13:11)



Now, if e AB denotes the error gradient and then unit vector; that means, the unit, that means, the unit, that means, a unit vector for that a way a gradient can be represented by dividing the magnitude of the vector. So, it is basically the unit vector representation of the gradient vector. Now, so, and then if it is a unit vector it is the unit vector this unit vector can be multiplied by any scalar quantity give the vector itself.

So, every vector which we have given an representation in terms of V i, W i and V i plus one W i plus 1 also can be represented in terms of his gradient representation, where eta is basically the constant which is like this and these are the gradient formula that we have discussed about it.

(Refer Slide Time: 14:20)



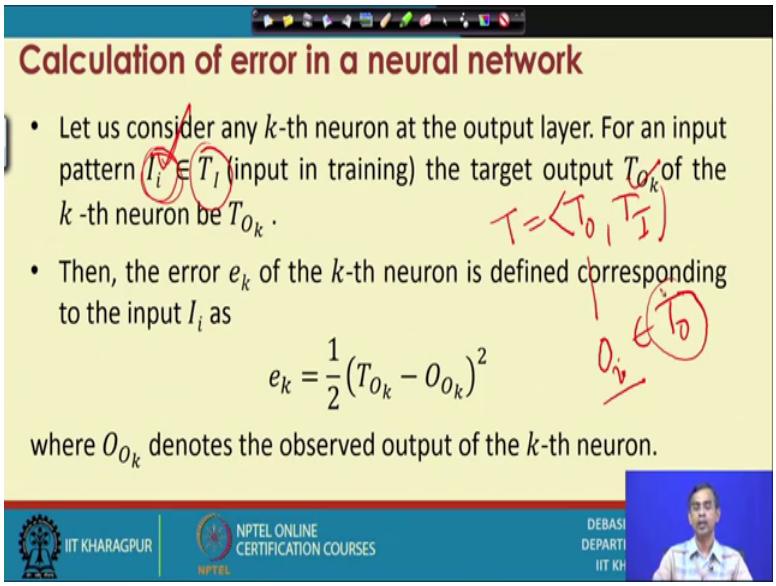
Now, so, having this is a representation and considering AB that we have discussing the previous vector representation and comparing this one we can readily write about this calculation as del V equals to this del V equals to del V equals to this one and del W equals to this one.

Now, this is a very important one observation of formulation. Particularly, this rule that we have just now obtained del V means what will be the change in V vector and del W means what will be the change in the W points. So, V and W are the two points then the changed either increase or decrease whatever it is if we represent it by del V then del V can be represent if E the error is known to there and then it is basically as a is a differentiation the first order differentiation of this error E with respect to V multiplied by eta. So, eta is a one constant, this constant is called the learning rate.

So, here eta is basically learning rate and this constant will be decided by the programmer. If the eta value is not chosen properly then network can learn incorrect way and incompletely. So, the eta value needs to be chosen very properly and it is the programmers responsibility.

Now, we have learned about delta rule and delta rule is basically the steepest descent method concept. Now, we will see exactly how the back proportion algorithm follows this delta rule to calculate this del V and del W value, that means, the network parameters.

(Refer Slide Time: 15:58)



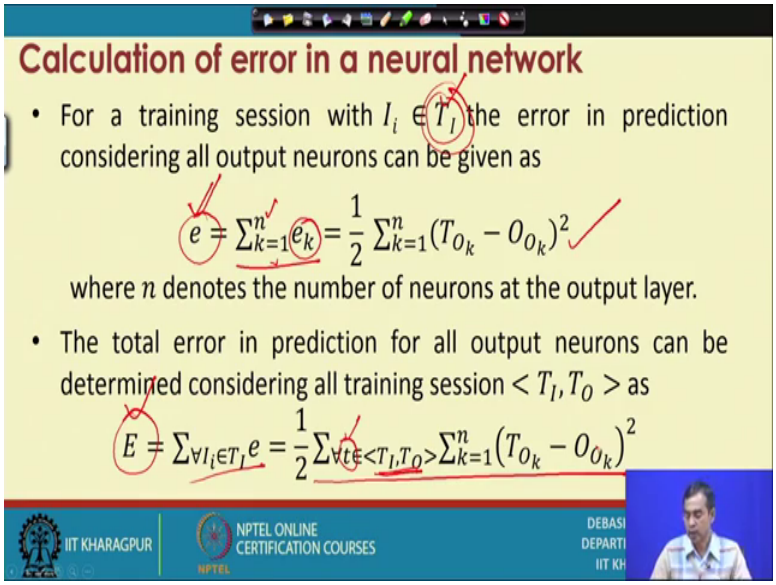
Now, so, we are basically discussing the calculation of del V and del W, that means, the change or updated values of V and W matrix. Now, as I told you it basically based on the error calculation now, let us see how the error can be represented. Now, we will consider any neuron let it be k-th neuron at the output layer. Now, for any input in the training set we decide that say I suffix i denotes the input which is there in training set it is a i-th input that belongs to the input training set and as you have already mentioned that this training set T is decided by T O and T I. So, I i is belongs to the T i and corresponding to this I i there is one O i which is there in T O.

So, it is basically the observed output due to the i-th input which is there in that T O sets. So, this is basically the target output there. Now, we can this way we can write say T O k denotes the target output; that means, the true output of the k-th neuron which is there in T O sets. So, this is the notation that we will follow it and then once we know the true output and then observed output of the k-th neuron in the output layer we represent or the back propagation algorithm represents the error of the k-th neuron is by this formula, it is basically average of the difference the square of the difference of the true output and then observed output.

So, true output and then observed output. So, e k is basically is the formula to calculate the error of the k-th neuron. Now, so, this way for all input that is there in T I can be known and hence the T the true output also can be note and therefore, for each input and

then for each perceptron the error can be known the error can be calculated. Now, this error can be used to calculate the error function e.

(Refer Slide Time: 18:34)



Now, we will see the error function e here. So, here basically the total error e it is denoted as this small e, that means, it is summing up for the errors from the all neurons that is there in the output layer. So, k equals to 1 to n, e k basically the error all errors this basically calculates all errors that can be obtained for a given training for a given input that belongs to a training sets and then in sum.

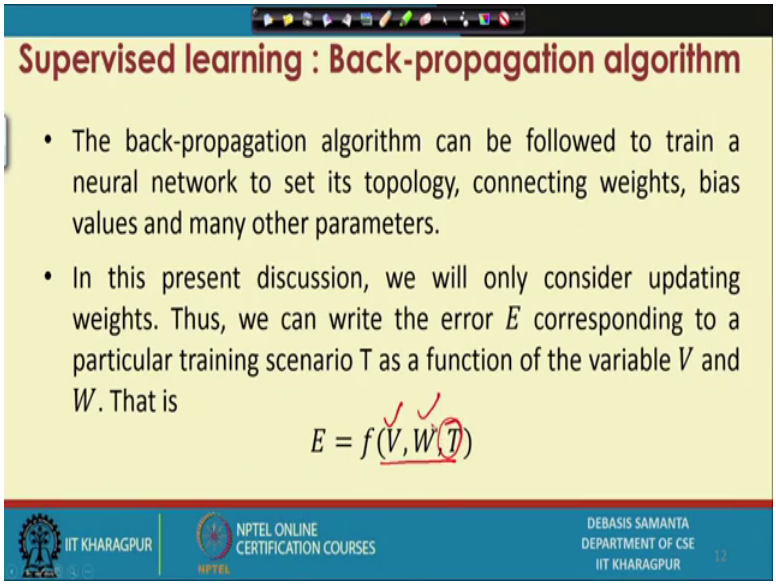
So, this is basically the summation form nothing, but. Now, this is the error for one input which is there in T I. Now, T I contents there are many other inputs also. So, considering all the inputs which is there in T I and if we take the sum of all the errors that can be obtained, then it will give you the error. So, this basically, the formula of error function that can be obtained which can be written in this form.

Where, T is basically the training sets that is there, for all T belongs to for all inputs that is belongs to these sets T is basically a training data that can be belongs to these sets and taking the summation of all the neurons for all inputs and then this one then the error function can be calculated. So, what I want to say is that given a training set to T I and knowing the knowing the output of the network we shall be able to calculate the error function at any instant and as I told you the steepest descent method is basically is to find

some values of the neural network parameters, for example, V and W parameter in our consider case so that this error value will be minimum.

So, now, we have learn about how the error of a network can be calculated. Now, this error calculation is a one important thing and now, we will see how this error calculation is useful and then the back propagation algorithm it is.

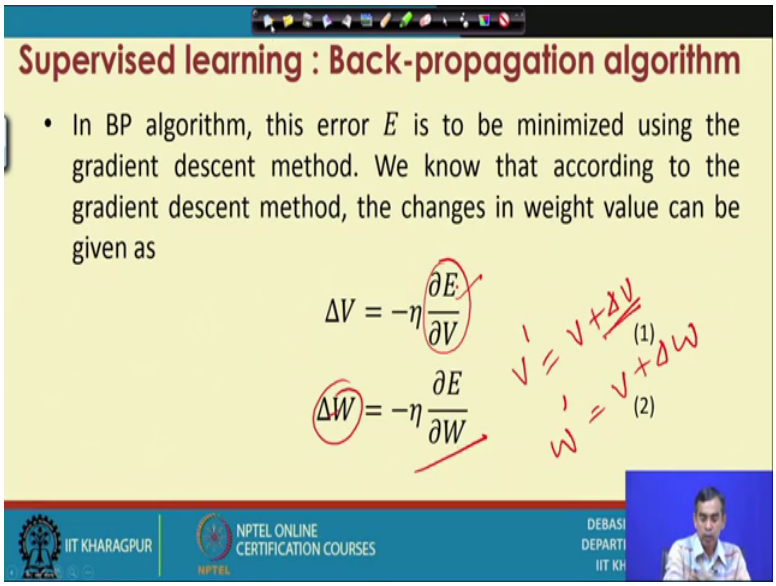
(Refer Slide Time: 20:43)



Now, so, in summary what we can say is that in summary the error E is basically a function of error E is a function of V, W and then training sets and then how this function look like we have discussed in the last slides. So, for given values of V and W and then for the given training sets, E can be calculated. So, in our again I repeat in our procedure or in the back propagation algorithm is to find the V and W value in such a way that for this training set T the error E is minimum.

So, it is basically minimizing error and finding the V, W and hence the neural network will learn it.

(Refer Slide Time: 21:53)

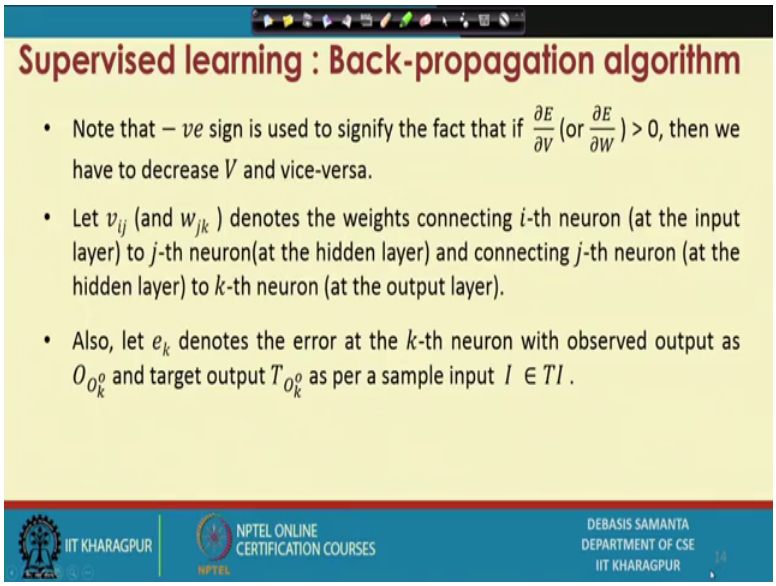


So, essentially it is a optimization problem, that means, how the V and W value can be considered. Now, let us come to the discussion of how this concept that mean error can be back propagated and then back propagation algorithm. Now, here is the idea about we have discussed about delta rule. So, this is basically the delta rule that can be followed there. Now, we can note that one negative number is used. Now, this negative number is for the function is that if error is increases then this value needs to be decreases.

So, that is why if it is increases and it decreases so, the opposite sign is used. On the other hand if it is decreases then it values to be increases so, negative sign is there. So, that is a convention that it will be there. Now, once this del V value is known to us then the next value V dash is basically V plus del V value. So, this del V value will be either incremented or decremented that depending on this what is called the slopes or gradient or first order derivatives if it is increases then it should be decreases if it is decreases it should be increase and vice versa.

Similarly, W dash the next the weight value if at any instant the V value is there then del W where del W can be calculated using this formula. So, these are the steepest descent method by which how E changes with V and then knowing this changes how the updated value updated value can be considered and then the revised value of the V or W matrix will be obtained. So, this is the idea that is followed here in a steepest descent method.

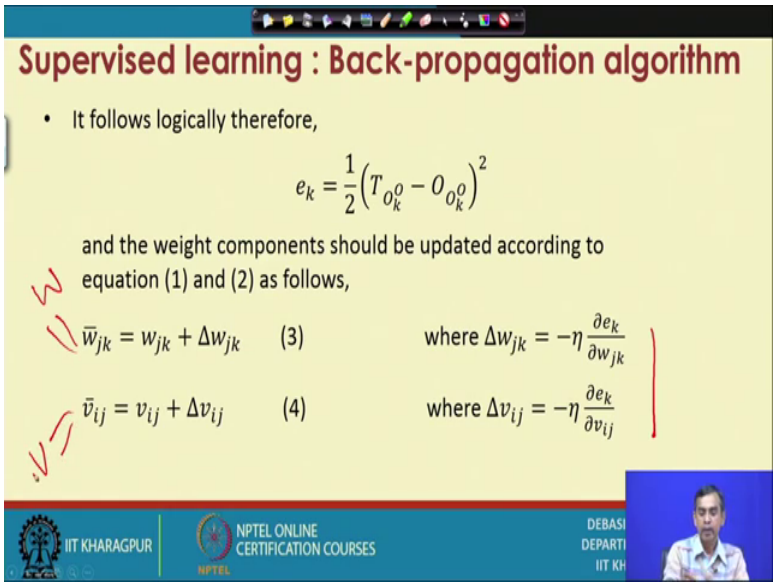
(Refer Slide Time: 23:42)



And, the delta rule is the most important useful is there as I told you the negative sign is to signify the fact that if del E by del V the first order derivatives of error with respect to V or the first order derivatives of E with respect to V is greater than 0, then we should decrease V and vice versa. And, v ij this is our usual notation that we have discussed about v ij and w ij denotes the weight values connecting from the i-th neuron in the input layer to the j-th neuron in the hidden layer and then from the j-th neuron in the hidden layer to the k-th neuron into the output layer.

And, e k denotes the error at the k-th neuron which is observed output is a difference sum is basically half of the sum of the different square of the observed output and the true output for each input i.

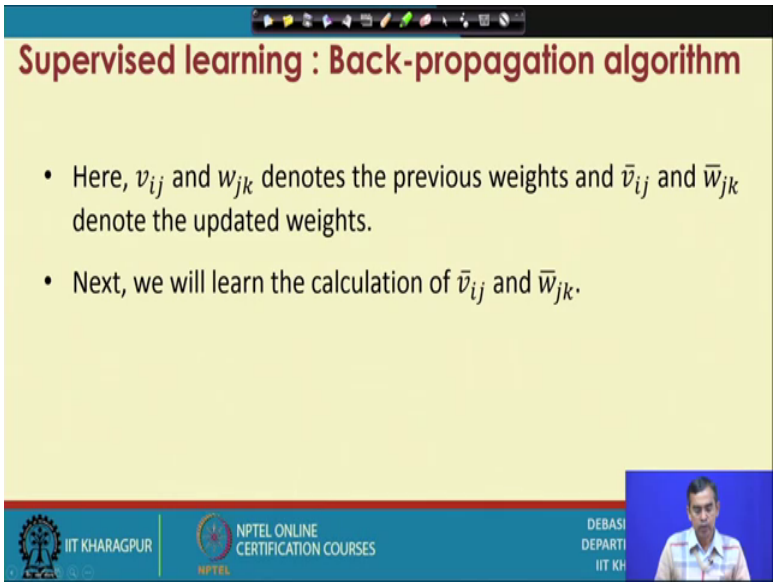
(Refer Slide Time: 24:54)



Now, now so, this is the concept that is there and ok, with this concept we will be able to learn about back propagation algorithm quickly. Now, as I have all learned about so, e k is basically the error of the k-th neuron in the output layer and we know that it is the formulation and if we know the error of the k-th neuron in the output layer then we shall be able to know what is the in incremented value of the del w jk. So, it basically if this is the current value of w jk or current value of v ij, then using this delta rule we will be able to calculate del w jk and then updated value del is a updated value this one. Similarly, using this rule we will be able to calculate this one and then updated value will be there.

So, this is the delta rule and using this delta rule we will be able to calculate the updated value and if we know any j-th and k-th weight or i-th and j-th weight we will be able to know them for the entire V and W matrix because this is basically give the W matrix and this is basically gives the V matrix. So, this is the idea that is followed in the back propagation algorithm and once the error is known to here then we will be able to find the updated value.

(Refer Slide Time: 26:10)



Now, how the updated value can be calculated. Back-propagation algorithm suggest one very clever method, this method is called the chain formula or it is called the chain formula in a back way. So, is starting from the input a starting from the output then go to the next previous layer in the next layer in the previous layer and so on so on. So, we will discuss about this is the propagation algorithm that is why it is called a propagation algorithm and it is propagation from the back.

We will discuss about this back proportion algorithm in the, next session, ok. So, back propagation algorithm we will be discussing the next session and we learn about how that network can be learned tool to if basic network can be trained to learn the values of V and W parameters values.

Thank you.