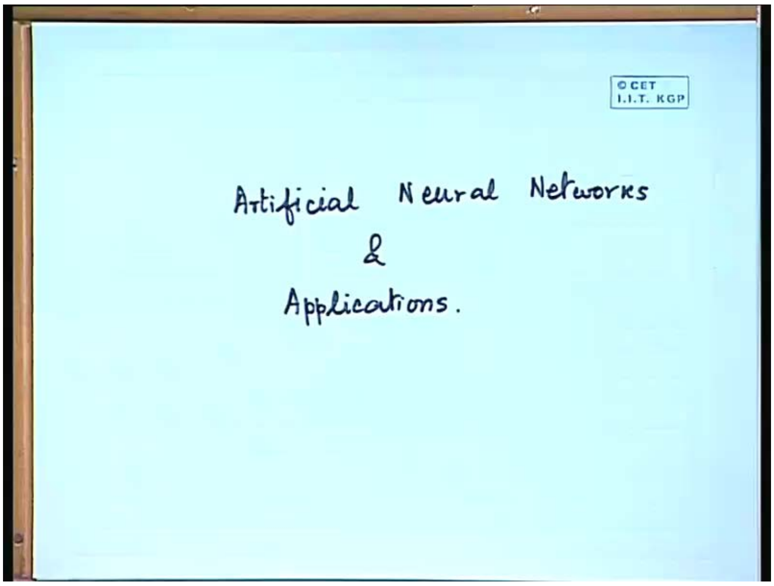
Neural Networks and Applications

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

Lecture - 01
Introduction to Artificial Neural Networks

Welcome to all of you in this 40 lecture course, on the Artificial Neural Networks and Applications.

(Refer Slide Time: 01:08)

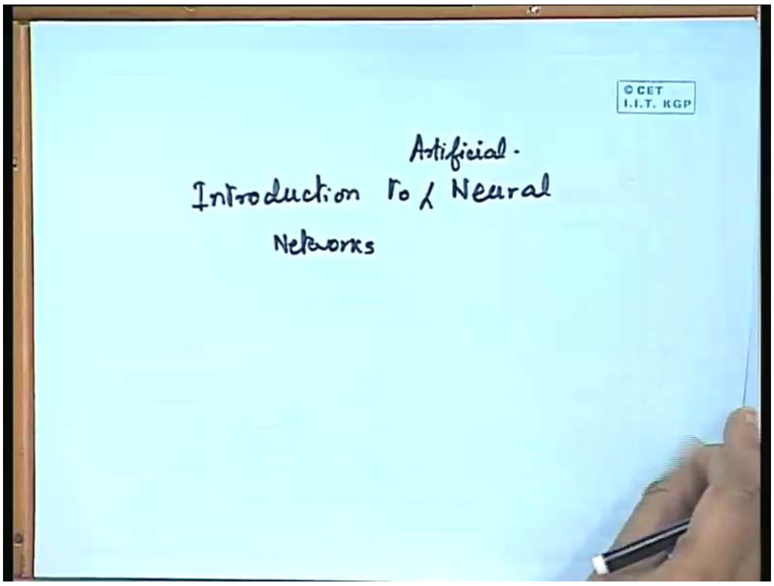


The subject of artificial neural networks has matured to a great extent over the past few years. And especially with the advent of very high performance computing, the subject has assumed a tremendous significance and has got very big application potential in very recent years. Now, this subject of artificial neural networks is going to be covered in approximately 40 lectures. And in that today I am going to begin with the introduction to the artificial neural networks.

So, we will be first defining what a neural network basically means. And as a name implies, actually the term neural networks derives it is origin from the human brain, or the human nervous system, which consist of a massively large parallel interconnection of a large number of neurons. And that achieves different tasks, different perceptual tasks, recognition tasks etcetera, in an amazingly small amount of time. Even as compare to today's very high performance computers.

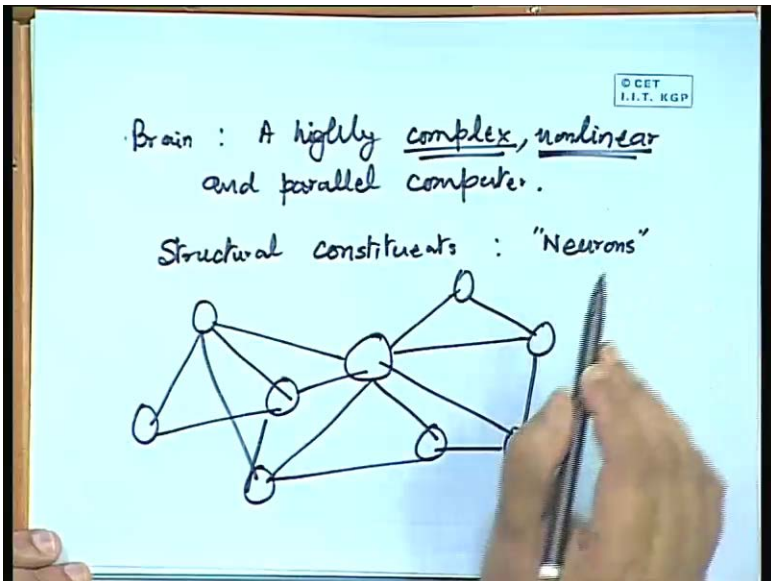
So, this is what inspired the researches to think that is there anyway, whereby a computer can be made to mimic the large amount of interconnections and the networking. That exists between all the nerves cells, can it be utilized to do some complex processing tasks where today's high performance computers also cannot do. So, this subject is the one that we are going to address. So, today we are going to study the introduction to neural networks.

(Refer Slide Time: 03:47)



That is going to be the topic of today, and specifically we are going to address the introduction to artificial neural networks. Now, whenever I say artificial the question that immediately comes to our mind is that, what is then the natural neural network.

(Refer Slide Time: 04:21)



Now, we know that our brain, the human brain is having a highly complex and non-linear parallel competitor. A human brain has a highly complex non-linear and parallel computer; and this can organize it is constituent structural elements. So, it is structural constituents of human brain, they are known as neurons. And the neurons are interconnected not in a very simple way, but in a rather complex way, so complex that many of the things we do not know yet.

Say, if we have got a large number of such neurons or nerve cells, which carry out the processing. They will be interconnected typically in a highly complex manner between each other. And there will be connections which exist from one neuron to the other, and that is how a network is realized. And this network is as told it is highly complex, as well as non-linear and is massively parallel. Because, our human brain has got, typically billions of nerve cells with trillions of such interconnections existing.

Now, let us I mean understand something which we can gives as an example, let say familiar task of recognition. Whenever we meet a person, a person who is known to us, we can recognize that what is the name of that person, because he must be friend of ours, he must be some known person. And we will be able to tell that who he is whenever be meet him and how is that we are going to do it, we are going to perform a task of recognition.

Now, we may be knowing 1000's of people around. And I mean what you have to do is to immediately instantly recognize that person. That I mean he his somebody whom I had met 5 years ago and his name is such and such, he works in such and such place or I had met him in connection with these works. We can immediately have this recognition task, this face recognition. Now, supposing instead of myself doing it I ask a computer to do this task.

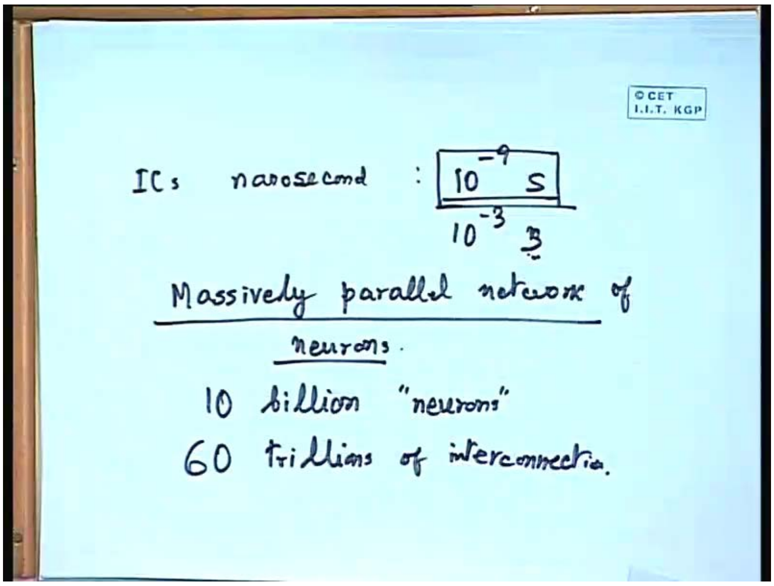
How much of time would you think that a conventional computer is going to take, it is not going to be of any easy computation for a computer. Because, firstly that you know that person, but a computer does not know that person. Firstly, that you have to teach to the computer that, there are photographs of different people. Supposing if I know, I mean thousand persons, then I should be feeding the photographs of 1000 people into the computer.

Now, I meet a person, so I now capture his photograph and then, I feed to the computer that, now you try to match the present photograph with all the 1000 photographs that you are going to have in the data base. Now, the computer is meticulously do that task, I mean you must have trained him, that how to perform a recognition, how to match the given image with that of the stored photographs. And then, ultimately at the end of all the computations and at the end of all the comparisons with 1000 people's photographs which is there in the computer data base.

The computer is going to give us results saying that, the person best resembles this photograph. Now, this could take I mean what long time may be a few hours, or who knows I mean may be several hours that would all depend upon, how many such images are we having in our database. If our database is very large, if we want to track down a person from a collection of let say, 10000 image, 100000 images, then the task is going to be really complex.

Because then, so many comparisons are involved, but how much of time are we taking we are doing it almost instantly. Now, how are we going to do that very instantly, is it that the computational capability that exists in the humans, is it enormously different from the way a computer is doing. Well, if we try to think in terms of the processing speed, we will be getting a different type of a result, like today's silicon IC's we know, it is response time is expressed in terms of nano seconds.

(Refer Slide Time: 10:36)



A nano second is time which is 10 to the power minus 9 of seconds, whereas if we are looking at the processing speed of a human neuron. That may be 5 to 6 order slower than that of a typical ICs, and that may take several milliseconds and milliseconds as you know is 10 to the power minus 3 seconds. So, it is 6 order slower, 5 to 6 order of slower, but in that case the question remains as very puzzling one.

That then, how is it that the neural processing within the human brain happens to be much faster than that of today's computer. Because, we are using the elementary block of the digital computation is all those integrated circuit, which have got processing times of the order of nanosecond. Whereas, the elementary block of the human computation the neurons, that is 10 to the power minus 3 seconds.

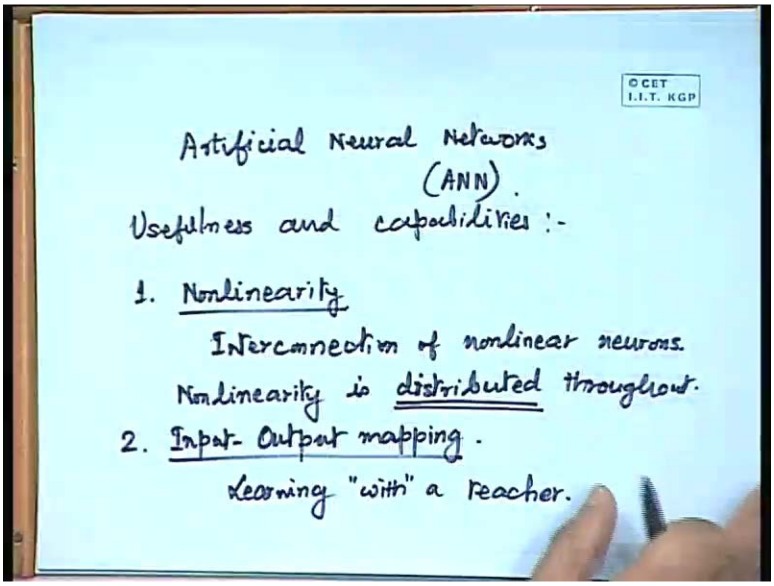
So, then how is it that this one happens to be much faster than this, the answer lies in the fact that the network of human neurons in the brain, that is massively parallel, there is a massively parallel network of neurons. And as I told you that the number is of the order of something like 10 billion, typically there will be 10 billion of such nerve cells or neurons. And this will consist of approximately 60 trillions of interconnections, so the answer lies in this massively parallel structure.

So, now the question is that is it possible for anybody to perform the tasks, that a human brain does is it possible to mimic that using the electronic components. Or is it possible to realize that task using a computer software. Well it is not that easy, because we do not have even in the age of parallel computers, we cannot really think of putting so many processing units and realizing it in a massively parallel scheme.

All that we can do within our limitation is that, we can interconnect a network of processes no doubt, and rather than considering the structure of a human brain in totality. We can only try to mimic a very small part of it, an extremely small part of it, in order to do some very specific task. That is the best thing that, we can do using electronic components and using the software we can do only to a limited extent.

We can make neurons, but that is surely going to be different from the biological neurons that we have talked about, so far. So, what we are going to study is the artificial neural networks.

(Refer Slide Time: 14:46)



By artificial we inherently mean that something which is different from that of the natural or the biological neurons. So, this is the subject artificial neural networks, in short form we very often refer to it as ANN. Now, firstly let us understand that, why at all we are going in for an artificial neural networks? What are the advantages that it is going to offer to us? So, let us list out the usefulness and some capabilities.

Number 1 that it exploits the nonlinearity, now I think all of you should be able to understand the terms linearity and nonlinearity. Basically, if there is a system where we give a set of inputs and we expect some output out of it. In that case, we call the system to be linear, if the relation between the output and the input can be best described in terms of a simple linear equation. If there are let say 4 inputs and 1 output, then if the output is a linear combination of all the 4 inputs, then naturally the system is linear.

Whereas, if we can write the output only in terms of, not only the linear terms, but also it is higher order terms, in that case the system is no longer linear, the system becomes non-linear. Now, lot of times for simplicity we consider linear computational models. But, if we are looking at the real life problems, most of the real life problems they happen to be highly non-linear in nature.

So, for that purpose we need non-linear computational units as well, and the neurons other once that happen to be non-linear. So, what we have got here is an interconnection of non-linear neurons. So, in artificial neural networks, we have got an interconnection of non-linear neurons. And another thing which should be noted is that, the nonlinearity is distributed throughout, in fact the very nature of computation as you can realize is highly distributed in nature. So, the nonlinearity that we are talking of is naturally distributed.

The second usefulness that one must talk about is the input output mapping. You see that you are providing some input to the system, and in response you are going to get some output. Now, I mean we can go in for a learning mechanism, a learning where a teacher is involved in which case what we do is that, we feed the inputs and then, we also say what the expected output is going to be. So, in other words we are specifying that for a given input, what is going to be the output or the desired response.

Now, it is possible that our computational unit that we having, that is not able to achieve the actual output that we get, may be different from that of the desired output. There may be difference between, what is actual and what is desired. What we can do is that, we can accordingly modify, if our system has got a set of free parameters, some parameters that we can adjust. So, if we are having such kind of free parameters, then we should be able to adjust the free parameters of the systems such that, for a given input or a set of inputs, we can obtain the output that is closest to our desired output.

In that case we may not be able to achieve that immediately. First time we feed a pattern, our system does not, I mean has not encountered that pattern before. We are feeding that what the desired response is going to be, but then the actual output will be different. So, the difference that exists between the actual output and the desired output, that adjust the parameter of the system.

Such that, the difference between the actual and the desired is minimized; and that we may have to do several times. So, there is a process of learning and this learning as you can understand involves a teacher, there is a teacher who says that corresponding to this input this is what the output should be. And if it is not, then the teacher is asking you to correct, the teacher is asking you to adjust the free parameters which are available in the system.

So that, next time you feed the input, the same input you can get an output which should be closer to that of the desired. So, this is something which is very very important and this is what is going to make the neural network, remarkably different from the conventional computational unit, and that is in the sense that it has got a learning ability. And in this case the input output mapping that we were referring to so long, is basically referring to learning with a teacher, here a teacher is definitely monitoring.

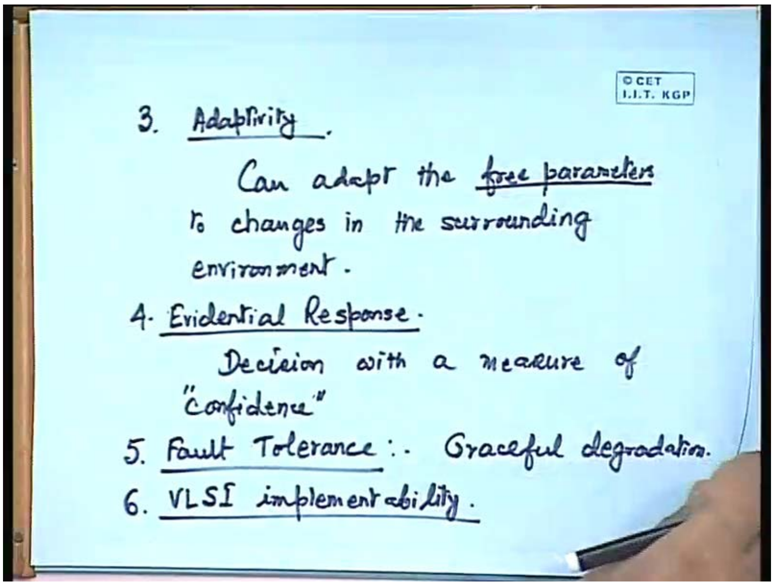
Now, note that all the time we can find a teacher, there may be some situations, where we may have to learn without a teacher also, may be from simple associations. You see let us look at the developmental process of a child. Now, a child is born with a brain and that brain has got massive interconnection of is neural processing units, but a child has to develop himself or herself with a process of learning. A child sees so many new things, when the world is new to a child, the child learns, the child finds out many things by himself or herself.

And that the child is able to do through some process of association, let us say that I mean a child sees so many animals around. Now, a child sees that group of 4 legged animals is called a cat, a group of 4 legged animals is called a dog. Now, a child may be making mistake initially, sometimes a child could get confused between what a dog is and what a cat is. So, he may feel little confused, but the parents are there as his or her teacher and parent corrects that.

Now, this is not a cat, this is a dog that your pointing to. So, now the child knows that dog has got some specific pattern characteristic, a cat has got some specific characteristic. So, when he sees more number of cats and sees more number of dogs, then it is possible for the child to distinguish. That this category of four legged animals are cats and these category of four legged animals are dogs.

Now, lot of times a child learns by himself through associations, he makes mistakes, he explores lot of things on his own, he makes mistakes he corrects. So, learning we have got two types of learning with a teacher and learning, without a teacher or a sort of auto associations that takes place. Now, the third thing that one has to talk about the characteristics of the neural networks is what is called as adaptivity.

(Refer Slide Time: 25:44)



Now, the neural networks they can adapt their free parameters to changes in the surrounding environment so, this can adapt the free parameters. In fact some, a few may be feeling little confused that what exactly do I mean by the word free parameters? I will come to that later on actually in respect of the human brain, the free parameter basically refers to what is called as the synaptic connection. And it all is tuned by the strength of the connection, but we will come to that later on.

So, at the moment let us accept this word free parameter, where the explanation of that would come little later. Or if it is not clear through this lecture naturally with the next few lectures, you will be able to automatically understand this. Now, this can adapt the free parameters to the changes in the surrounding environment. Well, you see naturally you can understand one thing that we definitely have to go through the process of learning throughout our life, in some sense are the other.

Now, the world that was there during our childhood is not the same world, that we are seeing today. There are so many changes, so much of developments have taken place in the scientific and technological world, our life style changed all together, our culture has gone through changes. There are changes everywhere and you see that, still we are able to cope up with this world.

Now, how is it, there are so much of changes in the surrounding environment that we are seeing around us, but we can adapt ourselves. That is a capability which we the human beings are having. And that we are doing by making some internal adjustments or the adjustments of the free parameters, that we will come to little later. Another characteristic of the neural networks is that, it not only gives us the response, now response I was referring to basically what I said was that, there is some kind of when I was talking of the input output mapping.

There is a definite input that we are feeding and we are expecting some response or the output from it. Now, at the end of learning, it is able to get the correct response, but a neural network can not only report what the response is, but it can also tell that it is the response with what confidence level. So, in that we can say that the neural network is able to give, what is called as evidential response. And just see that, we the human beings, lot of times we give the response with some evidentiality in it.

Like we can always say that, I think it is going to happen that way, we associated the word I think. That means, to say that we are associating some kind of a confidence measure. Now, that we are 100 percent confident, but may be when we say I think with a good degree of confidence we can say that yes, the feeling is that it is going to happen. So, that is to say associating a confidence with the decisions. So, it is not only a decision, but it is a decision with a confidence measure, with a measure of confidence.

So, all this we are telling as the characteristic of the biological neural network systems. Now, whether all these things could be mimicked into the artificial neural networks, or not is something that, we have to explore later on. Another very important characteristic which the biological neural network system exhibits is that, it is ability for fault tolerance. Now, what happens if supposing one particular nerves cell is malfunctioning.

Or let us say that one connection from one nerves to the other, a single connection is somehow not working. Is it so that our entire nervous system is going to collapse?

Because of that no, we can still carry on with our normal activities without any noticeable change. If too many neurons are affected may be that we will have some effect of it, but it is not something that is leading to a catastrophic failure. Whereas, unless you built in some fault tolerance into the computer system.

You know that unless that is purposely built in, then the failure of one processing unit could very often lead to a disaster, the entire computer system can collapse or the entire network can collapse. This sort of catastrophes can happen, whereas with the biological neural networks, if some neuron malfunctions or if some connections are malfunctioning. All that it leads to is some kind of a degradation in the performance, certainly not a catastrophic failure and that degradation is what is called as graceful degradation.

So, graceful degradation in the sense that, it all depends that how much of fault has taken place, if the fault is too many then the degree of degradation is large. Whereas, if the fault is not much, then the degree of degradation is small, so it is called as the graceful degradation. And in that sense the biological neural network system is highly fault tolerant. And it is possible to incorporate this fault tolerance mechanism, even in the artificial neural networks also.

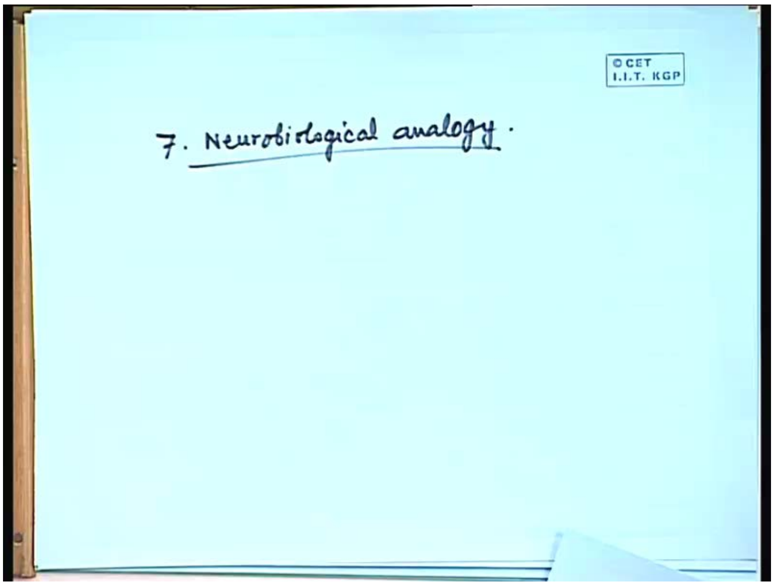
The next point that we must point out, in fact this is motivated by the fact that, there is a massively parallel computation that our brain is doing. Now, if you have to list about the capabilities of an artificial neural networks, then for the artificial neurons, we can list one of it is characteristic that the neurons, the artificial neurons that should be VLSI implementable. So, we can tell about its VLSI implement ability.

By this what I mean to say is that using the very large scale integrated circuit, it is possible to integrate a large number of neurons together. Now, naturally we cannot think of integrating 10 billions of neurons, if we could do that then we could have mimicked the human brain completely. But, we cannot do that much, but as I was telling you that, if it is utilizing network of artificial neurons to do some particular task, some particular application.

Then naturally we could be able to do that using the VLSI implementation and in fact, it is possible that way, because the neurons are absolutely parallel computational unit. Now, the neurons which are existing in a system, which are forming a network they can all do independent computation. It is one is of course, dependent upon the others, but there is a large number of parallelism, there is a good degree of parallelism that is involved with it.

And in fact, when we see the particular neuron structures in the subsequent lectures, this aspect will be more and more clear to us. And then, coming to the basic motivation of the artificial neurons.

(Refer Slide Time: 35:55)

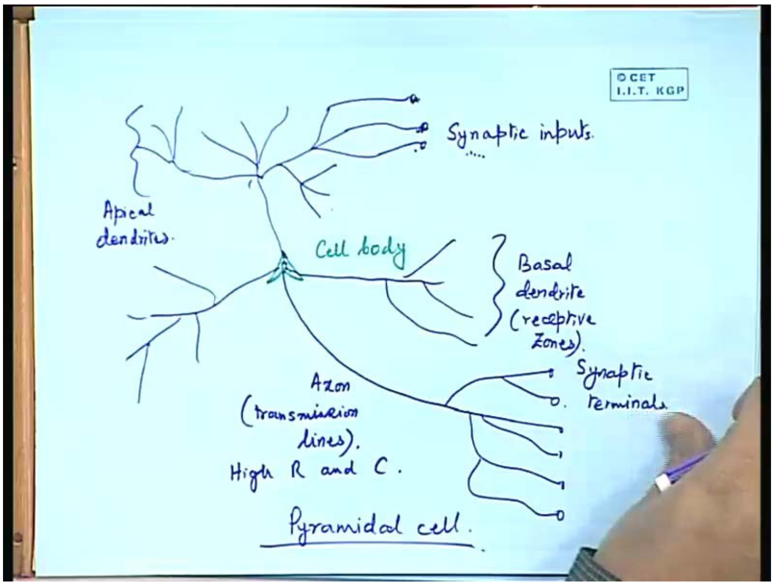


Another point that we should say is it is, neurobiological analogy, now everything as I was telling you was motivated by the biological neural network system. So, this 7 points are basically dealing with the properties of neurons, but all these properties could be imparted to the artificial neurons as well. They are the properties which the biological neurons fulfill and they are the properties which the artificial neural network is also can be made to fulfilled.

Starting from the aspect of nonlinearity, the input output mapping or the learning mechanism, the adaptability; the adaptability is a thing which we can do, but even a an artificial neuron also can be made adaptive, if we train it that way. Then coming to the evidential response, there also we can train a neural network, artificial neural network to do that. The fault tolerance, the VLSI implementation and primarily that everything is biologically, neuro biologically motivated.

Now, we should before going into the depth of the artificial neural network, we should see the structure of the human. Now, rather than drawing a human brain, we will be considering a typical nerve cell, which we refer to as pyramidal cell and that should look something like this, let us say that.

(Refer Slide Time: 37:59)



Now, here we draw, what is called as a cell body, so supposing this is the cell body and with that we now connect the following, we connect a thing which is called as axon. And this axons are basically acting like transmission lines, the lines that can carry electrical signals. And in terms of a it is electrical characteristic one can say that, it is having a high degree of electrical resistance and offers a large capacitance.

So, this has got high R and high C, now this axons they act as the transmission lines for carrying the electrical signals. And they end up with the, so here you can see that there is some tree like structure. And all this ultimately end up in synaptic terminals, what is called as synaptic terminals and these are basically used for making connections with the other nerve cells. Now, this basically is nothing but, the output part of the neuron, I should have shown the input part first, but let us draw that.

Now, these are the response zone, that is to say with the synaptic connections and that will lead to the other neurons. And then, we have here this is referring to the receptive zone. And we call this as basal dendrite, which is basically referring to the receptive zones from which we can get the inputs. So, these are the basal dendrite and then, we have the apical dendrite in fact, as compared to the axons the dendrites are having more number of branches.

Whereas, dendrites are much smaller as compare to axons, now typically this lengths, the lengths of axons are much larger than that of the dendrites. So, these are the apical dendrites and this ultimately would be connected to the synaptic inputs, these are all going to synaptic inputs. So that, it can receive signals from the other neurons. So, you see that this is the complete what is called as the pyramidal cell. So, now the pyramidal cell as shown here, this can receive synaptic inputs from other neurons.

And these will basically carry the signal all to the cell body. So, here the processing part will be done, where all these inputs will be combined. And how they are going to be combined, they will be combined in accordance with the strengths of this connections. Now, all the connections are not of the same strength, some connections are very strong and some connections are weak. Now, if the connections happen to be very strong in that case, the signals strength also will be large there.

Like here if this connection is strong, then the signal that will be contributed by this input will be much more than, if this happens to be a weak synaptic link, in that case the signal coming from here will be weaker. The signal from this may be weaker, so like that the strength of the synaptic connections will decide, that ultimately what signal, what is the net signal that will come to the cell body as the net input.

Now, that will ultimately decide that what a response is going to be and the response will be transmitted through the synaptic terminals to the other neurons. So, it has got a set of inputs, this is also connected to a set of outputs, because ultimately this is one pyramidal cell which is considered within a very large, massively parallel network of neurons. Now, as I was referring to the free parameters sometimes back, the free parameter essentially refers to the strengths of this synaptic.

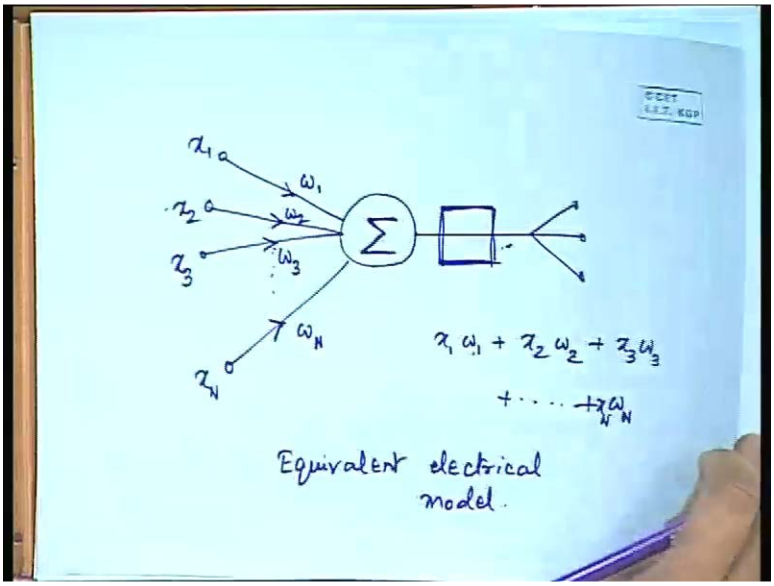
Now, as I was telling you that every input is associated with some synaptic strengths, it is connection strength. Now, supposing initially there is some a prior connection strengths that one can take. And then, accordingly there will be some response and that response could be different from that of what is desired.

So, if the actual response is different from that of the desired, naturally we have to adjust the internal parameters of the nerve cell. And what is that internal parameter it is this strengths of the synaptic connections. So, what we do is that we now alter the connection strengths and then, we feed the same inputs and find out that what the output is going to be. This time we may be closer to that of the desired, but still we may not be exactly equal to the desired.

So, that it will go through another round of synaptic in strength modifications. And this could go on alternatively till the actual response is close to that of the desired response, we may not be able to achieve exactly, but may be able to achieve close to that. So, the synaptic strengths are basically dictating that what the signal strength is going to be and these are basically acting as a free parameter to our biological nerve cell processing.

And equivalently in the artificial neural network, then we should be building up something, some electrical equivalent of such synaptic strengths.

(Refer Slide Time: 45:58)



In fact, in an equivalent electric model electrical model, we can represent a neuron like this that, a neuron or a nerve cell will be connected to several inputs. Let us say that these are the inputs, and it will be again connected to one or more than one outputs. And these inputs will be connected to the nerve cell, through some strengths of connection and we may be indicating the strengths of the connection, right on top the of this arrow we can write down, that what is the strength of this connection.

Let us say the strength of this connection we call as w 1, the strength of this connection we call as w 2, this as w 3. And supposing there are n number of such inputs connected and we call this strength as w n. And we are having here the input signal available to be that of x 1, here the input signal that is available let us say it is x 2 here, x 3 is available as input and here x n is available as input.

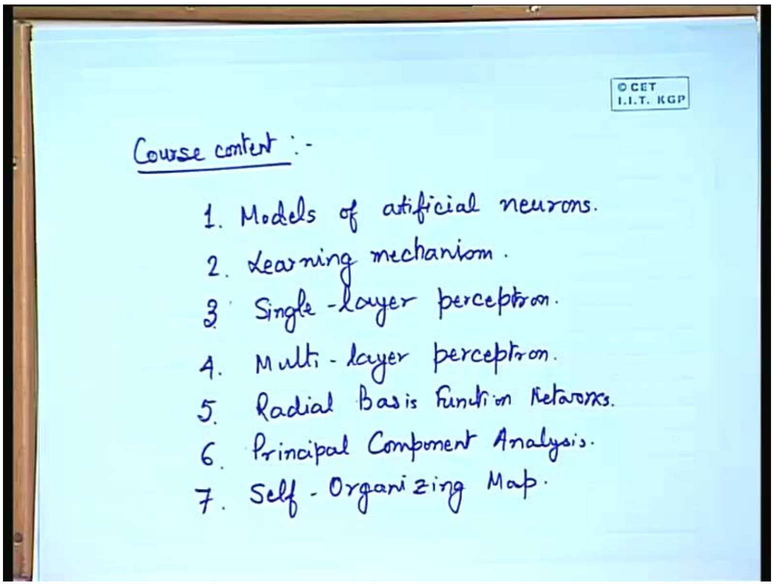
So, here what happens is that the net signal that will be available at this neuron is going to be x 1 into w 1 plus x 2 into w 2 plus x 3 into w 3 plus so on, up to x n into w n, there are n number of such neurons that is interconnected. So, this unit will be summing up all these responses, but then this is a linear summation. But, ultimately do we want a linear summation no, we must be wanting some decision out of it, that whether yes or no or a decision which is more quantifiable.

We do not want this competition alone, so in order to arrive at a decision this must be followed up by some non-linear processing unit. There must be some non-linear unit that will follow this summation. And effectively the output that will be available at the output of that non-linear unit whatever we have is going to be our output of the neuron. Now, if that output is different from that of what is desired out of it, then what we have to do, we have to simply change the strengths of this connection.

We have to change w 1, w 2, w 3 up to w n, so that with this same set of inputs, with the inputs still remaining at x 1, x 2 up to x n, we can then obtain an output that is close to the desired response. So, this is the equivalent models, so this is the biological model of a nerve cell with it is inputs and outputs. And this is the equivalent electrical model. In fact, we will be considering such equivalent electrical models of artificial neurons in our course.

Now, we give you an idea about what we are going to study in this course. So, our lectures series will be divided into some specific modules.

(Refer Slide Time: 49:50)



So, as the course content, we can that we will first begin with the models of artificial neurons. That is what we are going to cover from the next day, models of artificial neurons. And then, we are going to consider the learning mechanism, the different mechanisms of learning, the supervise learning unsupervised learning. The different conditions for learning, the auto associations and all this things we will be seeing in this chapter.

And then, we will come to what is called as the single layer perceptron. In the single layer perceptron what we will be doing is that, we will be considering a network of neurons where we will be having the neurons organized as just a single layer. So, we will be having the inputs, then we will be having layer of neurons and then, we will be having the outputs and the layer that is available is a single layer.

Whereas, this idea is later on extended to achieve, what is called as the multilayered perceptron where other than the input and the output, we are going to have some intermediate layers of processing. So, that is the realization of multilayer perceptron, which we will see in fact, we will be seeing that the single layer perceptron models have got a lot of limitations. And those limitations are can be overcome using the multilayer perceptrons.

Then, after starting the multilayer perceptrons, where one can use the multilayer perceptrons to solve what is known as the problems which are not linearly separable, they can be solved using multilayer perceptron. And they can also be solved using what is called as the radial basis function networks, that also we are going to cover in this lecture series. And then, we are going to study what is known as the principal component analysis, which is based on the eigenvalue decomposition technique.

And we will see that however, neural network can be very effectively utilized to perform the eigenvalue decomposition, into project given data into the eigen space for the purpose of dimensionality reduction. And then, we will come to class of hum unsupervised or self organizing networks which is called as the self organizing map. So, this will be under the scope of our study in the lecture series. So, now that is all for today and from the next class, we will begin with the artificial neuron model.

Thank you very much.