Neural Network and Applications

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

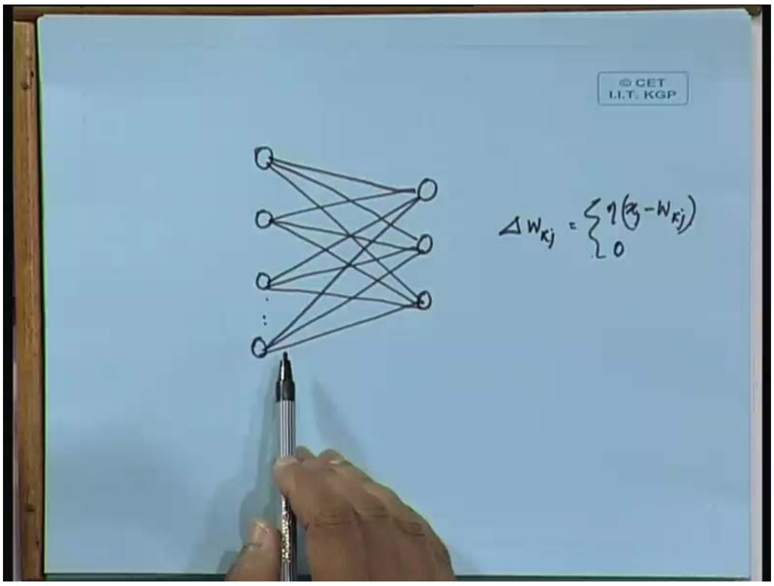
Lecture - 06
Associative Memory

Going to be on the Associative Memories, but of course, before we initiate this topic, whatever we were discussing in the last class, that is to say we were discussing above the learning mechanism. And in that we were especially discussing saying, towards the end of the last class, we were discussing about the competitive learning. Where, we have to see 1 or 2 more aspects of it, so that the concepts go very clearly in our minds.

And then, we are also going to consider the last aspect of learning, which is the Boltzmann, because as you remember that we had listed five categories of learning. So, we have done 4 of them and the fifth one that is Boltzmann learning is still the 1; that we have to do. And after, we finish of that, we will begin with the topic for the today, which is associative memories.

So, I do not know that, how much of it, I can cover today, but at least today I will at least make an introduction to the associative memory and today extend possible, I will go into the depth of it, so this is going to be the topic of today. And let us continue discussions that we were having in particular about the competitive learning. Now, there are a few questions, which I think can come to the mind of all of you and that is to set at well and fine.

(Refer Slide Time: 02:33)



We have a network, let us say certain numbers of inputs are there and they have got some outputs. In fact the number of outputs that we are going to have for the competitive network is competitive learning network is going to be equal to the number of classes that we would like to have. In fact, what type of network is it, is it a supervised or an unsupervised type of learning.

What kind of category would you put it into supervised, unsupervised, clearly unsupervised, why because here we are not providing any desired response to it. We are allowing the network to evolve on it is own. So, what happens is that, we will be having certain number of input units. We will be having some number of output units and the interconnections would be made like this. So, that these input to output will be having their own synoptic weights.

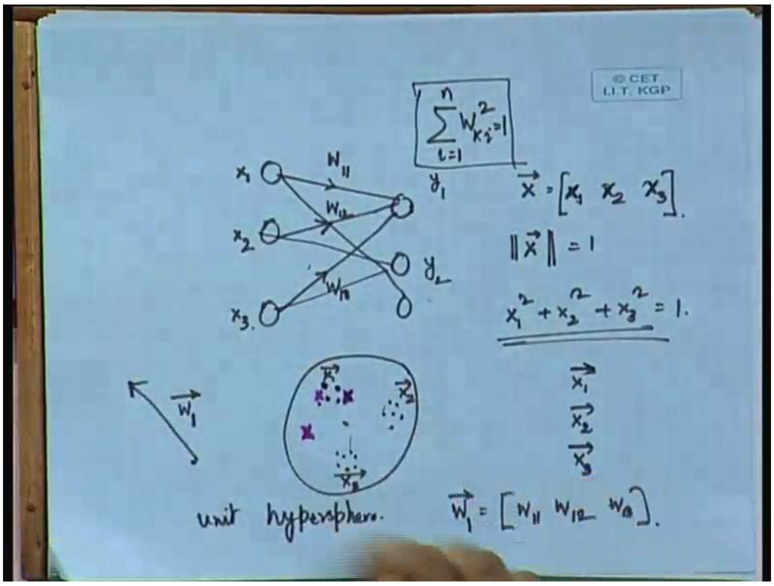
And then, the competitive learning networks as we said, that goes through a process of ((Refer Time: 03:51)) in the sense that, one of the output units will be in the competition and that will actually classify the pattern. That is to say for a given input pattern, one of these outputs will be the winner. So, it could be that, if you feed a different pattern next time, then another network could become the winner.

So, it all depends that how the interconnections weights between them are adjusted and the learning mechanism that we had seen in the last class, clearly saying that the delta W k j that is going to be equal to eta times x j minus W k j. So, that you had seen that there is a very clear preference or rather to say, we should put it this way, that the weights, that we are going to have is ultimately going to be aligned to the input pattern. Because, for a given output neuron k, if it is having connections to many different inputs, then the synaptic weights which will be given by W k j, that has to get align to x j.

Now, if it gets aligned to that; that means to say what, that means to say that next time you repeat the same pattern. Now, the weights are already aligned for that and there is a greater chance that or rather, it is not a matter of chance, it is inevitably going to happen. That if it is a winner in the earlier step, then for when the pattern is going to be presented for the next time, the same output units definitely going to be the winner.

In fact, it will be winner in a stronger sense, why because the weights have been already aligned. Now, let us try to get this concept in a form of some kind of a geometrical interpretation, let us say that, we have got three input neurons.

(Refer Slide Time: 06:08)



Supposing, we have got input neurons, which we are calling as let us say, the inputs are x 1, x 2 and x 3. And let us consider two outputs, let us say, that we have the output units to be y 1 and y 2, they are the respective outputs. So, we are going to call, all these weights as W 11. This one will be W 1 2 and this one will be W 1 3 and likewise for two we are going to have it as W 21, 22 and 23.

So, here let us also put some constraint, that in this case this X; that is going to be in the form of a vector. So, if we define X as a vector, then it is elements will be x 1, x 2 and x 3 and supposing, we define that the magnitude of this vector is equal to unity. That means to say what, if we define that this is equal to unity, this means to say that x 1 square plus x 2 square plus x 3 square, that is going to be equal to 1. But, we can have large combination of patterns, realized out of it, given the three inputs x 1, x 2 and x 3; we can realize a large number of patterns.

Now, what will be, where will be these patterns lie, what will be the locus of all these patterns, in which space are they going to lie. Sphere, it is going to lie on the surface of a sphere. So, in this case we have got three inputs x 1 x 2 and x 3, had they are been just two, x 1 and x 2. It would have been a circle, the locus would have been a circle and then because we have got x 1, x 2 and x 3, the locus is definitely going to be a sphere.

So, we can imagine that there is a sphere; the sphere is going to have a center, from which the vector will be originating. And we will be having for different patterns; we will be having the vectors positioned in different spaces. Let us say, that now we are going to designate the vectors by X 1 vectors, X 2 vectors, X 3 vectors like that. So, I am going to consider X 1 vector, X 2 vector, mind you these are different from x 1, x 2, x 3, because this x 1, x 2, x 3, let me write it down in lower case, so as to avoid confusion.

These x 1, x 2, x 3s are nothing but scalar quantities which composes the elements of this vector. But, in this case we are going to consider the vectors X 1, X 2, X 3 which will basically indicate the different patterns that we are feeding to a system. And how many such patterns can we feed, there is really speaking no restriction or there is ultimately going to be some restriction. That is according to the learning capacity of the network which we are going to study later on.

Right now, we do not bother; right now we assume that there could be many different types of patterns that you can feed to the system. So, when you imagine a sphere, may be that the X 1 vector will have it is edge located over here; that means to say from origin to this point on the surface. So, it is actually, we drawn as a 3 D surface, this is a sphere, which is a three dimensional surface.

So, supposing that, this is where we have the X 1 vector, this is where we have the X 2 vector, this is where we have the X 3 vector, like that. Now, we are not going to have just three patterns may be 4, 5, 6, 7 large number of patterns could be there. So, may be that here, we have got another vector here, we have got another vector here, we have got another vector, like that many vectors are there. And let us say that, here also near around X 2 we are having different patterns, near to X 3 we are having different patterns.

And let us not say that, we have only 2 output neurons, say that we have got neurons 1, 2 and 3. So; that means to say that, these type of a competitive network, competitive learning network can really classify the patterns into three different classes. Now, from this given example, we can clearly see that, there are three distinct clusters of patterns, which are existing.

We have got one cluster, which is around this where X 1 is lying; we have got another cluster which contains X 2 and also the patterns, which are nearing X 2. We have got another cluster of patterns which is close to X 3, three distinct patterns are there. So, the behavior that you should logically expect out of this neural network is that one of the neurons out of these 3 outputs.

One of the neurons will be classifying these cluster; that means to say that, no matter, whether the vector is this one or this one or this one or this one, it will classify according to 1, no matter. Then, another output neuron will be classifying this cluster and likewise the third neuron will classify this cluster; that is what it is going to do. Now, what exactly do we mean by alignment.

So, just like the way we had assumed that the magnitude of this vector is going to be equal to unity. Very similarly, we can assume that the sum of the weights that is connected to every output neuron that is also going to be equal to unity. Some of the weights mean that squared value of the sum. So, what t to say is summation W k i and this is summed up over i and summation W k i square in fact we should take.

So, supposing we have got n inputs, so that the summation W k i square summed up over i is equal to n; that is going to be unity. This means to say what, that supposing that for this pattern, X 1 or the cluster that we have shown, supposing we have feed the pattern X 1 vector, directly. So, X 1 is a vector and we feed that pattern and supposing that for this y 1 is the winner. So, y 1 is the winner means definitely W 11, W 12 and W 13 these are the weights, these are the synaptic weights, which coming to play.

So, what will happen is that, we are going to have W 11 square, plus 12 square, plus 13 square, equal to unity and because the output y 1 has now become the winner, this means to say what, it will go by the learning rule. It will definitely go by this learning rule and as a result, what happens, that the winners weights will be adjusted, all the losers weight remain the same.

But, the winners weight that is to say these ones will get adjusted meaning that this W 11 W 12 and W 13 will now be aligned to the X 1 vector pattern, is it very clear. That means to say, that if earlier we had, let us imagine in the 3 D way, that if earlier this was the W vector. And what is W vector, W vector is nothing but, we are defining it to be the vector consisting of let say W 11, W 12, W 13.

Let us call it as W 1 vector, because that is the vector that is associated with the output neuron 1. So, W 1 vectors was earlier like this, supposing this is the position or let me draw it with a different pen that will be better. So, supposing the 1 which I have marked right now, that is the position of the initial W 1 or initially W 1 may be somewhere here, but then what are we going to do. We have to steer this vector towards the pattern that has caused the output neuron to be the winner, you understand.

We have to this vector is somewhere here, supposing this is the vector, that we have got over here and supposing the winning pattern is this and the vector lies here. What we have to do is to steer this vector and align it towards the winning pattern, is this understood, anybody having any doubt on that. So, that is exactly, what we are going to do.

So, that means to say what, that initially if we have the vector, aligned from the center to this point. Now, with their adjustment of the weight, the vector will move closer to this. In fact, if for all these patterns this neuron is going to be the winner, then where would you expect the ultimate vector to get aligned, ultimately the weight vector to get aligned to the center of the cluster.

No matter, in whatever manner we take the center could be the centroid or whatever, somewhere in the center of the cluster of the vectors for which it is the winner, there the weight should ultimately get align to, this is what is going to happen. So, initially you start with random weights. The units 1, 2 and 3 the output units 1, 2 and 3 they will be having all random weights and you will keep on feeding patterns.

And now what will happen is that, with one pattern becoming a winner for one of the output neurons, that output neurons synaptic weights will now get aligned to the winning pattern. So, like this ultimately, there will be three distinct classifications, if you have more than three clusters, then of course, there is a problem. And I think, you can guess that problem, what is that problem, because we have got only three output units.

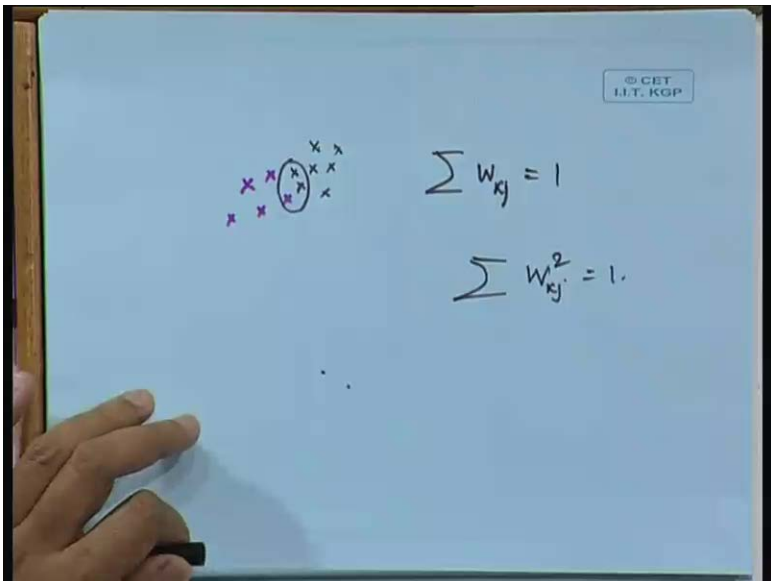
So, if we get more than three clusters, then the output units will not be very sufficient to make the cluster, so we will be having some errors. So, what happens is that, if there are four clusters, then the fourth cluster elements can get aligned to any of the other clusters and the results will not be very good. So, in fact that is the way, whereby we have to decide about the number of output neurons, when we have got a very clear cut idea about how many classes that we are going to make.

So, this is more or less the geometrical interpretation of the competitive learning network. So, it is indeed, going through a process of competition and the ultimate aim is that, the weight vector will be aligned to the winning pattern vector, that is the clear thing that we are going to do out of the competitive learning network, yes any question

Student: ((Refer Time: 18:17))

The two clusters having intersection, well there is a possibility, that the two clusters are having some kind of an intersection. So, in that case, what happens is that, near the intersection region, it can be, one of them will be the winner.

(Refer Slide Time: 18:42)



So, meaning to say, that the question that has been asked is that, supposing this is cluster 1, let us say and supposing we have got another cluster, which is cluster 2. And here said that the cluster 1 and cluster 2 will be having some kind of intersections. So, what happens is that, this is more or less the kind of an intersection zone. So, what happens is that out of this again, ultimately one of these will be belonging to this cluster, some neurons or some patterns will be belonging to the other cluster.

So, that will again depend upon the certain adjustments of the weights, but ultimately the two clusters are going to merge. So, there could be some little bit discrepancy in the results. The results are very much ideal, the classification is very much ideal, when the clusters are distinct. And the number of distinct cluster that we have is equal to the number of output neurons that we have in a system.

As long as that is clear, the classification is very accurate, otherwise the classification is not that accurate, there could be some sources of errors in that. Any other questions that you may be having in your mind

Student: Sigma W k j

Sigma W k j, yes

Student: Equal to what

Yes, so that was one of this question is coming that last class, I said sigma W k j is equal to 1 and this class, I am saying sigma W k j square is equal to 1. This square thing, I have only said from a geometrical interpretation point of view, so that we can imagine that this is on a sphere. That is the only reason, why I took this to be W k i square is equal to 1.

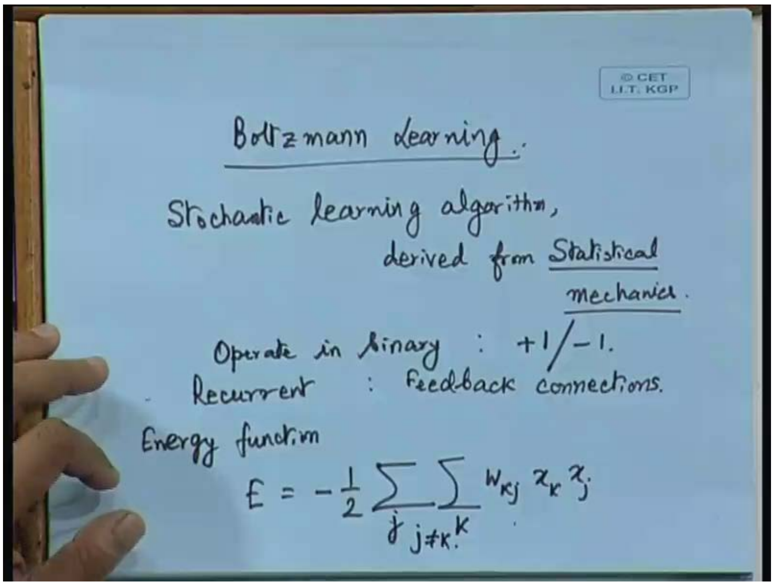
So, ultimately the weights are going to have some constraint, whether you define it in the terms of summation equal to 1 or summation square equal to 1. The weights are in fact having some kind of constraints between them; it is not that you can keep on increasing the weights indefinitely. There is a limit; you have to keep a check on that and another thing I should have told you that, in this case I have taken example of a three dimension.

So, does it meant to say that, I cannot take more than three dimension, we can definitely take, we can take four dimension as input, we can take 4 inputs we can take 5 inputs. In which case our input pattern is going to be a five dimensional vector or multidimensional vector, it is going to be n dimensional vector, it is going to be. So, in that case, what will be the geometrical interpretation of that, where will the patterns lie, on the surface of unit hyper sphere.

So, this will lie on the surface of unit hyper sphere, if we imagine that summation of X i square is equal to 1; that means, to say that the magnitude of the input vector is equal to unity and also this is hyper sphere, because it is more than three dimension. So, we cannot really conceptualize, we cannot really visualize, the existence of this sphere, but it is definitely a hyper sphere.

So, now let us go over to the next topic that we were going to cover on the learning aspect and that is called as the Boltzmann Learning.

(Refer Slide Time: 22:24)



Boltzmann learning and we are going to cover this shortly, so now Boltzmann learning is actually a stochastic learning algorithm. So, we have already seen about the stochastic learning aspect, the one which is not deterministic simply that. So, this is a stochastic learning algorithm and this is in fact derived from statistical mechanics. Minding you, at this stage, I am not going over to any discussion on this statistical mechanics aspects or the details of Boltzmann learning, that we will not be doing a here.

We will only see in a very nutshell, about the learning mechanism in it, without going into much of details. We will cover the details, if we have time towards the end of the course. Now, the neurons basically constitute a recurrent structure and the neurons in a Boltzmann leaning, they constitute a recurrent structure. And what is meant by a recurrent structure that means to say that, there is a feed back. There are feed backs that exist between neurons, they are all interconnected and they are having feedback.

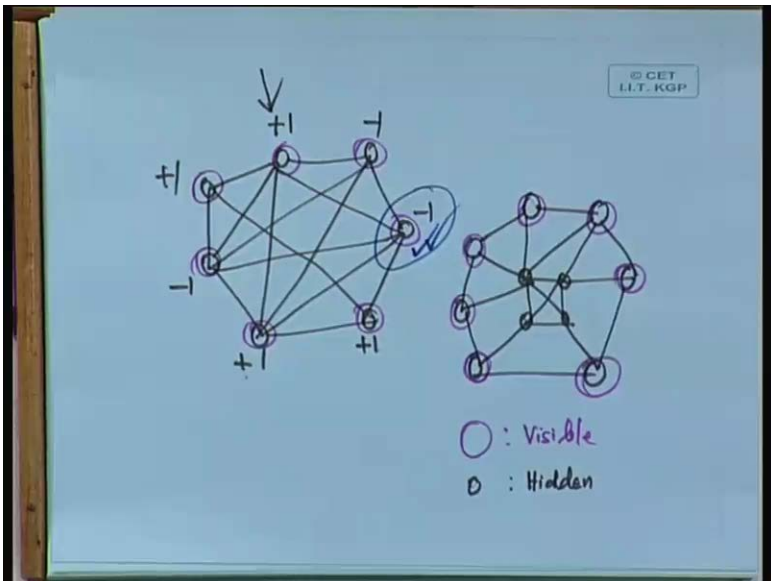
So, neurons constitute a recurrent structure and they operate in the binary mode, so meaning that it is excitation could be either plus 1 or minus 1. So, now we can characterize this kind of a machine, by an energy function. So, this is recurrent, recurrent meaning that, we are going to have feedbacks interconnections amongst the neurons. Feedback connections between the neurons, as also self feedbacks, that is also included.

But, of course in the energy function that we are going to define, that the self feedback is not included. So, the energy function that we are going to define for such a kind of a network is given by E, which is equal to minus half summation, this is by the definition. Summation of W k j x k x j and we sum it up over all the j's and all the k's, but with the condition that j is not equal to k.

All the j's and all the k's with j not equal to k, meaning what, that the self feedback is not considered in this case in the energy expression. So, we are not considering any W k k term or W j j term and just simply interpret what it means. You take a neuron labeled as k and you take another neuron labeled as j and you have a connection from the neuron j to neuron k and that connection is W k j, k j is from j to k, do not make that mistake.

So, it is W k j x k x j, so x k x j are the inputs and inputs could be either in the level of plus 1 or minus 1 and there will be some W k j, some weight will be associated. And we are summing up all these products and together that will be signifying an energy function for this entire network. Now, what it means is that in what sense is it stochastic, because we said in the beginning itself, that it is a stochastic learning algorithm. So, it is stochastic in the sense, that in this kind of a network.

(Refer Slide Time: 27:03)



So, imagine that you have got network like this, may be some neurons which are interconnected amongst them. So, like this the neurons are interconnected amongst themselves and not only that, we have drawn this picture, where you can say that, we can take the outputs from any of these neurons, there are how may 3, 4, 5, 6, 7 neurons, that I have drawn.

So, it is a seven neuron system and we can pick up any neuron as output unit or rather to say that any output pattern could be defined as a vector of all the seven states put together. But, in a general Boltzmann learning network, we could be having some neurons, which do not take part in the output. So, meaning that we could be having a set of neurons, which are there in the outward outer layer and then some set of neurons, which are there in the inner layer, which serves more likely hidden there.

Hidden layer in the sense, that they do not really take part in the output, means we cannot specify, we cannot specify their states in the outputs, but we can take all these neurons, for the output. So, basically what happens is that, in that case whenever we have got such kind of layering's that some of the neurons are available as output and some of the neurons are not available, they are hidden. In that case, we just segregate the neurons into two categories.

The one that is available at the output, let say we mark them with a different color. So, all these neurons which I have marked with a different color, those things we are calling as the visible neurons. And once which are there inside, the once which are not colored, which are marked as black only, they are hidden neurons. So, we will be typically having visible neurons as well as hidden neurons.

But, what I want to discuss is that or say for example, we may be having in this case, there is a Boltzmann network, where all the neurons in the output layer, they are visible, all the neurons themselves are visible. So, there is no hidden layer in that in this Boltzmann network that I have drawn. Now, what the Boltzmann network is really saying is that, when you have energy like this, this energy is defined for what; this energy is defined for a particular state of the network.

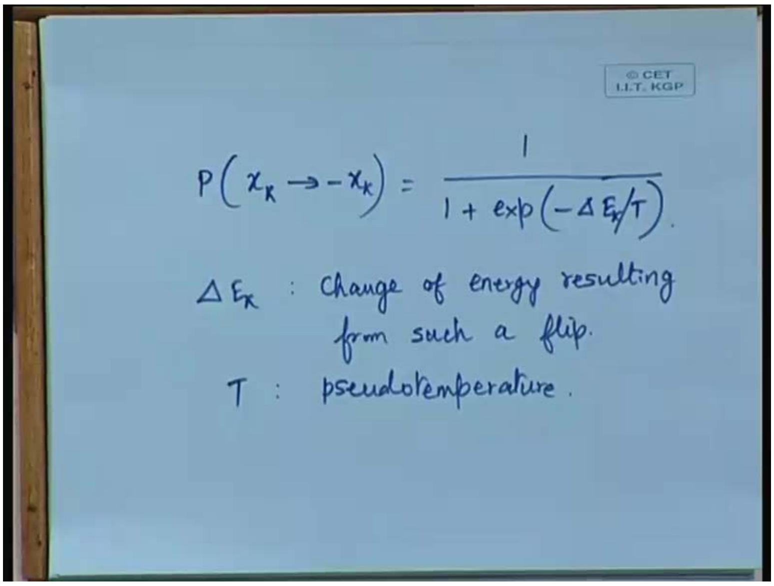
And what is meant by the particular state, you take a network like this, let us say. And supposing you have got some particular state, say this is at plus 1, this is at minus 1, this is at minus 1, this is at plus 1, this is at plus 1, this is at minus 1, this is at plus 1, something like this. So, this is a state of the neuron that we can describe, supposing we begin with this, so plus 1, minus 1, minus 1, plus 1, plus 1, minus 1, plus 1, this vector, if a form it into a vector that will define the state of this.

So, for this state of the neuron, we will be having some energy function, now what happens, if I randomly pick up any one of these neurons. Supposing, I just simply decide to pick up this particular neuron, supposing I pick up this 1 and then I after picking up this neuron, I change it is state, I flip it is state, from minus 1, I make it to plus 1. Then, what happens and then we have to recompute this energy, because it was computed on an earlier energy, based on the earlier state.

Now, that I decided to flip it, the energy is going to change and what happens that, there will be a total change of energy, it could increase, it could decrease, but I can pick up anyone and I can make it cause the change of state. Now, what it means in a stochastic sense is that, what the probability is that there will be a change of state for any picked up neuron k.

Supposing, if I pickup any neuron k, then what is the probability, that it is state will be flipped that probability is stochastic in nature and it is given by this.

(Refer Slide Time: 32:02)



So, probability that a neuron k flipped, it is state from it is current state x k to minus of x k, this probability is given by 1 by 1 plus exponential to the power minus delta E k by T. And what are the meanings of these terms, what is delta E k; delta E k is nothing but the change of energy. It is the change of energy resulting from such a flip meaning by applying this equation, because we decided to flip, one particular x k.

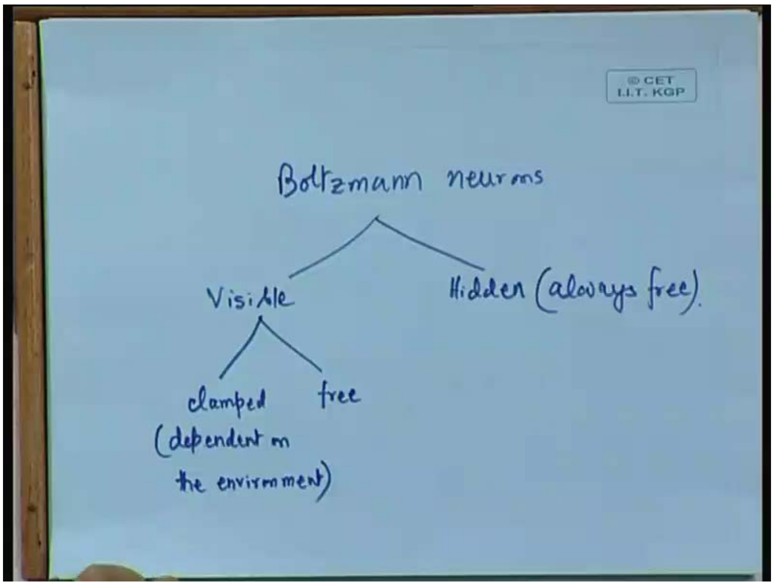
Then, what is the change of energy that is there and what is T, as usual is the pseudo temperature. So, T is the pseudo temperature, which we had defined earlier, when we talked about the stochastic activation units, there we came across this concern. So, what happens, that depending upon this delta E k and T. T will definitely signify the noise in the system.

Because, basically that concept of the temperature comes about that with increasing temperature, the network becomes more noisy, more stochastic, less deterministic, like that. So, ultimately what happens is that, there will be a condition, when the Boltzmann network will have a steady state. That means to say that ultimately, see initially, when you allow the network to settle.

In that case, all the vectors will be random; there will be some random pattern that we are feeding random weights that will be there in the system. And now it goes through such change of states and change of states will also happen, because every change of state will mean that, there will be significant change in energy, significant change in delta E k will be there. As a result there will be more probabilities of change of as we have shown.

As a result the change of state will happen with more frequent probabilities, but ultimately there will be a state condition, when the probabilities of flips will get reduced, which means to say that the network will attain some kind of a stability. We can say that the network has attained the thermal stability. But, one thing that one can tell about the Boltzmann network is that, firstly that we have already classified the Boltzmann neurons into two classes.

(Refer Slide Time: 35:21)



So, Boltzmann neurons we have already classified to be the visible, one that is available for the outside and hidden. Now, one thing that we have to understands that can we change the states of all the visible neurons, as per our will or can we allow. All the neurons, which is available at the output layer or which is available in the visible way, can we afford to change everything, perhaps we can, perhaps we cannot.

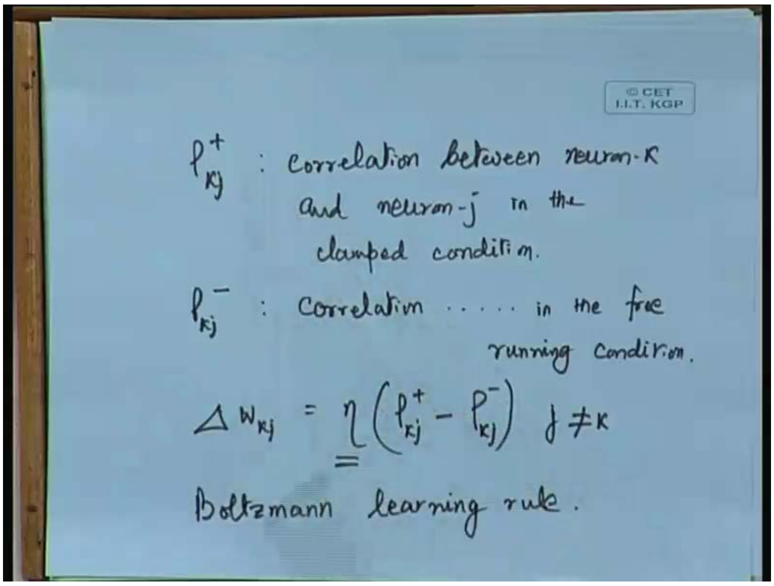
In which case, we cannot if we happen to put some constraint on the system, because ultimately this kind of a neural network, Boltzmann network, you are going to uses some application, know that application may put forward some constraint. That this, this, this, this neurons, I do not want change of state to be there, may be that if you have 100 neurons visible, you can decide that out of that, I will not be allowing change of states for this 15 neurons, these are my constraints.

I am clamping it, that some of these will be permanently as plus 1, you can never change a state, some of these will be permanently in minus 1, you can never change a state, like that, we clamp the states of those neurons. May be out of 100, I specify that 15 are so clamp and remain 85 are free, free parameter adjustments are possible with remaining 85.

And when it comes to the hidden, then there is no question of my specifying constraint, because the hidden neurons do not take part in the output, so the hidden neurons should always be free. So, there is another classification that we are doing, that is the visible network, the visible neurons, we are classifying into two groups. One is the neurons, the visible neurons which are clamped, clamped in the sense, that their states are pre decided, so this says dependent on the environment.

So, depending on the environment, we are going to decide that some of the neurons will be clamped and there will be some other neurons, which will be free and hidden neurons, they will be always free. Now; that means, you say that, when we have some free neurons and free and some clamped neurons, then how does the system learn, that is what we are going to study, so the learning could be defined like this.

(Refer Slide Time: 38:20)



That, we consider, let us say that we denote rho k j plus to be the correlation between neuron k and neuron j in the clamped condition. Any suggestion, how can I mathematically define a correlation in this case, any simple mathematical way; that you can suggest. Two neurons, one is neuron k, another is neuron j, so we are going to define a correlation for that.

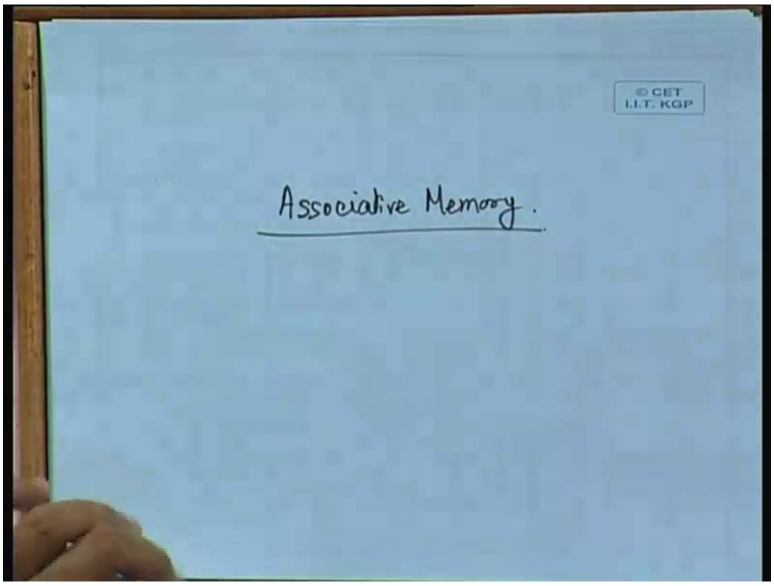
So, what is going to be a simplest form of correlation, when you are going to say that neuron k and neuron j is correlation, when their states are same, say both of them are plus 1, it is correlated, if both of them are minus 1, then also it is correlated. But, if one is plus 1 and another is minus 1, it is uncorrelated. So, what is the measure of correlation, you simply multiply their states.

So, if the multiplication gives you the value of plus 1, then it is correlated, if the multiplication gives you minus 1, it is uncorrelated. So, we can consider that this rho k j's also are binary quantities, plus 1 or minus 1, plus 1 for correlated case and minus 1 for uncorrelated. So, we are saying that rho k j is the correlation for the neuron k and neuron j in the clamped condition, mind you.

And we again define another correlation parameter, which we are defining as a rho k j minus, which is the correlation between the same neurons in the free running condition. In that case, delta W k j will be defined by again, whatever we do for learning, this fellow is unavoidable, the learning parameter, it will be always there. This eta, the learning rate times rho k j plus minus rho k j minus and again, this will be defined for j not equal to k and this is our Boltzmann learning rule.

So, it is definitely a stochastic learning rule, that we have decided, so based on this factor, the W k j, the weight adjustment will be there. But, more than that, I am not discussing about the Boltzmann network at this stage, because Boltzmann network really needs some more detailed explanation, which we will see later on. So, now we can afford to begin the topic that we promised for today that is about associative memory.

(Refer Slide Time: 41:53)



In fact, in the remaining time, the only thing that I can do is to give you the concepts about the associative memory. Now, just know the two words in it, I said associative and I said memory. Now, how are these two things really related, in fact, if you see the learning mechanism, how do we learn, forget about artificial neural networks, let us come to our case.

How do we human beings learn, we learn a lot, using association, we can always relate, based on our experience, we can relate that, this is something that I had seen there. Supposing, in front of me, I have got students from IIT, who are attending this course. Now, there are large number of students, I may not be able to recollect the names of every students, there are 80 or 70, 80 students in this group.

So, then what happens, that supposing, I see one of the persons in the market. Now, what I am going to do, I am certainly not going to recall the name of that person, because I may not know the name, unless he is a very familiar person, a student whom I have observed for 1 year or 2 years or he has attended, he or she has attended my earlier course. So, I may be able to recollect the name immediately, but then what I am going to do is that, I will think that yes it is a very familiar face and where I have seen this person.

So, then I recollect that yes, I have seen him in the class, he is the person, who is attending the neural network course. So, there is a kind of an association that I am making, that for the neural network course, I have got an association, whenever that face is presented to me as a pattern. So, what is that face, that face acts as a as an input pattern to me and then I put him into the category, that I have seen him in the neural network class, so there is a kind of patterned association that we are doing.

So, it is actually, whenever we are associating anything, we are recalling from our memory and then only we are associating. Now, what is the concept of a traditional memory, because memory all of us understand because, nowadays everybody is using computers and we know that computer has got large amount of memory. We have got memory in the form of RAM; we have got secondary memory in the form of hard discs and what not.

So, then what is the difference between this memory and that memory, then there we feed the address of the memory, there every memory, the memories are actually arranged in continuous locations and every memory location will be having some address. So, I will be giving the address of that memory and then, I will be retrieving the data corresponding to that address. And I will use that; that is what we are doing for all our computational purpose.

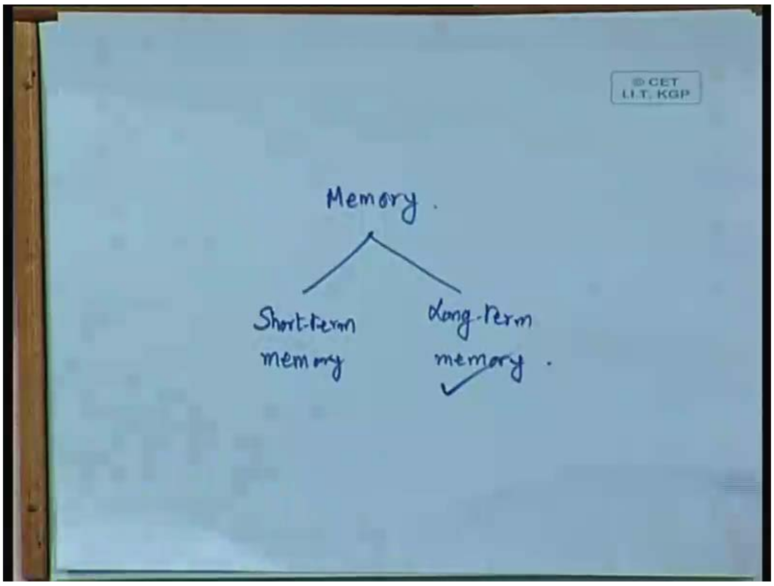
Now, here what I am doing is that I am not feeding the address, I do not really know that, if my brain has got a large amount of storage space, I am not really remembering or it is not possible for me to specify that. Where, in the memory, which address location in my memory, did I store this persons face, it is not possible for me to know. I see the face, that face acts as a pattern, that face acts as a stimulant and based on that stimulant, I go over to the memory and then get it.

So, here it is something like the retrieval that I am doing from the memory is based on content, the pattern is content. I am accessing that and then I am able to retrieve the data out of that. So, really speaking we are feeding the content and then we are getting the response from that memory. This content I have seen in a neural network course, this content I have seen him in the department office.

This is the pattern content, I have he is a shop keeper somewhere, I have seen him, so this is the kind of an association that we are making. So, it is a memory that involves very much the association and I think from all these discussion, you can very easily, understand that it is in fact the process of association and the process of memorizing, they are very much interrelated to each other.

The process of learning and the process the memorizing, they are very much interlinked to each other. Now, let us see that what kind of memories do we have, especially the memory of our brain.

(Refer Slide Time: 47:35)



Now, it is seen, that the memory, that we use, can be divided into two broad classes, one is a short term memory and the other is the long term memory. Now, if I ask anybody in the class over here that, what was the menu for today’s lunch? You are definitely going to tell me, that the menu was this, that, we had rice, we had dhal, we had chapatti and we had some subji, curry something. So, you will be able to tell me, from your short term memory.

Now, if I ask you the question, that what the menu was for you on Monday, will you be able to give me the answer. Simply say that no, we do not remember, what was a menu for Monday lunch or what was the menu for Tuesday lunch, 3, 4 days back, whatever menu was there, even yesterday’s, we will not be able to tell. So, why cannot we tell because, it is in the short term memory.

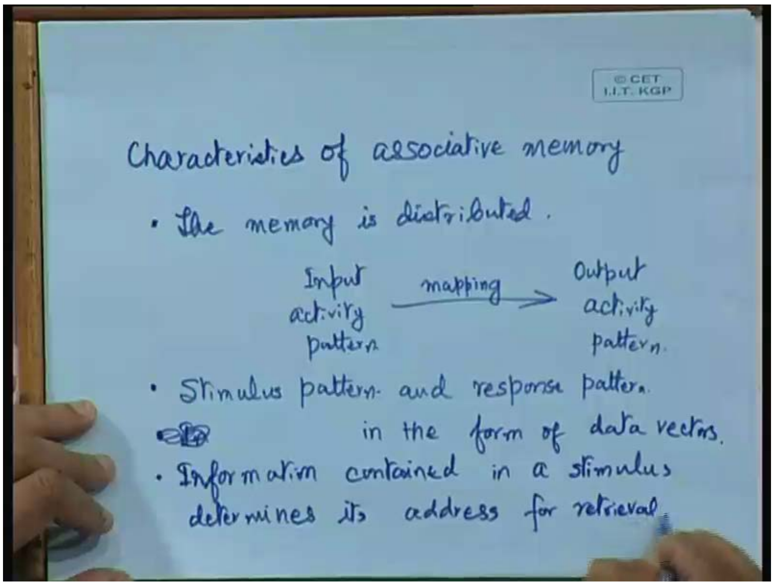
But, if I ask you that in which year, did you appear for your high school leaving examination, are you going to make a mistake on that, never. If you have passed out your schooling in the year 2000, you are going to say that, yes, it is year 2000. Even, if I ask you 15 years from now, that in which year you had your school leaving examination, you are going to say that it was in the year 2000, you are never going to forget that.

In which year, you got into IIT; you are never going to forget that, that means to say that this is something that goes into your long term memory. So, we have got these two categories, the short term and the long term. Now, one thing that, why is the short term, really a short term. Let us also, try to understand that, because the short term memory is all the time indicating some current information, the very latest and current information, which is going to be updated and corrected, all the time.

Know, why cannot you remember the menu of your lunch, you can remember only the menu of today, which you have taken just a few hours back, you took your lunch. So, you can remember the menu, but why cannot you remember for 1 or 2 days back, because it is getting updated all the time with current information. So, it is not possible for you to retain all these things into your long term memory and there is no point.

In fact, there is a kind of intention also that goes on, I do not think anybody will deliberately, try to remember the lunch menu and that to when it is hostel lunch, anyway. So, let us see the aspects of associative memory, so all that we can cover at this stage is some characteristic of the associative memory.

(Refer Slide Time: 51:10)



So, we list out some characteristics, the first characteristic is that, the memory is distributed. And what is meant by distributed, very simply, that it does not lie in one particular location, means all the memory locations are not laying in one particular place, they are distributed, even in our brain. The neurons are distributed all over the brain and this is highly distributed in nature.

In fact, what happens is that, together we are feeding some input activity pattern. So, there is some input activity pattern, that we are feeding as a stimulus to the system and then the system or our brain or whatever you say or artificial neuron, neural pattern whatever you think of, so it gives some output activity pattern. So, there is a mapping that takes place from the input activity pattern to the output activity pattern and this happens in a distributed memory environment.

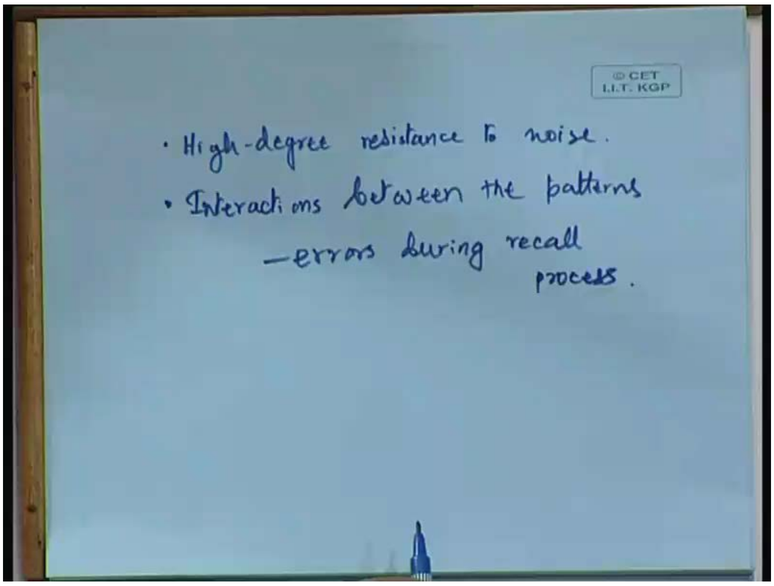
We will come to that, we will come to the details of that soon and then the second characteristic that we can say is that the stimulus pattern. What is meant by stimulus pattern, the input, like little while back the example, that I was giving you, the stimulus pattern is the face of the person, whom I am seeing right now and I am trying to associate. So, stimulus pattern, that is one pattern and then we are also going to have a response pattern.

Now, what is the response pattern, again it is not only that, I know that, I have seen this person in neural network class, but also more or less, I will be able to predict some more things. May be that normally, this particular persons sit is on the left hand side of the class or normally this is the person, who sit is on the front bench, normally this is the person, who will sit somewhere in the middle row. Normally, he prepares to sit on the extreme right hand corner of the class, like that.

So, there is a response pattern that also we get, that once a stimulus is presented, we are also going to get a response pattern. Then, the information contained in the pattern, so stimulus pattern and response pattern, they are in the form of data vectors. So, this is quite obvious, why is the stimulus pattern a vector, because stimulus pattern is in this example, for I may let us take that example, the face image, it is a stimulus.

So, what is that it contains a data, it contains a pixel data, at every pictured element and we are having some intensity, some color and all that. So, that is a data vector, at the input and response vector, again that is also a data, it is giving that in which particular location, this person seats and all that. Then, the information not only contains the storage location, but also it is address for retrieval. So, information contained in the stimulus, also determines it is address for retrieval.

(Refer Slide Time: 55:57)



And then, another very important aspect is that, it has got high, it has got high degrees of resistance to noise, what it means. This means to say, that although the neurons, the neurons are doing the ultimate processing. The neurons could be operating under noisy conditions. We could be operating the neurons in a stochastic way, but even then our associative memory works in a mode, that it has got very high degree of resistance to noise.

So, although the neurons themselves will be noisy, but still the classification that we are making the association, that we are making is highly resistance to noise. And then, the last point that we would like to mention about the characteristic is that. There is a high degree of interactions between the patterns high degree of interactions and that often, leads to errors during recall process.

Now, I give a simple example to that, I think, it happens many times in our real life. Now, somebody had once asked me, that I think, you are working in Syndicate bank, Bombay branch, I have seen you are over there. Now, I am surprised, that no I am certainly not in an officer or employee in syndicate bank and that too posted in Bombay. I am a simple person, living in IIT Kharagpur; I am just teaching over there, I am not a bank officer.

He said that no, I think I have seen you, so what happens is that, may be that my face looks very similar to somebody who is working in Syndicate bank, Bombay branch. So, they say that, somebody recollects, wrongly. So, what happens is that, there is a very similarity in the pattern, may be that, because my face resembles another person, people make a wrong conclusion, it is quite possible.

So, there are interaction between the patterns and also another thing, which is going to happen, that there is a definite learning capacity up to which we can learn and if we try to go beyond that, there could be errors in the recall process. In fact, we are going to see all these things, in terms of artificial neural network behaviors, in the next class. So, this is just only an introduction to the associative memory.

Thank you.