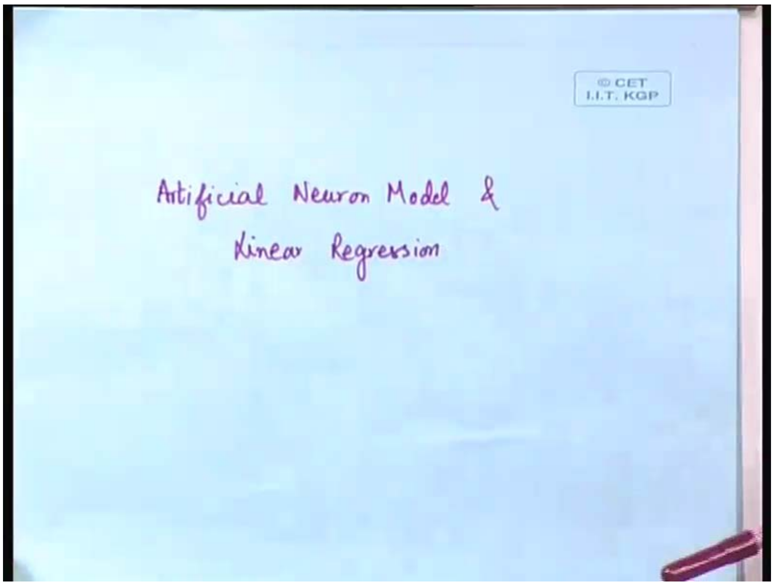
Neural Network and Applications
Prof. S. Sengupta
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 02
Artificial Neuron Model and Linear Regression

Artificial Neuron Model and Linear Regression.

(Refer Slide Time: 00:55)



Now, in the last class we had given you an introduction to the biological neurons. What we discussed there was that, the human brain is able to do enormous amount of computations. And in a very short time utilizing the billions of nerves cells or the neurons, which are existing in our brain. And as I explained in the last class that, these neurons are interconnected to each other in the form of a network

In fact, a very complex network and in it is not that we understand fully about all the different connection mechanisms of biological neural networks in a complete way. Whatever we know is mostly from the ((Refer Time: 02:08)) perform, and this is mostly the scientists observations and believes, which has lead to some kind of a model making for the biological neurons. And essentially it is the biological neuron model, which we will have to adopt in order to do the computations for our purpose.

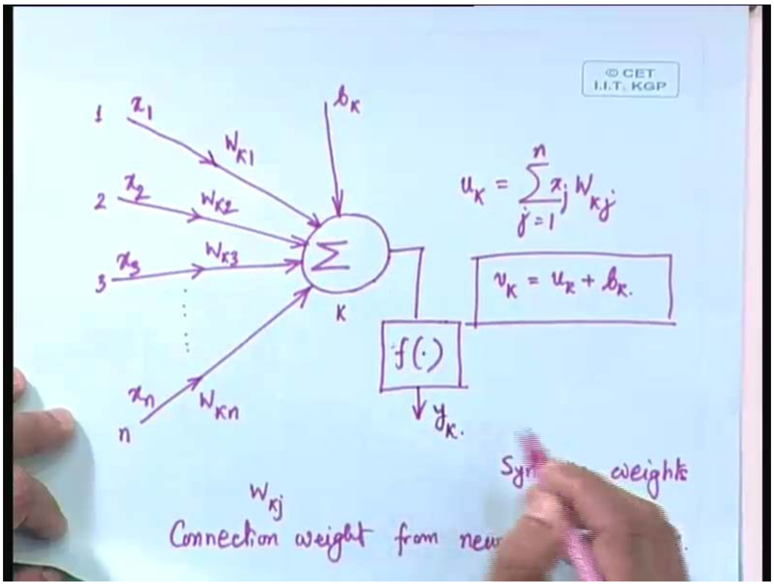
So, what we are considering is the artificial neuron model. And the name artificial itself means that it is different, from that of the biological neurons. The first thing is that we in a typical artificial neural network structure, we are not going to consider so many number of neurons, it is only small manageable number of neurons that we are going to consider. That too for solving specific problems and as I told you that if I have to list out the different applications of neural network.

Then that list is never going to be complete, because we can imagine the applications of the neural networks in any application domain. I mean be it science, be it technology, be it any type of market forecasts, weather forecasts aim many applications. Signal processing, biological applications, where we can apply the neural networks in a very big way. And in fact, neural networks along with the fuzzy systems, there are going to be very popular for the domestic appliance controls. I think all of you must have heard about that the neural fuzzy controllers.

So, there are large number of applications which one can think of, now coming to the model of the artificial neuron it needs some introduction. Now, as I was discussing in the last class that, the biological neurons are interconnected to each other through what is called as the synapse. So, to mention that it is the synapse which is basically deciding about the strength of the connection, how strong or how weak those connections are.

Typically and nerve cell or a neuron could be connected to several other neurons, or it might be receiving inputs from several sources.

(Refer Slide Time: 05:00)



So, if we look at a neuron in isolation, supposing this is a neuron that we are going to consider. And there we are basically receiving inputs from a number of sources. So, there are plenty and I am just drawing a few, this may be the input number 1, this may be the input number 2, input number 3 and so on. We may be having n such inputs, so I just happened to number them from 1 to if n. And they are connected to this neuron which is going to do the process.

Now, this neurons are connected to these ones using the synapses. Now, synapses are existing in the biological neural network as I explained, but a very similar concept. we drew up for the artificial neurons also where we model the strength of the connections this way, so we call the strengths of the connection as the synaptic weights. So, the strength of the connections will be known as synaptic weights, weights they decide that how much of strength it is.

So, larger the value, more strong the connections are smaller the values less strong they are. In fact, there is no binding that whether synaptic weights, must be only positive or not, as a matter of fact it could be positive as well as negative. Because, even for biological neurons also or biological thus the synapses associated with the biological neurons, which we discussed in the last class. We said that they could be either excitatory or could be inhibitory image.

Now, we will be having such synaptic weights which will be associated. And those synaptic weights will be defined for the connection between the input and the neuron under consideration. Now, mind you the input may be received from other neurons also, or the inputs may be coming directly from any of our perceptual units. So, no matter whatever the sources of the inputs are, basically we are receiving the inputs from some source.

And there are a large number of neurons existing and let us pickup small, one particular neuron out of in, and let us say that we index this particular neuron by the letter k. So, this is a particular neuron k, that is under consideration and this is receiving a number of inputs. And these are having connections or the synaptic weights, which will be written as W k 1, W k 2, W k 3 and so on and this one will be W k n.

So, they notations that we are using is that W k j, basically indicates the connection strength from the neuron j to the neuron k, if you are considering all these inputs to be neurons, which in typical cases it will be like that. So, W k j will indicate the connection weight from neuron j to neuron k. So, please note this order it is not k to j, it is not always the case that the connection which is existing as W k j will be equal to W j k. In fact, the feedback path may or may not be existing.

So, the way we are indicating this does not mean that there is any connection between k and n back or k and 3 back. If there is a connection, then we will be indicating that by a separate way, for some applications W k j may be equal to W j k. But, k j always means connection from the neuron j the superscript that we are writing after and k the superscript that we are writing before, so k is this neuron.

Now, when these sort of strengths are defined and we can define the inputs as x 1, x 2, x 3 and so on, up to x n. Now, we are linearly combining all these things. So, in effect what it means is that, we are summing up all these inputs multiplied by their appropriate weights. So, that the sum total, which we can is call as the u k, u k will be equal to summation of j equal to 1 to n W k j and this has to be multiplied by x j.

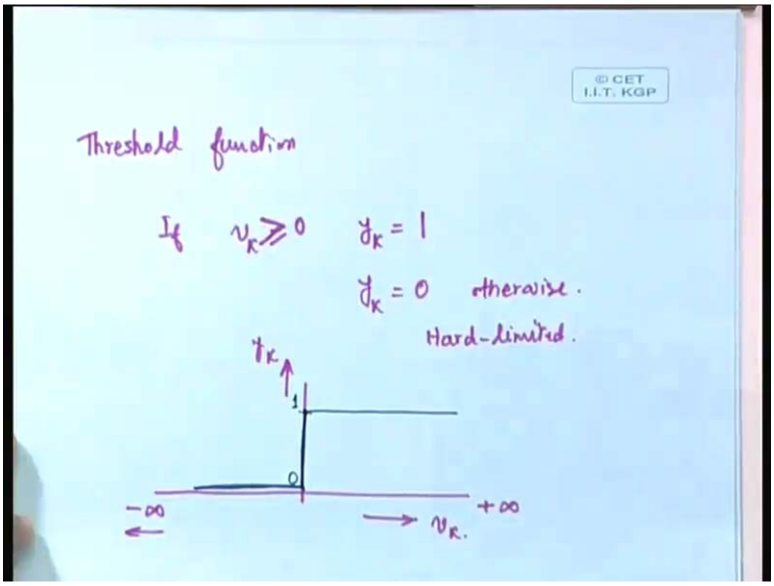
So, it is summation of x j W k j and the summation will be from j 1 to j equal to 1 to n, is this clear. So, we are adding up all these n different weighted inputs, effectively linearly combining them together. Now, to this we often required to, so this will give us some kind of an activation value, some kind of weighted addition that we have made. And we may like to move this whole thing up or down, just like adding some DC offsets to it.

Those who are familiar with the operational amplifiers the known that whenever we are combining some inputs to the operational amplifier using it has a sum, we can add some DC also to it. So, that the whole summed input can be raised up or if the DC is added in the negative direction, the whole thing can be brought down. So, it is a offset or what we can call as a bias that can be applied. So, we can in fact say this bias, that also will be used as an input to this system.

So, this bias I can call as b k, again the subscript k for u k and b k means, that they are pertaining to the neuron k over here. So, that the combined output of this will be v k equal to u k plus b k, so it is the bias with that we add up all these weights, weight multiplied by this inputs. Now, the thing is that, this is the v k and we are interested in finding that what kind of a response is this neuron going to give us, that is very important.

Basically it would determine that only we have considered this artificial neuron model. Now, there are several ways whereby we can define this function and the simplest of that is to consider a threshold function.

(Refer Slide Time: 13:31)



The threshold function can be defined like this, that you see that the combined output from all this is what this one, v k equal to u k plus b k and u k is this summation and b k is the bias that we have added. Now, we can say that this will result in the activation like this, that if v k is equal to, is supposing greater than or equal to 0. Then we can say that, we are going to design the neural network or design this neuron in such a way, that followed by this, if we take this output.

And we just pass it through some function, let us say this is a function of v k and this will be the overall output. So, we are going to define this function now as a threshold function, how that if this v k that is the total response over here is greater than or equal to 0, then the final output if we can call this as y k the final output of this. If v k is equal to 0 is greater than is equal to 0, we define that y k is equal to 1 and y k is equal to 0 otherwise.

Otherwise means, that when v k is less than 0, that means to say what that v k can assume any value in the range of could be a positive, it could be a negative value. And in this case it have not really put any binding on the value of v k, v k could anything theoretically I should say it can go from minus infinity to plus infinity. But, the output is hard limited we call this as the hard limit, so the output is hard limited to 1 or 0.

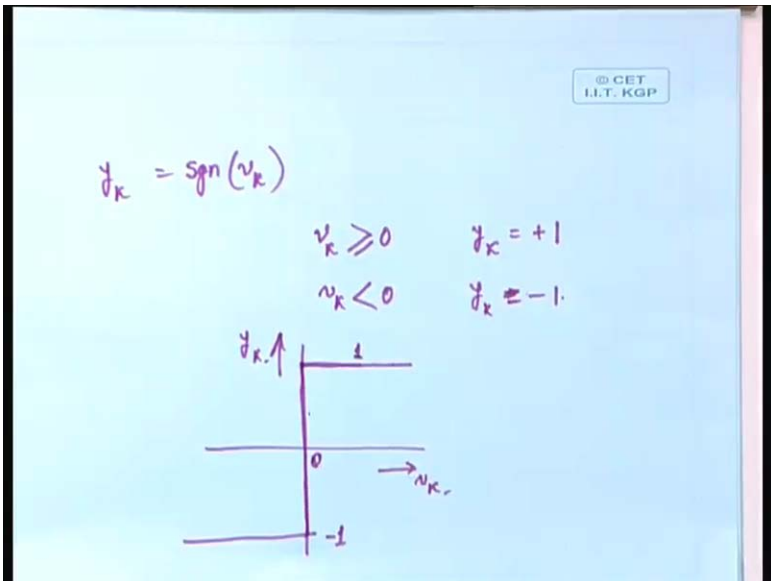
So, if you plot the response of this threshold function that, would look something like this, that on this axis we plot v k and on this axis we plot y k. So, this necessarily means that at v k equal to 0, we are going to see a discontinuity in the function definitely, as long as v k less than 0. We are going to have the values of y k to be equal to 0, so this is y k equal to 0. And this is the point where we are going to make the y k equal to 1.

And this is the positive direction of v k, so this is towards plus infinity and this direction is towards minus infinity. So, as long as it is less than 0, y k is equal to 0 and when v k is greater than or equal to 0 with a discontinuity happening at v k equal to 0, the y k is going to be equal to 1. So, this is a threshold function that one can use, in which case the response of this neuron is going to be absolutely binary in nature.

Now, in this case we considered that the outputs are going to be 0 or 1. But, there are some applications, where we do not require the binary values modeled as 0 and 1. We quite often need applications where the output is to be modeled as minus 1 or plus 1.

And if it is minus 1 or plus 1 in that case we have to define this function f to be a signum function.

(Refer Slide Time: 18:16)



In that case, we are going to define the threshold function like a signum function, signum of v k, meaning what that if v k is greater than or equal to 0, then y k is going to be equal to plus 1. So, y k will be signum of v k, meaning that v k is greater than or equal to 0 y k is going to be plus 1 and with v k less than 0, y k is going to be minus 1. So, here the function would look very similar, it will be minus 1 up to here, then a discontinuity at v k equal to 0.

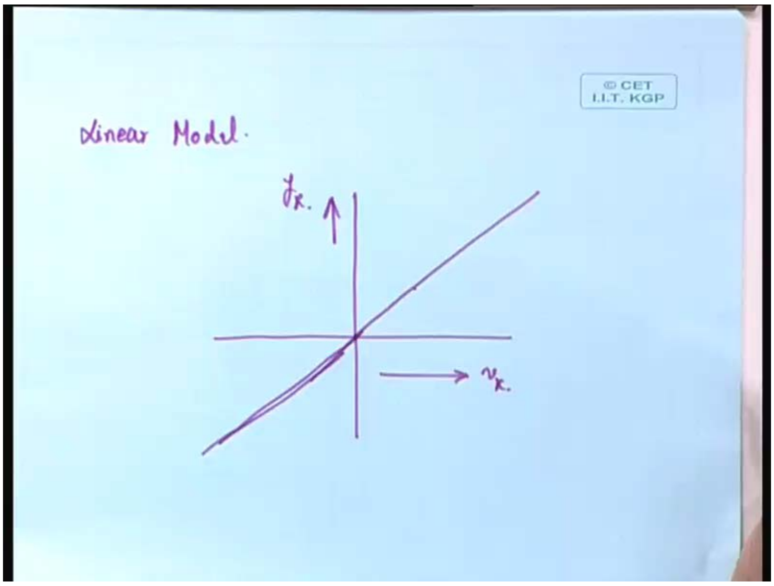
So, this axis is v k as before, this is y k and beyond v k greater than 0 we are going to have y k equal to 1. So, this is the value of 1, this is the values of minus 1 almost the same thing as that of the threshold function. So, this is also a hard limiting only restricting the values to 0 and 1 or minus 1 and plus 1. In fact, this particular neuron model, where the response y k can be either 1 or 0 and it is defined as a threshold function like this.

This neural models are known as McCulloch and Pitts neural network model. So, this was originally proposed by them McCulloch and Pitts, so this is often known as the McCulloch and Pitts neural network models. Now, it is not always mandatory that the neurons will behave in this manner only, because after all the kind of problems that we are going to solve with the neural networks, where a wide range of variations in their response is quite often needed.

So, if we are leaving with the analog world say for example, here we did not put any restriction on the values of this x 1 to x n. They are analog they can vary in any manner you like and the v k, on the v k on the some there also we did not put any restriction. But, we are only restricting the output, but why do we restrict it always, we need not have to do that. So, what we can do is that we can propose a different model also.

Now, for a very simplistic model why do not we consider the linear model itself, by linear model to say that, what is the problem if we do not choose any f function out here. And we take this output directly that means, to say that we make y k equal to v k. If we make that way, then our model gets quite simplified, in that case how will be the input output characteristics pretty simple.

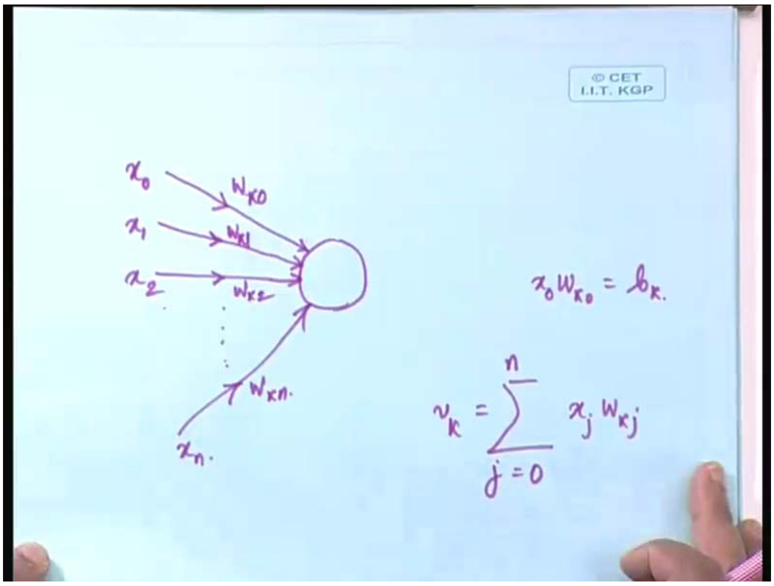
(Refer Slide Time: 21:48)



Because, if we take a liner model in that case, we can plot it like this, this axis v k and this axis y k. So, when v k is equal to 0, we will have y k equal to 0 and when v k is something let us say a value is equal to a, y k also will be having a value equal to a, because we are going to have y k equal to v k only. So, that it will be a perfect straight line with a slope of 45 degree and passing through the origin it will be a perfect straight line. So, that will be our perfect linear model for that.

In fact, just to represent the neural network in the form of a linear model, we make some simplification in it is structure, you see so far we are taking the bias and the inputs separately. In fact, it is possible to include this bias as a part of the inputs only, means that we can model it like this, you see here we have got the numbers ranging from x 1 to x n, they are the inputs. So, we can have an input which we can call as x 0.

(Refer Slide Time: 23:23)



So, x 0 and that will be connected to the neuron and then, we can have the other inputs as it is that is to say x 1, x 2, x 3 and so on, up to x n, this is x n and this weight we are going to call it as W k 0. So, that we have x naught W k naught to be equal to b k. So, what we are simply doing is that, this b k rather than representing as a separate bias connection we are representing it in the form of an input only, after all bias also is an input, but we are just modeling this way.

So, W k 1, so this is W k 0, this is W k 1, this is W k 2 and this is W k n, so that what happens is that when I write the expression for v k, I can simply add it up like this x j, W k j and in this case I sum it up from j is equal to 0 to n. So, that is the only difference, earlier I was calling j equal to 1 to n and then, adding the bias to it in order to realize the v k. And in this case I am simply describing it by a single equation that summation x j y k j.

Now, in this case actually in a linear model, if it is a perfect linear model, then we are not making any restriction about the values which v k and y k can assume. Because, like before we are taking the value of v k in the range of minus infinity to plus infinity. So, likewise y k also can be in the range of minus infinity to plus infinity, but quite often for practical consideration, we tend to put some hard limiting threshold beyond some range.

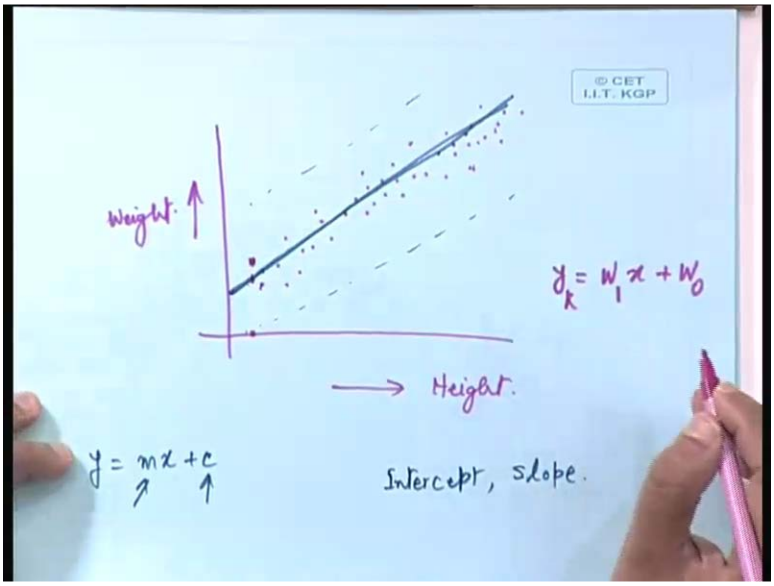
So, if it is such that, from a practical realization point of view, that this is the point beyond which our hardware cannot process the signal, we can put a hard limiter out there. So, that between some range, between this range we can use it as a linear model. So, when we talk about a linear model that necessarily means that, within our expected range of operation it is going to follow a linear response.

It is going to fallow this equation which I have just now mentioned, that is to say summation of v k is equal to summation of x j W k j and summation ranging from j is equal to 0 to n. Now, this is linear model and let us try to see that, what could be the application of that, where are we going to use this kind of a linear model. Basically, we can use it for any data fitting, where we have to fit straight line to a given set of data. May be that we have conducted some experiment and the experimental data is showing a large number of variations.

(Refer Slide Time: 27:26)

Let us say that this is the input, we have noted some input output characteristics. So, that we have set some inputs and for those inputs there will be some output values, may be that when the input is this much we have got output this. May be when the input is this and output is this, may be for input this, so like that there could be a set of points that we can obtain. So, this may be the data that you have collected from any experiment, can be a scientific experiment, can be any experiment that you would like to do.

(Refer Slide Time: 28:24)



Let us take very simple example, supposing you are trying to plot the weight of a person against the height, so on this axis we have weight and on this axis we plot height. Now, it is known that generally the persons having, the persons who are taller they tend to have more weight as compare to the persons who are short. Not always the case, because the weight of a person is not dependent upon height alone. The weight in fact, will be depended upon so many factors, his height, his basic physic, then his food habits, his calorie intake, fat intake, his dietary habits.

Then also how much of calories he burns, how much of exercise he performs every day, on so many factors the weight of a person will be depended. But, we are only studying that how is the weight changing with height. So, may be that we conduct an experiment on 1000 people and we make a plot like that, that against height we plot the weight and this may be set of data that we obtain.

Now, our objective may be to fit a curve, fit a curve that passes best through these set of points, we do such kind of curve fittings in experiments data. Now, whenever we are applying a linear model, we would necessary fit in a straight line to this set of observations that we have made. Now, whenever we want to fit a straight line, then we have got two free choices for that straight line.

What are the free choices, the intercept and the slope, a straight line will be characterized by the intercept and the slope. If we are writing in the popular form of y is equal to m x plus c, m being the gradient or the slope and c being the intercept, then that is the way whereby we can modeled a straight line. And in fact, we got a free choice that we can vary m, we can vary c in any manner we like, that means to say that if we do not have any a priory guess work available.

What we are going to do is to fit a tryout with wide varieties of m and c and finally, coming up with the best type of straight line fitting. We know that the best fit straight line would be somewhere over here, but somebody may like to fit the straight line like this, somebody may like to fit the straight line like this, in which cases we know that it is not an appropriate straight line fitting.

Now, supposing that we have fitted a straight line, we have managed to fit a straight line like this. The best set up m and c we have chosen, from this set of experimental data. Now, there are some points which are lying above the curve there are some points which are lying below the curve, so like that there are a large number of observations that we have made. And if you see, if you take any particular point like say for example, I take this height.

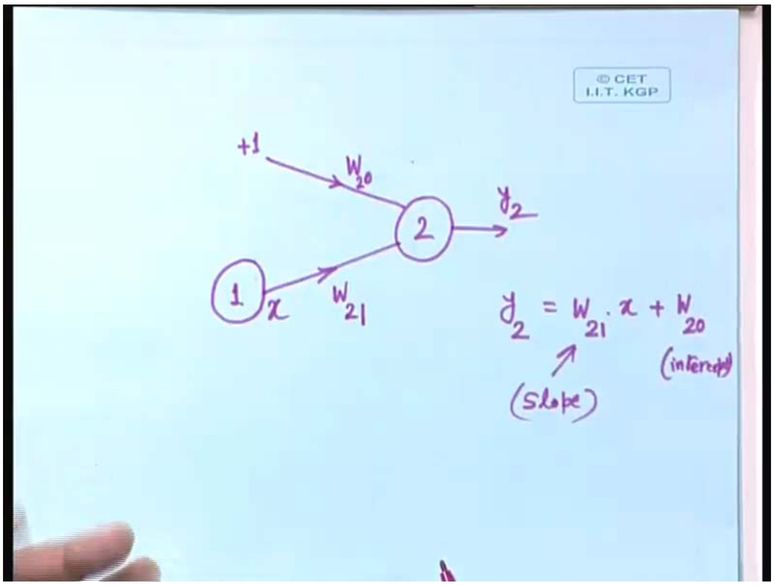
For this height I have got this to be the difference or the error between the actual value, the actual value is this and the fitted value is this, so this is an error. Whereas for this one the error is here, so if I take this to be a positive error, this one is a negative error. Now, I have ways to make the error positive, because if I take the absolute value of the error or if I take the square of the error. And if I add the square of the errors or the square of absolute errors, then I will get a combined errors measure.

And if I divided by the total number of points, that I have used in the experiment I will be getting the means square error also, which is a measure that how good or how bad my curve fitting is. So, in this case also it is going to be a straight line fit, that we are trying to do and very interestingly the equation that we are going to write for this is of the form of y is equal to m x plus c only.

Only thing is that instead of writing it has y is equal to m x plus c, according to our notation, we will write it as y call it y k. Because, again if we are processing this using the k th neuron call it as y k, y k will be equal to w 1 x plus w 0. If I write it in this form what does that mean, then w 0 indicates the intercept and w 1 indicates the slope. What is x, x is the input, x is the input to the system what is the input the height, we are seeing the persons height and against that we are plotting the weight.

So, x is the input to it and w 1 can be defined as the synaptic weight between the input and the output. Now, in this case, that means to say that we will be having only two neurons.

(Refer Slide Time: 34:49)



So, it is the neuron 1 and it is the neuron 2, let us not call it as k anymore, we have neuron 1 and neuron 2. And neuron 1 is connected to neuron 2 by the synaptic weight, which we are now going to call it as W 2 1 or 1 2, 2 1, because it is W k j as we had said. So, we are going to write it as W 2 1 and this it is the neuron 2, which is doing the processing. And here we are going to have an input equal to plus 1 and this will be multiplied by W 0, not exactly W 0 this is W 2 0.

So, that the overall response that is y 2 will be equal to W 2 1 into here the input will be, although I have written it as 1, 1 is the neuron number, but it is input will be x, so it will be W 2 1 into x plus W 2 0. So, that the whole thing is modeled as a very simple neural network, whose response will be in this case y 2, it is a linear response that is being made. And the input to this is x multiplied by W 2 1 the synaptic weight and W 2 0, this indicates the bias.

So, the bias in this case indicates the intercept, so the bias stands for the intercept and W 2 1 that is the synaptic weight that stands for the slope. So, this is the slope and this is the intercept. So, we can so that means, to say that fitting straight line to this set of data, means that modeling it to a neural network. Because, after all what we have to determine, just like the way to fit a straight line, we have to make a good choice of this m and c. Very similarly we need to have a good choice of this W 2 1 and W 2 0.

If we can make a good choice of W 2 1 and W 2 0, then the job is done. And what job is done, like in this case you see that if somebody gives me the height of some unknown person, means this is the set of data that I have gathered during my experiment, during my trials I have gathered this data. So, now that I have got this data and I have fitted a straight line, if someone else data is put in, or if I know height of somebody.

Then I can easily use this curve, in order to say that if it is given that, the height is this much, then the corresponding weight will be this much. That is what we are going to say, because this will be the weight that we are going to interpolate out of it. So, likewise here also what happens is that, once the neural network is trained with some patterns with some inputs. And you feed some unknown input to this system, in that case it will be behaving with whatever weights with which it has already trained.

So, the neural network has already trained itself with some W 2 1 and W 2 0 that means, to say that it has already fitted a straight line like this. So, now if we just feed some unknown input to it, then we will get the corresponding output which will be as per the fitted data.

Student: ((Refer Time: 39:05))

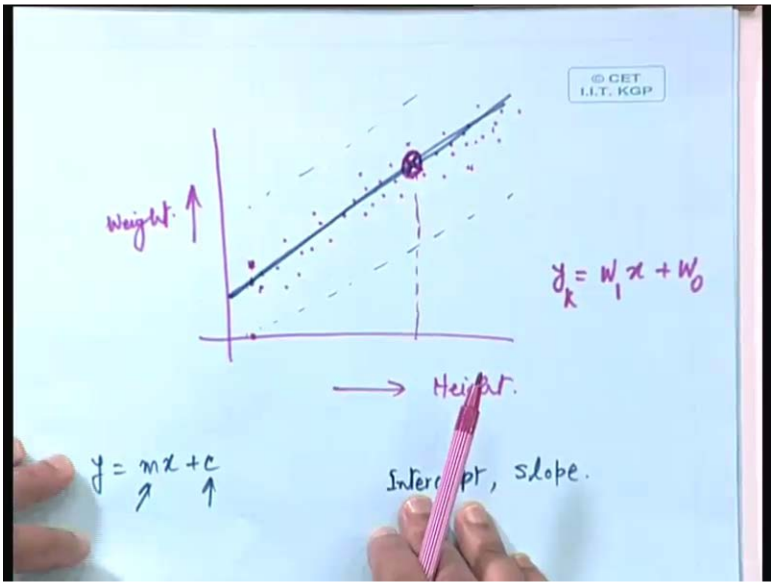
That is right

Student: ((Refer Time: 39:09))

Well, let us not talk in terms of probability at this point, all that we are trying to say is that this is the fitted data. So, because it is an unknown data to it, we can at best conclude that we are going to have this particular point. Now, how good or how bad that is with respect to the actual data that we are not knowing yet. So, it is after conduction a large number of experiment, in fact to be very honest with you, this experiment is not at all a good experiment.

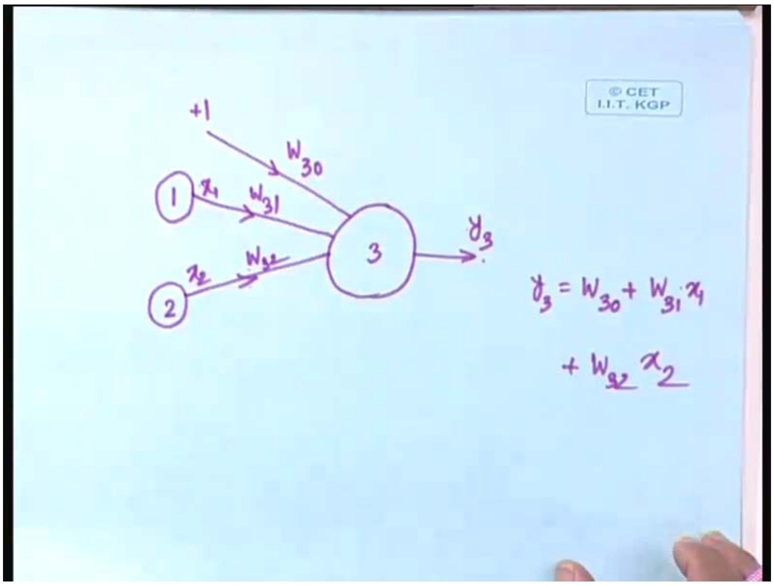
Because, as I was telling you that here if have assumed that the weight of a person is dependent on only one factor that is the height, which is not at all true it is dependent on many factors. So in fact, if we now start listing the different factors to which it is related to.

(Refer Slide Time: 40:20)



In that case, we are going to get different curves like if now for say for example, I plot a curve as weight as plotted against the average calorie intake of a person. Then also I will get a separate curve out of that or all the parameters with which something is dependent upon, if I plot all these together. Then effectively I will not be getting a model like this, but for each one of the inputs, I am going to get a straight line fit of this sort of the nature. So, that ultimately it will be a combined fitting of linear response.

(Refer Slide Time: 41:16)

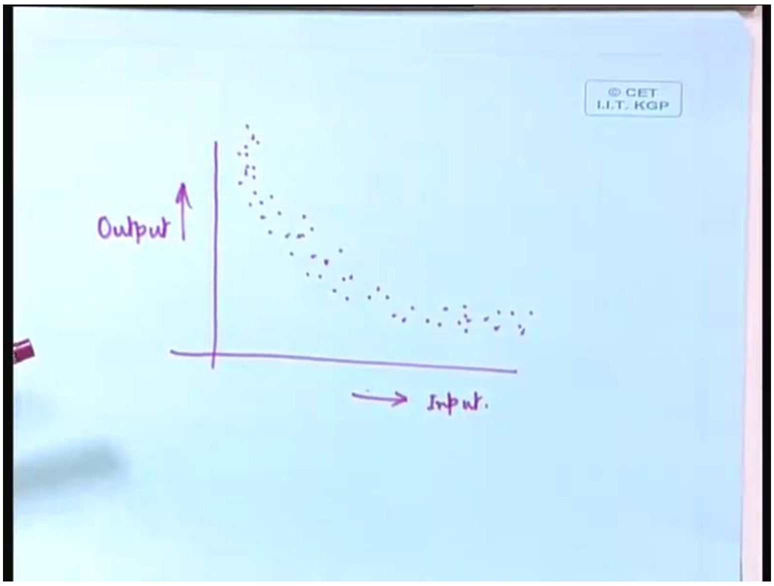


In fact, it will be like this that supposing, there is some output, let us say that the output to this neuron is dependent upon this two set of datas. So, I say that these are coming from the neuron 1 and the neuron 2 and supposing this is the neuron 3 whose response we are trying to find out, and also there is a bias. So, that I say that this is plus 1, so this is W 3 0, this is W 3 0, this is W 3 2 and then, this one will be the y 3.

In this case what happens is that your y 3 will be equal to W 3 0, that is to say the bias plus W 3 0 we call it as x 1 plus W 3 0 we call this as x 2, so this will be the combined output. So, that we get, so now that y 3 is dependent upon two inputs, y 3 is a function of x 1 as well as x 2, so we get two slopes pertaining to that one associated with x 1 and the other that is associated with x 2. So, this is a two dimensional problem now, or rather modeled it as a two input network and like this we can extend.

So, essentially it is a data fitting problem that we are doing and let me also tell you about, it is not that always, it will be linear curve fitting that we are going to do.

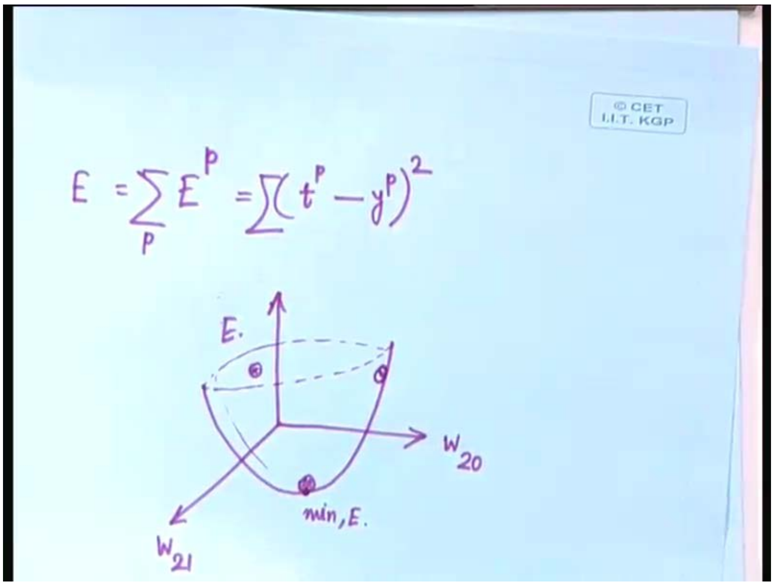
(Refer Slide Time: 43:10)



Supposing the observed data is say of this pattern, Supposing this is the input and this is the output and the observed data pattern is like this, supposing this is the observed data pattern. Now, somebody tells me to fit a straight line on that, we can do the straight line fitting, but the straight line fitting will not be certainly and appropriate fitting. If I try to fit a straight line like this, you can see that particular straight line how best you can think of, we will be having lots of deviations from the actual point.

So, in that case we have to consider or we have to go in for some non-linear model. So, do not think that the linear model is able to solve everything, but for the time being we restrict ourselves to linear modeling at least in this lecture. So, now let us think that what is going to be, because of the straight line fitting, we are certainly going to make some error. There are number of points which are in error, so how do we represent those error.

(Refer Slide Time: 44:45)



Let us say that the errors is represented as E and error for a particular point let us say E p, I write E p will be equal to the target output for the point p. Like say for example, again in this curve, supposing I am interested in finding out the error that is happening at this point. Now, supposing this is point p is equal to 1, the first point I consider. Now, this one is having this as the actual output or what we can call as the target output.

So, this is our t 1 the target output and this is the response that we have got after curve fitting. And after curve fitting means what, that is what the neural network is giving us, because neural network is doing that curve fitting job essentially. So, this is the response or rather in our case it will be y k or the output, so if I say that the response for the point p is y p. Now, let us not use the subscripts, let us use the superscript, because the subscripts we were using for the neural network index.

So, I think it will be creating unnecessary confusion, if I use the subscript there, so instead let me represent it as superscript. So, I say E p as the error pertaining to the point p, so which will be now defined as the target output for the point p minus the actual observation, the actual output which will be the fitted output. The fitted output will be here this one ((Refer Time: 46:50)), so this will be our y p, so this will be the y p which will be the fitted output or the neural network response.

And if I square it up, in that case I will be getting the error term or rather the squared error. And what I have to do is that in order to determine the total error, for all the set of points that I have got, I have to define as summation of this E p and to be summed up over p. Means whatever points are there under the set of p, I have to include all of them, which means to say that it will be summation of this t p minus y p whole square, this will be the summation of errors.

Now, why do we measure this error, because this error measurement is very much is necessary. Because, if we start with any arbitrary intercept and slope for this straight line, we should measure that error. And then, we will find out that in what direction we should adjust the slope and the intercept, so that next time the fitting that I do can be a better fit. So, if I start with some initial assumption about the intercept and the slope, or to talk in terms of the neural network, if I start with some initial assumption of this W 1 and W 2.

Then what I have to do is first measure that how much of error am I doing in the process. And then, dependent upon that error I have to adjust these W's or in term of this I can say I have to reorient the straight lines intercept and the slope. So, that next time the fitting is better. Next time also I do the same measurement that with the better fitment, I calculate the errors, the error overall the points p. And I this time come up with a new error measure, this new error measure may be less or better than the earlier error that we had got.

Accordingly we reorient the straight line, say for example, supposing this is the initial, this is the final straight line that we have fitted. But, supposing with the initial set of points I placed my guess to be like this that this is the initial slope. Now, I slowly I turn this, so that after a few iterations I get the straight line fitted as follows, where the error becomes minimum. Because, what happens is that if I try to orient this straight line more, then the again error will increase.

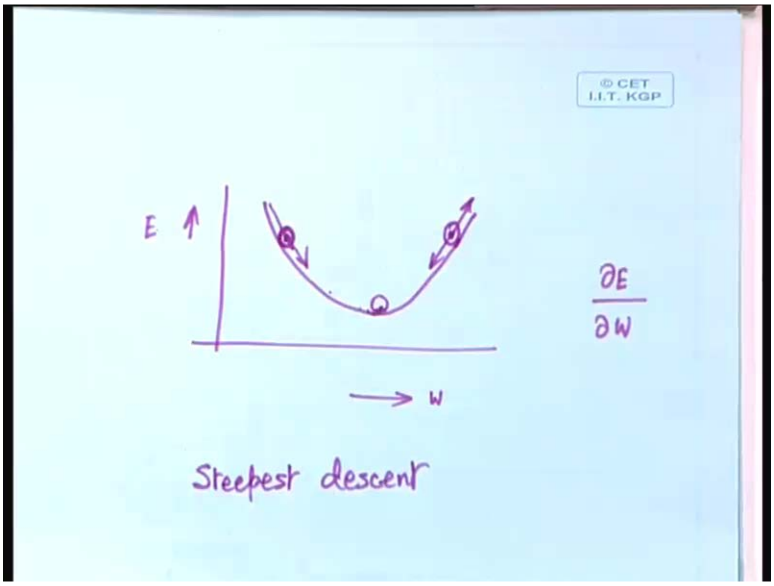
So, we are orienting from here to there and then, we are finding out that the error becomes minimum. So, our objective in a neural network also to be to find out that where are we going to find that minimum error. If we now start making a plot of let us say, in this case, in this simple kind of neural network that we have considered out here take this model.

What we are doing we have got x as the input and we have, y as the output. And we are going to adjust which parameters W 2 0 and W 2 1, in order to determine the minimum error. So, what we can do is that, over all these set of points, we can calculate that how much of error, how much of combined error measure that is E. We obtain for different values of W 2 1 and W 2 0 and supposing we make a 3 D plot, 3 D where let us say on this axis I put say W 2 0, on this axis supposing I plot W 2 1 and on this axis I plot the error.

So, it is a 3 D plot now, can you imagine it is becoming a 3 D plot you vary W 2 0 in its entire range, you vary W 2 1, in it is all possible range. And there is one particular solution, where this straight line is going to have the best fit, and they are the best fit means that where the error is going to be minimum. So, it will be ultimately a surface like this, it is a 3 D surface that can be imagine to be like this. If this is a surface, then you can imagine that this is a 3 D surface, which will result out of all the possible values of W 2 0's and W 2 1.

And there will be one particular point where we are going to get the minimum E. And that is the point which we want to reach is it followed, that is the point where we want to go. Now, imagining at 3 D surface is little difficult, we find it difficulty in visualizing a 3 D surface. So, let us visualize in a better way, let us put a cutting plane and just try to observe the 2 D projection of that on that particular plane. In that case our problem in a 2 D could be modeled as follows.

(Refer Slide Time: 53:24)



Like if we have a curve of E, supposing on this axis we are plotting E. And let us say that instead of two variables here, W 2 0 and W 2 1, we have got only a single variables, some variable which we have plotted in this direction. Now, we are going to reach the minimum point over here, but what could be our starting point to start with in this example. In this case for example, we could start with any value of W 2 0 and W 2 1.

We could be on this part of the surface, we could be on this part of the surface, we could be anywhere on the surface, but ultimately we would like to reach this as our destination. If we could the destination, then I can say that the problem that I am going to solve with this kind of a neural network model is going to converge. If it does not, then really I have to doubt about the usage of this kind of neural network models at all, it has to reach this.

So, what happens is that if this is our starting point on a 2 D error profile, if this is the starting point. Then just see what happens is that, if you are considering this error curve, this error characteristic, error versus some parameters, supposing this is the parameter W that we take. Then, if you are taking the partial derivative of E with W, assuming that a very similar plot you also get for the other W's also. So, E is not dependent on this W alone, so that is why it is a partial derivative that we decided to take.

Now, in this case what happens is that, what is the slope of these dou E, dou W. The slope is like this it is in this direction, but if we are in this position, then on this curve with respect to this curve we should slide down in this direction, we should make a slide down in this direction. So, that by sliding down further, we ultimately come to this point. We may not be able to reach this point, this minimum point in one iteration, we may take several iterations in order to reach it, but ultimately we should reach it.

Something you can imagine like this, that you get a ball where you have just placed some small ball over here. And you have allowed that ball to roll down the surface of that ball, where that ball will ultimately reach, the ball ultimately comes to the point of minimum. If we allow the ball to slide down freely, if we do not allow it to slide down freely, if we try to forcibly make it things may be good, things may be bad. Like if the surface is having lot of fraction, let us say in that case the sliding process will be very slow.

It will very slowly slide down, so we may feel in the inpatient about it and we may just put some taps, so that the ball slides down faster. But, if we are too much ambitious about it is fall, that now the fall should be much faster the that, then we are going to shake it so vigorously that it may ultimately leave the surface completely and the just the ball will fly away, it can happen. So, we have to be careful about that, but the objective is that if you allow it to roll down freely.

Then it roll down in a gradient descent way, if the ball lies here it comes down this, if the ball lies in the opposite slope, that means to say here the slope is this is in this direction, but the ball has to roll down like this. So, it is always against the slope that it is sliding down, so this technique is known as the steepest descent approach. So, this will be something that we will discuss further in the next class.