**Data Analisys and Decision Making - II**
**Prof. Raghu Nandan Sengupta**
**Department of Industrial & Management Engineering**
**Indian Institute of Technology, Kanpur**

**Lecture – 58**
**ANN**

Good morning, good afternoon and good evening to all of you my dear friends. Welcome back to this course and my dear students also. Welcome back to this course DADM II, which is Data Analysis and Decision Making II under the NPTEL MOOC series. And as you know this total course is basically this course DADM II is for 12 weeks which is 60 lectures, which convert into hours becomes 30, because each lecture is for half an hour.

And this we are in and we have already completed 11 weeks, which means 11 into 5, 5 being the number of each week number of lectures that 55 class has already been completed. And you have taken assignments 11 assignments had already been taken by you and solved. And we are in the 12th week. And as you can see from the slide, we are in the 58th lecture; that means, 58, 59, 60 will complete this course; we have completed 56 and 57.

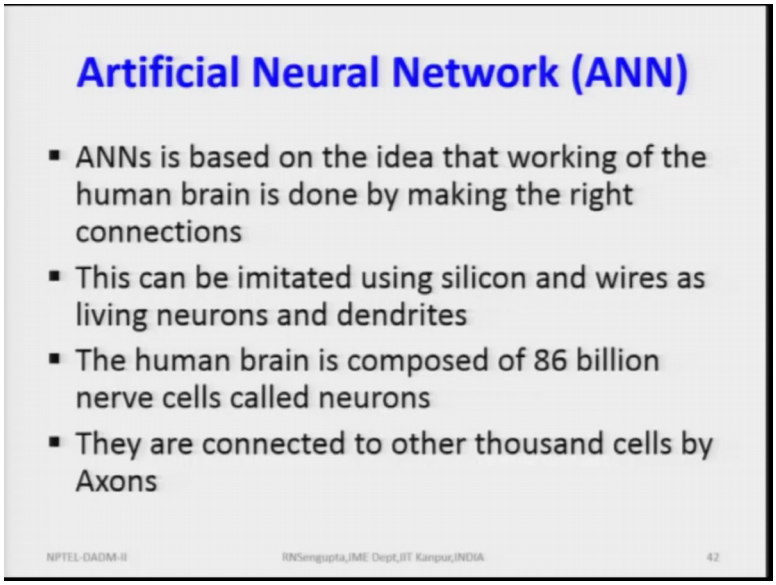
So, in 56, 57 we basically completed the concepts of AIS. And in the last lecture, the total duration who you will find it a little bit less in number like not 30 minutes. So, in on an average, the classes are 30, 31, 29 minutes plus minus same. So, that last lecture was 21, because I finished on the 21st minute in the concepts of AIS. So, I did not want to start immediately the new concept about the one which we were going to discuss.

And my good name is Raghu Nandan Sengupta from IME department at IIT Kanpur. So, we will discuss the concept of artificial neural networks and we will utilize artificial neural networks in a problem solving case again from the area of finance for prediction, some prediction not prediction actually prediction of from the concept of change point detection.

So, I will come to the concept of change point what we mean by that. As I speak I will draw some few diagrams also as I did for that problem where I discussed that how the prediction alpha, beta errors can be minimized and what the errors are, I do the concept in the hypothesis testing case to. So, this background again for artificial neural network

would we kept simple. So, these are as a concept I will basically discuss and they can be learned in details by the people who are attending this course.

(Refer Slide Time: 02:47)



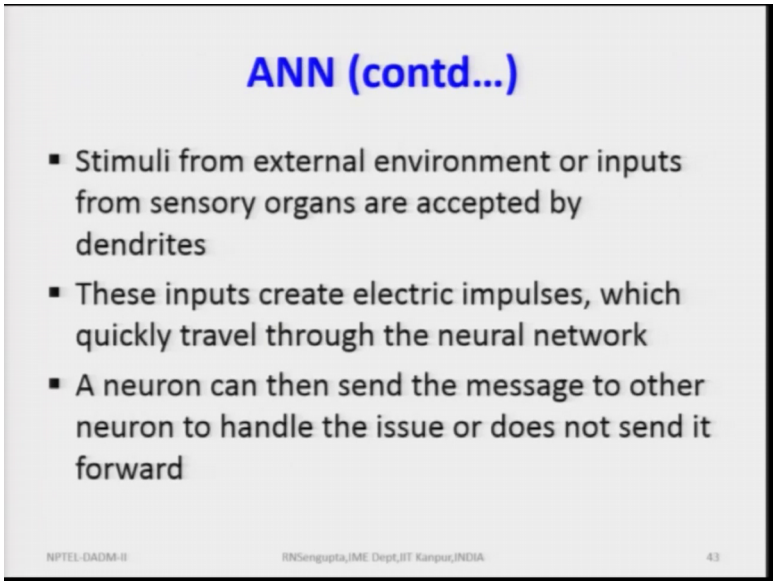

So, artificial neural network ANN is based on the idea that working of the human brain is done by making the right connections. So, there are a nervous system, neurons are there, synapse are there, the joining portions and electrical connection is made and the information is passed in is electrical signals. So, suddenly if you get a pain here or some you touch a very hot thing on that and your body ache is there, so that nervous system basically transmits that information very fast.

So, if there is dead cells no neurons and then obviously, you do not feel the pain. So, obviously, I mean when the barber cuts your hair or you clip your nails and you do not feel the pain because the concepts of feeling in ANN is not there. So, if you consider the sole of your feet, the obviously, there are few feeling is there, you feel it, but the level of sensitivity is much less.

Now, this human brain is done as it works if this can be this is the nervous system what I am talking about the passing of the information through the neurons how the take place that will be replicated as ANN as we did for the artificial neural system. The second point is this can be imitated using this silicon and the wires as living neurons and dendrites.

So, they would be converted into the nervous system into neurons, artificial neuron systems where information can be passed and processed. The human brain is composed of 86 billion nerve cells called neurons, and we will try to replicate the neuron in a very simplistic sense. So, they are connected to other thousand cells by the axons and information is passed through the pulses or electrical connections.

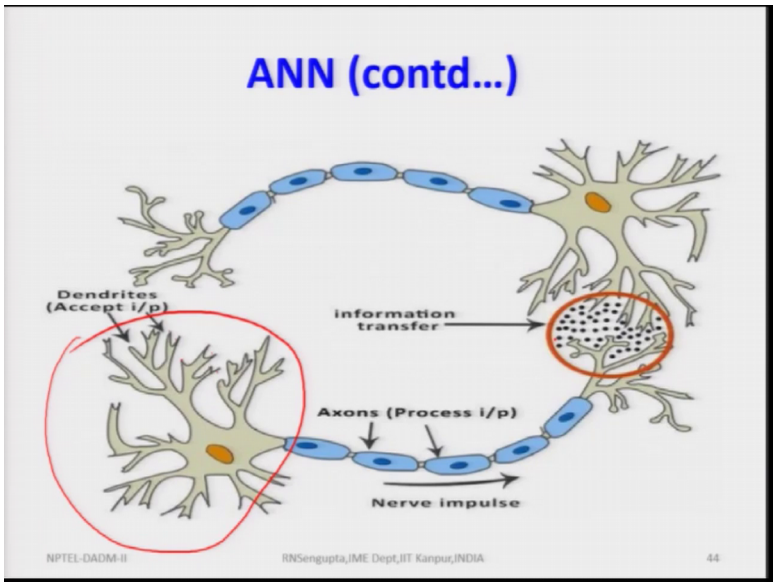
(Refer Slide Time: 04:47)




Stimuli from the external environment or inputs from sensory organs are accepted by the dendrites and they are processed. Suddenly we feel the pain we will immediately remover our hand or in coldness, hot, pain, so all these things are transmitted. These inputs create electrical impulses as I said which quickly travel through the neural network and basically goes to the brain and the brain basically passes the information what we should do, should we remove the hand or should we do not remove the hand.

So, if the response is very solve, obviously, it means the nervous system is not working properly. A neuron can then send the message to other neurons to handle the issue or does not send it in forward, that means, it will basically take the decision as I said. If I pinch myself and suddenly I feel the pain, I somebody pinches myself I will basically remove the hand. So, because the information is coming immediate that yes it is hurting remove it.

(Refer Slide Time: 05:47)



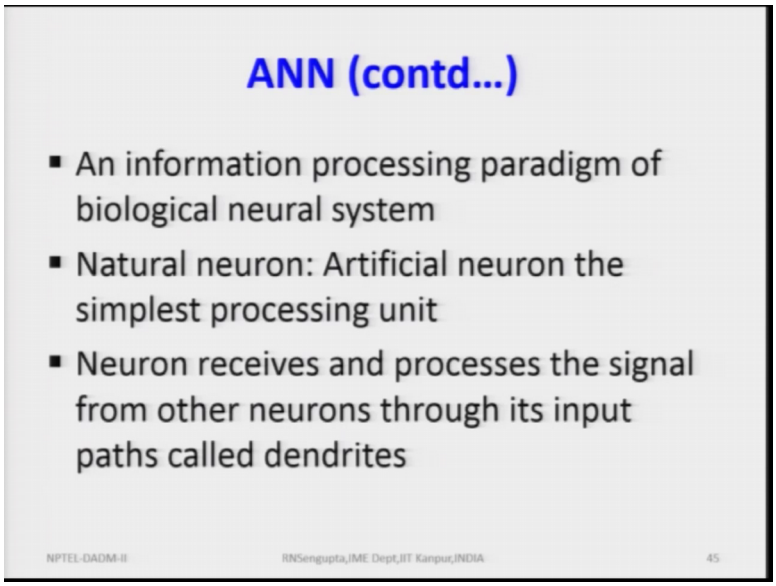

So, this is how very simply and if somebody who wants to draw this from on the biological perspective on the bio science students, I may not have given the right diagram in detail, but as please bear with me I have trying to give the ideas as simply as possible. So, if you consider the dendrites are there, so they are like small this I would not say that if you see if we zoom in to this plant, there are many leaves, small leave, branches. So, this is the area.

So, the dendrites accepts the information there are many stimulus points. So, these are the stimulus points, we gets the information, shocks. And then the nerve impulses through the axons, so which processes the inputs. So, these are the cells which passes on the information.

And these electrical informations are, so they do not connect basically this electrical information passed on from cell to cell through the whole system from dendrites axons, and it goes on to synapse and into the brain. And based on that, the brain takes the decision what to do. In case if it does not want to see them then the transformation information we stopped there only.

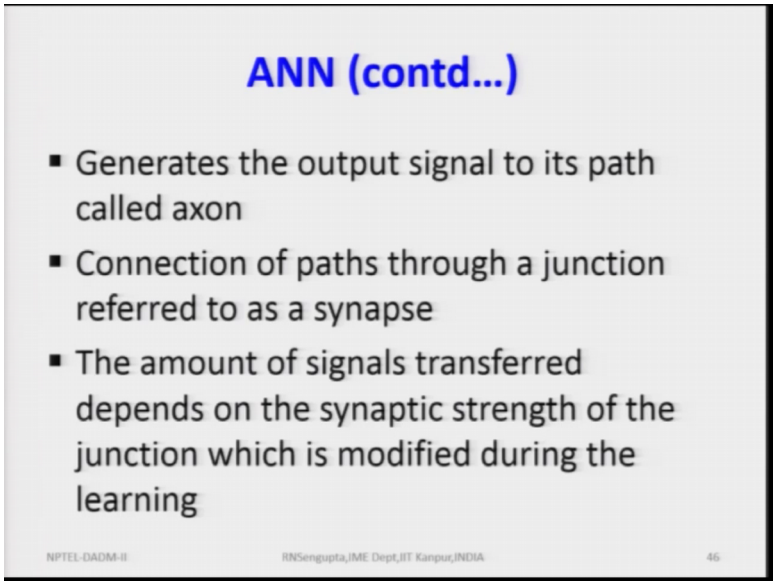
So, no need of passing on the information; in a very simple sense, dendrites, neurons, axons and then it goes through neurons and to the brain. So, this synapse is the whole portion, dendrites are the point at once where the information comes. And these are the information transfer is happening through the electrical pulses.

(Refer Slide Time: 07:37)




An information processing parent paradigm of biological neural system is basically ANN. The natural neuron is basically the artificial neuron has been will be made which is the simple processing in unit based on which the information is passed. Neuron receives and processes, as I already said neuron disease and process the signal from other neurons side by side. So, they are going in series I get the I am in neuron, I get the information from my previous source, process it and pass it on. Neuron receives and process the signals from the other neurons through its input paths called the dendrites and it goes forward it.

(Refer Slide Time: 08:17)



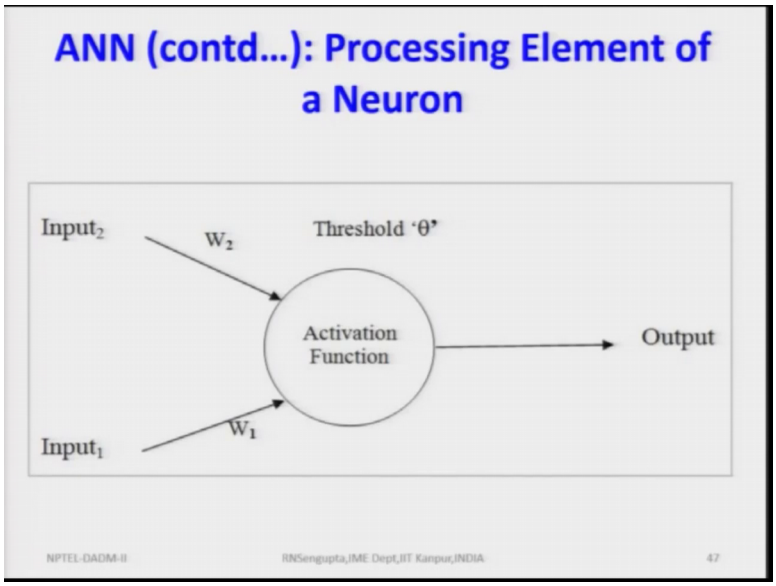
In a in general on neural network system, it generates the output signals to its path called the axons and the impulses are given. Connection of path through a junction referred to as synapse. So, these are where the connection happens the dotted points where there signals are going. The amount now which I had forgot the first last point is basically that the amount of signal, signal, the intensity of the signal would give you the importance of the information processing which is happening. The amount of signal transfer depends on the synaptic strength of junction which is modified during the learning process.

So, as in the artificial immune system, the cloning which was happening or the binding of the white blood corpuscles the fighting cells with the pathogens which are coming, the bad cells which are coming they bind. So, how they are replicated the clonal selection both under the positive and the negative selection process, so they generate faster cloning is faster, matching is faster affinity is faster, so they generate faster and more in numbers.

So, this synaptic strength would also be depending on how the information flow is there. High impulse, low impulse will give you the intensity and the importance of the information which should be passed on from the dendrites to the neurons through the synaptic phase space and then into the brain for processing.

So, this is already mentioned. So, these are natural neurons; neuron receives the process and signals them and pass it to the dendrite. So these generates the output signals to its path in the axons. Connection of the pass through a junction referred to a synapse. The amount of signal transfer depends on the synaptic strength.

(Refer Slide Time: 10:23)




Now, I come to the, so this is a very simple background that how it should be done in the actual I am not going to the biological details. Now, the processing element on the neuron which will try to replicate in artificial neural network system is like this. You have an act, so technically we will consider this as the processing unit which will process. So, signal is coming.

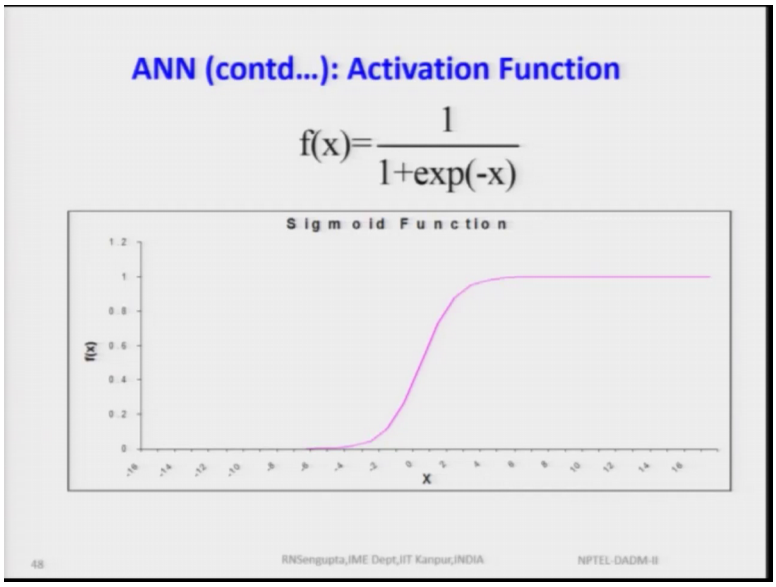
So, input is coming with some w 1 2 w 2 are so called the weights based on which can also be probabilistic they will change. So, this input one, input two may can be many inputs. So, you will basically have stages, first stage, second stage, third stage and each stage would have such activation function on the neurons which is being pictorially given as a circle which is activation function.

So, this activation function means inputs are coming inputs can be say for example, the price of a product, quantitative product, flow process I have happening of fluid in a channel or some functional form. And this activation function basically processes it depending on activation function. It will basically activate the whole process it, and pass on to into the next cell they would be our replica of this neuron or the circle of the joint which would be there on the next stage, so is that the input which is processed by activation function. So, the active function functions can be different depending on the parameter values. I will come to one simple example activation function.

So, these activation function basically processes the inputs and gives them as the output again a sort of singular into the next stage. So, say for example, you want to predict. So, there would be first layer, second layer, third layer, they can be hidden layers also. So, the number of layers is not limited, but obviously, if there is a rule based on which in the number of layers are kept minimum, because too much of processing would definitely not help in my prediction or whatever work I am trying to do.

So, they would be processed, and they would be given as output based on which we will analyze that there are many subjective and objective sets of informations inputs which are happening and depending on the activation functions which are used at each node will get the output such that output would clarify that based on the decision variables the parameters what should my output.

(Refer Slide Time: 13:03)

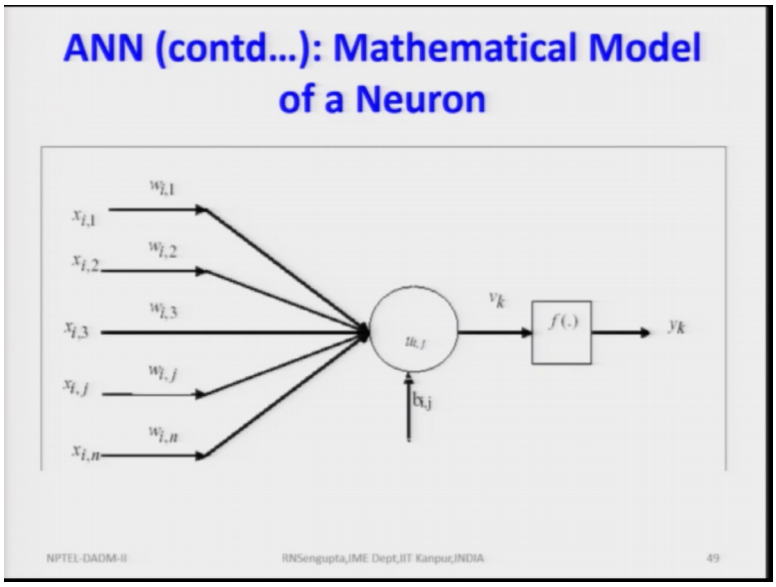



So, basically the activation function is a sigmoid function. So, f of x is 1 plus 1 in the denominator you have 1 plus exponential to the minus x. Now, keep increasing. So, the x factor which we say is the amount of information is coming, higher and lower; obviously, you to mean e to the power minus x would be if it x is very high e to the power minus x would be a very big number.

So, the ratio would almost be 1. And in the case if e e to the power and x is very very small number, so obviously, the value would change accordingly. So, this sigmoid function and the activation function will detect what type of processing should be

utilized. So, you are you will get different functions the inputs which are happening with the weights will basically multiply them accordingly to the get the output. I am just giving one example of a activation function only.

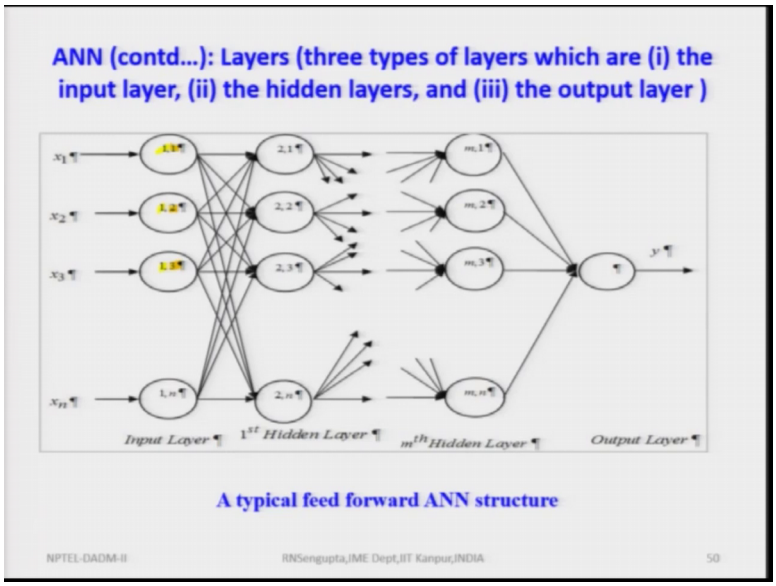
(Refer Slide Time: 13:59)




Now, this is what we basically try to denote the mathematical model of a neuron. So, there are inputs x 1, x i 1. So, there are different levels. So, the second number 1, 2, 3, 4 till n are for the ith unit ith unit of the node. And the weights are given by w again I am ignoring i am just giving 1, 2, 3, 4 till n. So, these weights and the inputs are basically coming to the node u i j and they can be. So, b i j means basically these are the external influences like white noises you want to predict the prices of something and they are white noises which are not under your control.

So, those would be considered as the b i j for that particular stage or the node. Then based on that you are getting your output which goes on further on through the functional form and based on that you get the output y k; so they can be different such neurons each stage, so that I will basically try to discuss now.

(Refer Slide Time: 15:09)




So, this is a very simplistic way of defining. So, you have the input layer and then and the nodes unmarked like this. The first number, so if you see this so called node, so this first number is for the first layer input layer. Second number which is there it is for the node number in that layer. So, 1 1 means the first layers first node; 1 2 means first layer second node; 1 3 means first layer third node. So, you can basically have n such nodes for the first layer. In the second one, the nodes and marked as 2 1, 1 1, I am not I am mentioning the first number and the second number 2 1, 2, 2 3 till 2 n.

So, number 2 with the first one is basically the first hidden layer of the after the input layer. And the second number is 1, 2, 3 till n are basically the node number. Similarly, I can go to the third layer which is the second hidden layer and the numbers are 3 1, 3 2, till 3 n. Similarly, the fourth layer or the third hidden layer 4 1 4 2, this third and fourth are not drawn here. The hidden layer for third one which is the fourth there would be 4 1, 4 2, 4 3 till 4 n. Similarly, the sixth seventh and I come to the mth hidden layer the numbers are marked as m 1, m 2, m 3 to m n, and then I basically have the combined effect and then I get the output layer.

Now, in the input layer, there are inputs happening. So, I will consider the inputs as x 1 is the input which is happening for node 1 1, x 2 as for 1 2, x 3 as 1 3. So, this it can be also made that there are multiple inputs for each nodes, but the reason it is kept as different nodes is that the sigmoid function which I am trying to utilize for each node the values

and the parameters of that function would be different depending on the input which you have.

So, if you have an input x 1, the sigma function would be depend consider it will be f of x 1. The basic structure of the sigmoid function of the processing function remains the same, the parameters may change. Similarly, for x 2 again the sigmoid function is same, but the parameters are different, because that is why the nodes which you have 1 1, 1 2, 1 3, till 1 n are specific to the sigmoid and or the consider the input function which you are using.

Now, as the output comes from the node one 1 1 or 1 2 or 1 3, they these would be going into different amounts or different effects would be therefore the first hidden layer that the second layer. So, the output happening from 1 1 can have consider a percentage wise or the probability of its effect would be of different values. So, it they can be as I will consider as if I say the probability; probability is p and the base numbers I will mean I can mention as 1 1, one the first one is basically it is coming from the 1 1 node from the first layer and the second one is basically is going to the first node in the second layer.

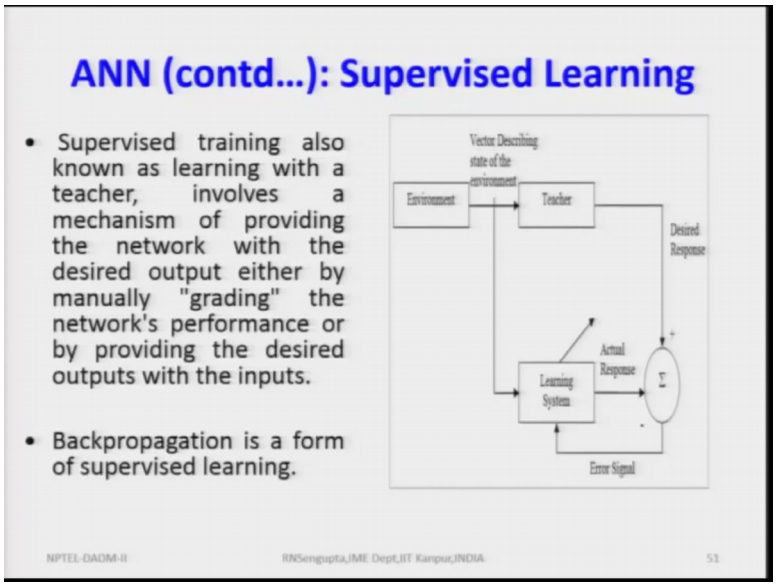
Similarly, if it is p i j it means basically it is coming from the ith hidden layer to the jth hidden layer which is next side by side, and from which node it can also be specified. So, if you see the diagram here, the connectivity is there for each and every node. And interconnection is there between each node in the input layer with each node in the first hidden layer, similarly for each first hidden layer connection would be there for each node in the second hidden layer and so on and so forth. The second, third, fourth are all omitted due to positive space, I draw the mth hidden layer. So, the mth hidden layer all the nodes 1 to n are being connected individually with all the nodes which are there in the mth minus 1 hidden layer.

The other point which should be mentioned is that the number of nodes which I have for each layer I am taking as 1 and as n. So, it can change also. So, depending on number of such sigmoid function which are applicable it will change. So, these each input layer and nodes of the input layer, each in a node of the first hidden layer to the mth hidden layer would have the sigmoid function depending on the functional value which we have.

So, we will basically give some input find out the weights process it time and time and as add as it changes the probabilities and the processing probabilities will change, the

weights will chain and we get the output. So, what is supervised. So, this is the structure. So, we will have basically have some supervised learning. If you remember the cloning process, they basically clone themselves depending on the probability of affinity.

(Refer Slide Time: 20:17)




So, we will have some supervised learning concept also. So, supervised training also known as the learning with the teacher involves a mechanism of providing the network with the desired output either by manually grading the network performance on by providing the desired output with the input cell that it trainings itself and learns it better. Map book there is a way that back propagation. So, if you consider the environment, so it is giving some information and it goes to the teacher that teacher means the training set which is there. So, they give a output desired output.

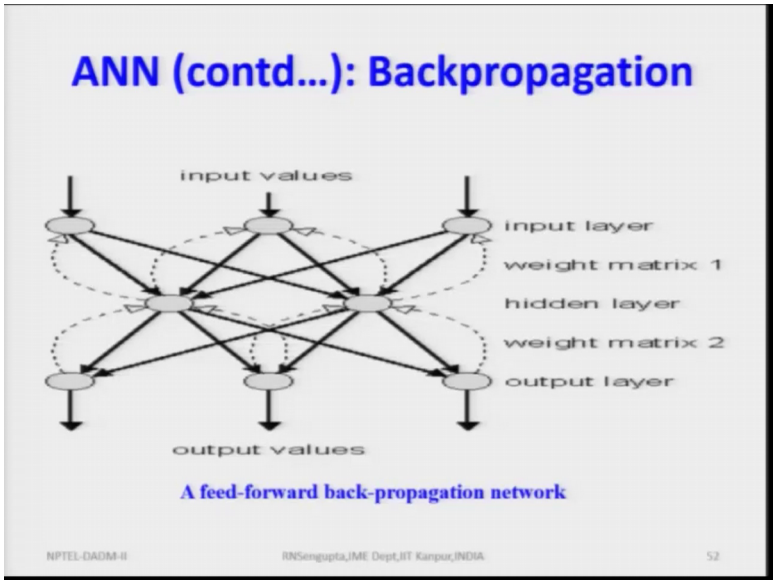
So, the desired output is basically processed and based on the functional form. And once the function form and design output is there, the error signal if the error signal is very high or low the error signal goes. So, if the error signal comes here and the error is very high, then the actual the learning system will modify it and again given the actual response to the processing unit. So, it will continue in this loop till it is the best and if the error signal is within limits, so that the learning system has basically learned itself to that state, so that you can pass on that information for the next level of processing.

Now, if you consider these loops which are happening here; loops which are happening here, so it is something due to the Q GERT and GERT process the graphical evaluation if

you technique and the queuing graphical evaluation rule technique. So, these rework which is going on from one node to other, where the nodes are the jobs. So, this concept is similar to the what we are considering at the supervised learning.

And to just to remind the GERT and GERT processes are looping which was there was not there in the POT and CPM method which we have done the POT and CPM method we have done in the project management course under the NPTEL series which I taught, but if you have taken that you will would have understood. So, we did not discuss in details about GERT and Q GERT in that class we did, but not in that details.

(Refer Slide Time: 22:31)




Now, we come into the back propagate; back propagation means information is happening and an and flow is going back to fine tune in the prediction model. A feed forward and back propagation mode method is that input variables are coming. So, this goes vertically down. So, inputs are coming. These are the nodes at this level. So, I am the basic it can be the mth layer hidden layer mth middle hidden layer.

So, those are immaterial. So, input layer is there and the weight matrix is given to what weight ages does the connection would be. So, you would have one weight here. I am just marking and not writing. So, the weights would be w 1, w 1 1, w 1 2, then w 2 1, w 2 2, w 3 1 and w 3 2 that means depending on from which node to which node.

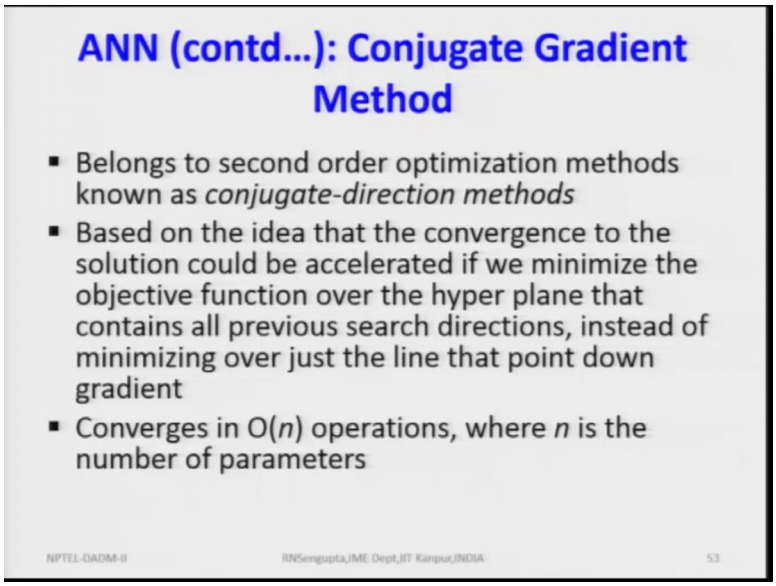
So, this is the hidden layer and there is the weight matrix also there for the output layer again the weight should be defined accordingly. So, be careful about the suffix of the weights. Now, what happens is that as the information goes to the hidden layer, so obviously, hidden layer will also give some information set which is the feedback loop which is coming into the initial input point. So, this in this feedback loop are shown as dotted.

So, from the second node, I am considering the numbering of the nodes being starting from the left to the right. For the second node for the hidden layer, so this will give a feedback to the third node for the first layer. It will give a feedback to the second node for the first layer, similarly can give a feedback to the this is not drawn, so it can give a feedback not running this is the dotted line is not there it can give a feedback to the first node in the first layer.

Similarly, the output layers which are there, it gives us it will give the output the back propagation concepts of the feedback would be there; from the third node, in the output layer to the second node in the hidden layer. Similarly, it can give feedback to the first node in the hidden layer. Similarly, the second node for the output layer and the first node of the output layer can give information back information to the second node in the hidden layer, similarly both of them both of them means the first and the second node for the output layer can both of them can give feedbacks to the first node for the hidden layer.

So, again I am repeating the number one nodes I am using from the left to the right and the input layer is the top. So, it could have been drawn in 90 degrees or so. My input layer would come like this, first hidden layer, second hidden layer and the till the mth hidden layer as I have discussed. So, this is a feed forward back propagation network based on which you can do.

(Refer Slide Time: 25:33)




The conjugate gradient method it would be basically belongs to the second order optimization method known as the conjugate direction method. So, based on the rate of change of the function, we will consider on that.

So, if you remember whenever you are trying to do optimization problem in when it is a maximum or the minimum, what we actually consider is the rate of change of the functional value with the change of the decision variables. And in which case the decision variables rate change is the maximum in the positive direction or in the negative direction will precede that in that direction depending on whether the problem is a maxima or the minima.

So, this would not be discussed here, it will basically be going into the DADM III course. Based on the idea that the convergence to the solution could be accelerated if you minimize the objective function over the hyper, so the minimization and the rate of change of the function in the hyper plane. Why it is an plane, because if he is a two dimension one that finding out dy dx is very simple.

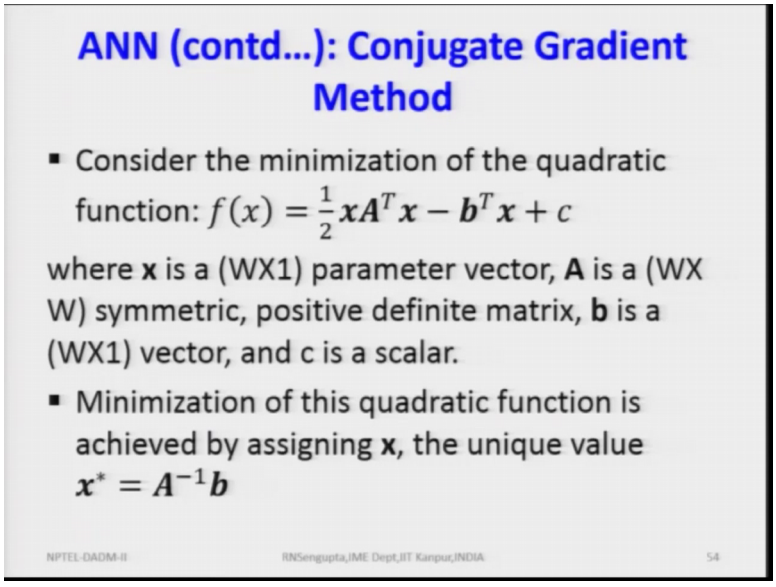
Now, if it is that, if you want to find out the rate of change of the function in that three dimension one, you have to basically take slices from any direction and find out the partial rivet different differentiation of the functional form with the decision variables x 1, x 2, x 3 whatever they are which means that we are keeping if you are finding on the

partial derivative or the functional form with respect to x 1 it will mean that we are keeping x 2 and x 3 and other variables fixed.

So, we are only trying to find out the rate of change of the function in one direction only. So, let me continue reading it. If you minimize the objective function over the hyper plane that contains all previous search direction instead of minimizing over just the line that point that downward direction. So, we will basically combine the partial derivatives to find the rate of change. The convergence is of order O order n where n is basically in the number of parameters based on which we are trying to say.

So, depending on the half time complexity, space complexity, all these problems can be considered from the computer science point of view, but I am not going to go into that. I will just simply consider the concept of artificial neural network, how it can be utilized. So, we will consider the minimization of say for example, the quadratic function it is not required, but I will just briefly mention it.

(Refer Slide Time: 27:51)




So, when we are you want to find out here A is basically the W cross W symmetric matrix W the size of the weights which you are going to consider x is the W cross 1 parametric weight vector, and b is the W cross 1, again the vector of parameters and c is a scalar. So, what we want to find out is that value of x star based on which we can minimize or maximize the functional form such that we achieve the objective function as maximum or minimum.

Now, what we want to do is that depending on the objective function and rate of change of the function, we will basically have a sigmoid function or the input function which will process the input, try to find out the optimum value whatever the optimum value and then combine and basically give of a feedback loop depending on how close or far the functional form is there with respect to the achieved target.

So, if the error is less than some threshold value, the feed forward would now then stop, because then the information flow back would not happen because it is already cross that threshold and it can go into the next hidden layer. So, the input from this layer would be passed on the next hidden layer.

And based on that this process will happen and they would learn themselves. As I said the clonal concept which was in AIS or in this supervised learning concept can be tailor made depending on the functional form which we use. With this I will close the 38th the 58th lecture and continue discussing about the ANN and its application in a simple area of finance about the change point direction about which I will definitely explain in somewhat details as I have been doing for AIS and an artificial neutral network system.

Thank you very much for your attention and have a nice day.