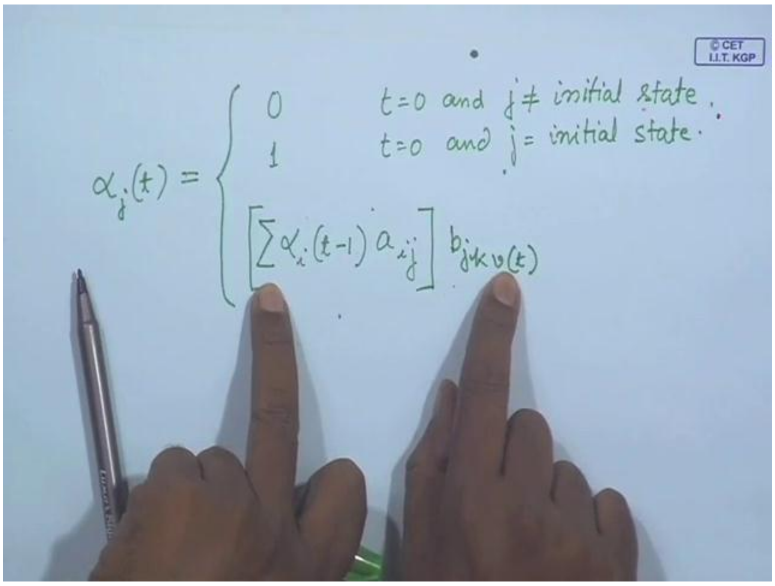
Pattern Recognition and Applications
Prof. P.K. Biswas
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 39
Hidden Markov Model
(Contd.)

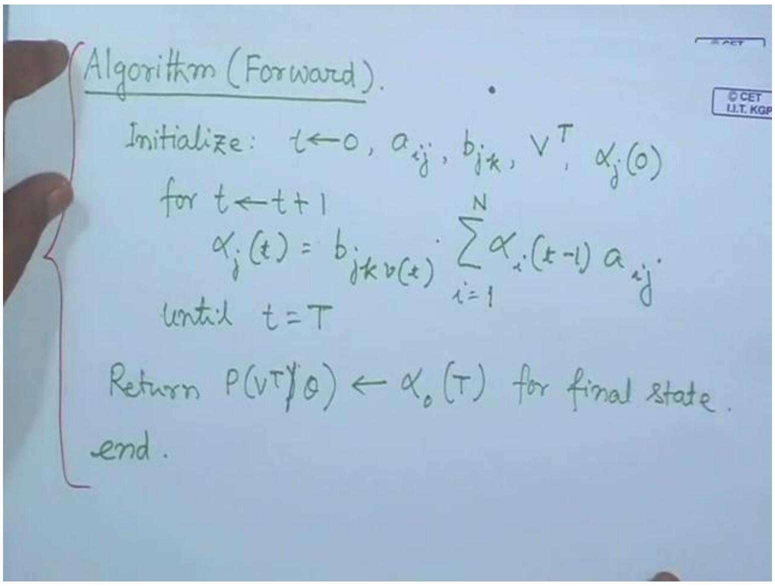
Hello. So, in the last class, what we have done is we have determined a probability that the hidden Markov model is in a particular state, say omega i at times step t after generating first t number of visible symbols from the given sequence of visible symbols b t.

(Refer Slide Time: 00:59)



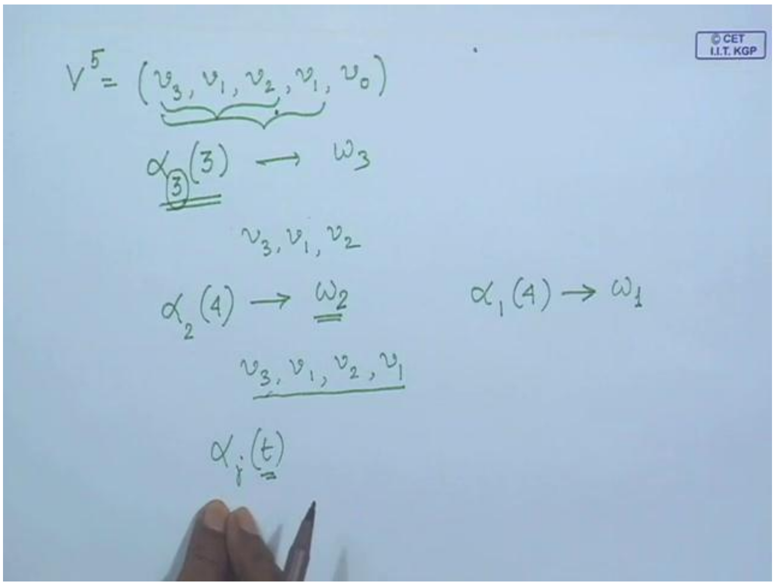
So, for doing that, what we have done is we have defined a term alpha j t, while we have said that initially alpha j t will be equal to 0, if the model is in time step t equal to 0 and this j is not an initial state. Whereas if we are in initial state at time step t equal to 0, then this alpha j t will be equal to 1 and otherwise alpha j t will be given by this particular expression that is sum of alpha i t minus 1 into a ij times b jk v t. So, this actually tells you that that what is transition probability from all the states at the previous time step t minus 1, so that multiplied by the transition probability or the emission probability of the t th symbol from state omega j. So, that is what is alpha j t?

(Refer Slide Time: 02:14)



We had also written algorithm, which we have termed as forward algorithm to find out this value of this alpha j t and alpha j t when the machine which is the final state omega naught; that is what is the probability that the machine theta or hidden mark of theta model theta has generated the given sequence V T. So, let us likely rubber it on this concept what does this alpha j t mean.

(Refer Slide Time: 02:49)

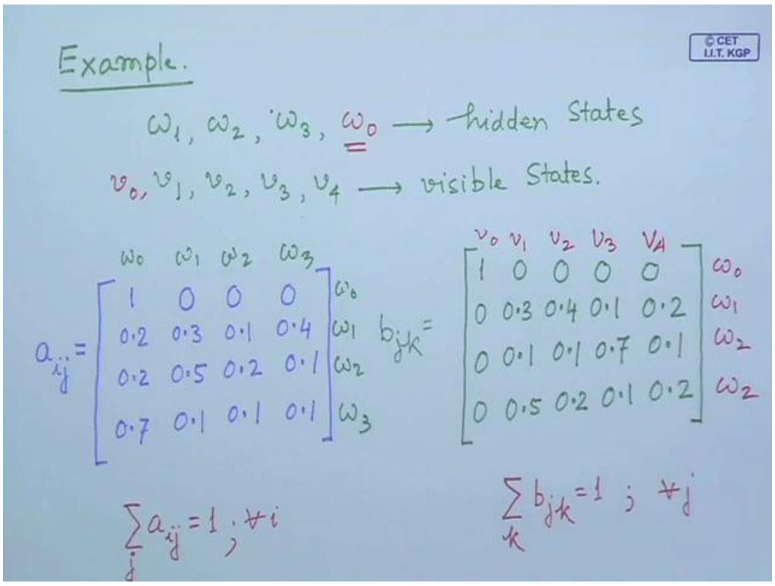


Suppose, the given V T. So, let us take that we have sequence of five symbols and this sequence of five symbols are nothing but say v 3, v 1, v 2, v 1 and the final one as we said has to be v 0. So, this is the sequence of symbols V 5 that we have. If we want to find out the term say alpha 3, sorry, alpha 3 at t equal to 3, this says what is the probability that the machine or the hidden Markov model will be in state omega 3 and it will be in state omega 3 after generation of first three symbols that is v 3, v 1, v 2.

So, this after generating first three symbols, what is the probability that the machine will be state in state omega 3? Similarly, if I say what is alpha 2, 4, this says what is the probability that the machine will be in the hidden state omega 2 after generating first four symbols? That is it has already generated v 3, v 1, v 2 and v 1. So, these are the symbols sequence of symbols, which have already been generated and after generation of these four symbols, what is the probability that the machine will be in state omega 2?

Similarly, alpha 1, 4, this will specify, this will actually indicate that after generating these four symbols, what is the probability that the machine will be in state omega 1. So, this is what this alpha j t means. So, this alpha j t means that after generating t number of symbols, what is the probability that the machine will be in state omega j. So, let us elaborate on this with the help of an example.

(Refer Slide Time: 05:31)



So, we consider an example. So, first what I have to specify is I have to specify the model theta and to specify the model theta, we have said that we have to have what are the numbers, what are the hidden states, what are the visible symbols or visible states and what are the state transition probability tables that is a ij and b jk. So, I assume that the machine has let us say four hidden states including the finals. So, I will have omega 1, omega 2, omega 3, and the final state, which is omega 0, let me write it in a different color.

So, omega 0 is the final state and in addition to this final state, it has three more states omega 1, omega 2 and omega 3 and all these are hidden states. At the same time, I also have to have that the transition probabilities for the visible states or visible symbols and for that, let me assume that there are four different visible symbols, which are v 1, v 2, v 3 and v 4. So, these are the hidden states and these are the visible states and in addition to these visible states, we also said that when the machine is in hidden state omega 0, which is the final state or absorbing state, in that state, the machine emits only one visible state. So, that particular visible state let us say that it is v 0. So, these are the visible states that I have v 0, v 1, v 2, v 3 and v 4 and we also have the transition probabilities. So, the transition probability is a ij. So, let me assume that the a ij is specified as 1, 0, 0, 0, then say 0.2, 0.3, 0.1, 0.4, 0.2, 0.5, 0.2, 0.1, 0.7, 0.1, 0.1, and 0.1. So, these are the transition probabilities among the hidden states.

Similarly, I have to have the transition probabilities for the visible states, which actually said that it is the emission probability that is b jk. So, I write b jk, which will be given by say 1, 0, 0, 0, 0, 0, 0, 0.3, 0.4, 0.1, 0.2, and then 0. So, these are the two different transition probabilities tables. The first one is the transition probability table or a ij as shown over here. This is the transition probability table for the hidden states and b jk, which is shown over here, this indicates that the transition probability table for the visible states. You find that both these tables are indexed from 0. So, this is for omega 0, omega 1, omega 2 and omega 3.

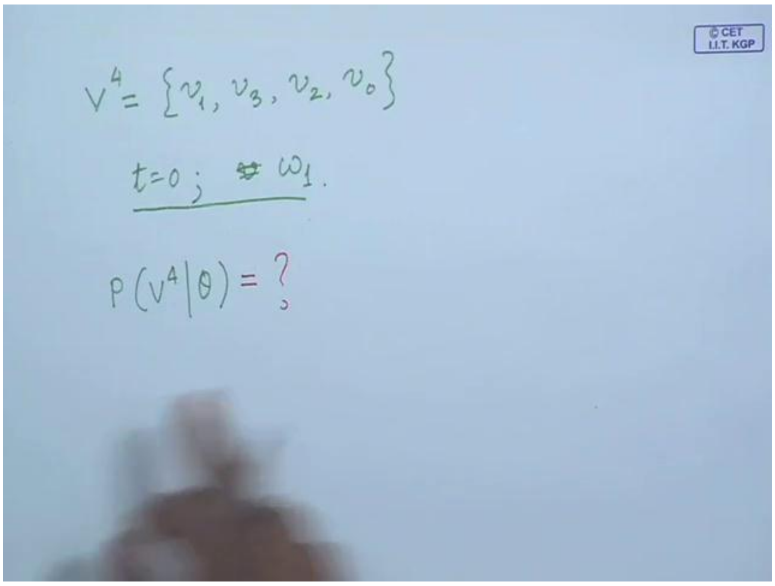
Similarly, the rows are omega 0, omega 1, omega 2 and omega 3. Similarly, for b jk, this is for v 0, v 1, v 2, v 3, v 4 and this is omega 0, omega 1, omega 2, and omega 3. So, as we said earlier that once the machine enters the final state omega 0, it remains in omega 0 only. It cannot come out of the 0. So, here you find that once it is in omega 0, a 0 0 is equal to 1, but a 0 1 is 0, a 0 2 is 0, a 0 3 is 0. That means that once the machine is in state omega 0, it will always remain in omega 0. The transition probability to omega 0 is equal to 1. Transition probability to any other state is 0, so which indicates that from omega 0, the machine cannot make a transition to any state other than omega 0. So, it will remain in the accepting state only.

Similarly, if you look at this transition table, which is actually the symbol emission table in omega 0, here it says that b 0 0 0 is equal to 1 that in state omega 0, the machine will only emit symbol v 0. It will not emit any other symbol, whereas from any other state, there is a finite probability that machine can emit any of the symbols except v 0, say omega 1 0 is equal to 0; that indicates that from state omega 1, the machine cannot emit those visible symbols v 0, v 0 will only be emitted from omega 0 and this is the only symbol that is emitted from omega 0. We had also put two more constants.

We had said that sum of a ij has to be equal to 1 when you take the summation over j. So, here you find that if you take the sum of any of the rows, the sum will be equal to 1 and this is true for all i that means for all the rows, this has to be true. Sum of the transition probabilities in a particular row is always equal to 1. Similarly, the other constants that we said that sum of b jk, when you take the summation over k that will be equal to 1 and this is true for all j.

So, here also, we find that if you take the sum of all the elements in any of the rows that is also equal to 1. So, these are the restrictions on these transition probabilities that we have to have. Now, given these transition probabilities, my problem is I want to find out that if we are given sequence of visible states, what is the probability that that sequence of visible states is actually generated by this model theta? So, for that let me take a sequence of visible states…

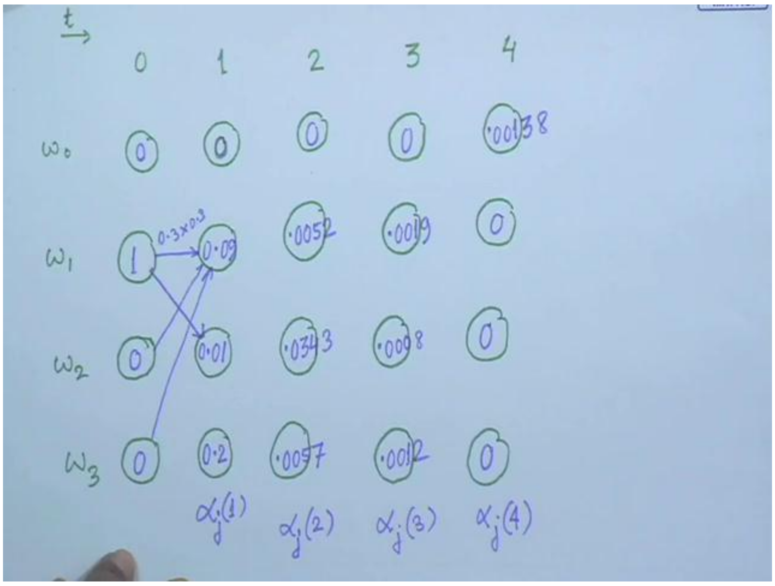
(Refer Slide Time: 14:17)



Say the sequence; we have the sequence of four visible states V 4, which is given by v 1, v 3, v 2 and v 0. As we said that we have to have an initial state, we assume that the machine is initially at t equal to 0, the machine is in one of the hidden states. So, I assume that at t equal to 0, the machine is in state omega 1. So, this is my assumption and based on this, I have to find out what is P V 4 given theta, where this theta is specified by state of hidden states, the state of visible states, the transition probability tables a ij and b jk.

So, this part is a ij. Let me just have a demarcation between these two. So, this is the ij and this is the b jk. So, these four quantities, they specify my model theta. So, what I have to find out is that given this sequence of visible states, what is the probability that that machine theta has generated this sequence of visible states? So, this is what I have to find out and that is what is my evaluation problem. So, in order to solve this problem, let us have a diagram, which is called trellis diagram.

(Refer Slide Time: 15:57)



So, so at different instants of time, I will put t in this direction at t equal to 0, t equal to 1, t equal to 2, t equal to 3 and t equal to 4. At t equal to 1, I have different states, t equal to 0, I have different states. This is omega 0, omega 1, omega 2 and omega 3. So, let me put omega 0 in the first row, omega 1 in the second row, omega 2 in the third and omega 3 in the fourth row. So, this is the situation I have and as we said that at t equal to 0, we have assumed that the machine is in state omega 1.

So, naturally alpha 1 0, alpha 1 0 has to be equal to 1 as per our definition because we have defined this alpha in that form. If you come to this definition at t equal to 0, if j is initial state, then alpha j t is equal to 1. If it is not an initial state at t equal to 0, then alpha j t is equal to 0. So, just by going, just going by this definition as I have assumed that at t equal to 0, the machine is in state omega 1. So, I have alpha 1 0 that is equal to 1, alpha 0 0 will be equal to 0, alpha 2 0 will be 0, alpha 3 0 will be 0. So, this is the initialization, initial states that I have. Now, let us see what happens at times say t equal to 2. So, here again, I have all these four different states. Now, find that coming to state omega 0, I have various possible paths. I can have a transition from omega 1 to omega 0, I can have a transition from omega 2 to omega 0, and I can have a transition from omega 3 to omega 0.

So, if I do this transition, then as we said over here that alpha 1 0 at time step zero, what is the value of alpha 1 that will be given by this expression and in this expression, I have multiplied by this transition probability. This transition probability is dictated by v 1 and what is the value of v 1 if you look at this b jk v 1, if you look at this and look at transition, the sequence of symbols that I have at times step t 1, the visible state is v 1. If you look at this, this is my hidden state 0 and the probability of emission of v 1 in state 0, in state omega 0 is equal to 0.

So, whatever path I choose, the probability that this state will exist, the existence of this state will have certain probability at time step at t equal to 1 that is equal to 0. So, as I am using this color, let me use this color. So, that is equal to 0. Then let us come to this omega 1. You find that I have various ways in which I can have a transition to state omega 1. I can have a transition to state omega 1, from omega 0, I cannot have a transition because this transition probability is equal to 0, omega 0, omega 1, and this is equal to 0. So, I cannot have a transition to state omega 1 from 0. I can only have transition to omega from omega 1 to omega 1. I can have transition from omega 2 to omega 1. I can have transition from omega 3 to omega 1.

Now, this transition probability from omega 1 to omega 1, if you look at this, that is equal to 0.3 multiplied by what is the probability that in state omega 1 at time step 1, it will output a symbol which is v 1 because my first symbol is v 1. So, the probability that from state omega 1, it outputs the symbol v 1 is equal to 0.3. So, over here, it will be 0.3 that is the transition probability from omega 1 1 to omega 1 multiplied by 0.3 again, which is the emission probability of the visible symbol v 1 from state omega 1. So, this is 0.3 into 0.3. If you come from here, this probability is 0.

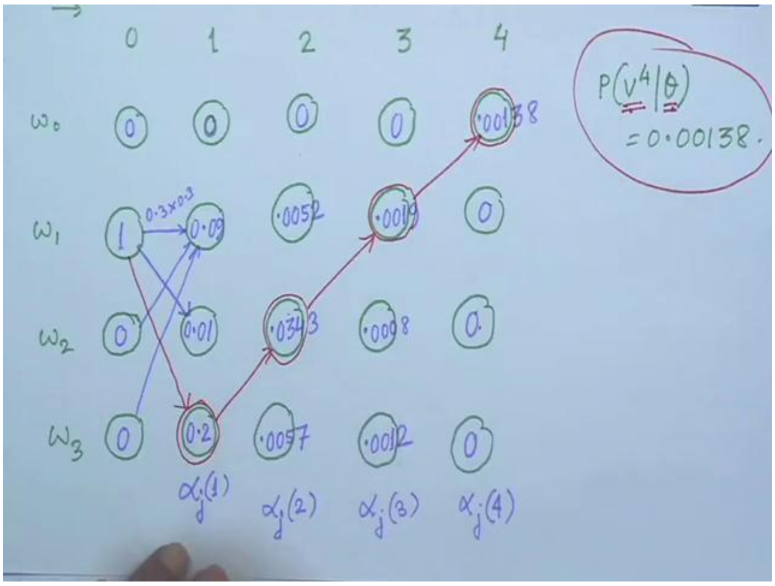
So, the contribution to this will be 0. This probability is 0. So, the contribution to this will be equal to 0. So, this alpha 1 1m the value of this will be 0.09. In the same manner, here what I have is omega 1, a 1 to that is the transition probability from omega 1 to omega 2, a 2 to a 3 2. Now, the contribution of these two to this will be equal to 0 because alpha 2 alpha 0 2 is 0, alpha 0 3 is 0. So, that contribution from this to this is 0. I will have only contribution from here to here and this again, if you look at this one, the transition probability from omega 1 to omega 2, omega 1 to omega 2 that is 0.1 and the emission probability of v 1 from omega 2, v 1 from omega 2 that is also 0.1.

So, this contribution will be 0.1 into 0.1, which is nothing but 0.01. In the same manner, I can compute that this will be 0.2. So, these figures in this circle says that this is the probability that my model will be in hidden state omega 1 at t equal to 1 after emitting the first symbol, which is v 1. If you continue like this at different time steps, so I will have omega 0, omega 1, omega 2, omega 3, time step 3, omega 0, omega 1, omega 2, omega 3. This is also omega 0, omega 1, omega 2, and omega 3. If I continue in the same manner, you will find that here also in time step 2, alpha 0 will be equal to 0, in time step 2, alpha 1 will be equal to 0.0052. You can do this computation. This is nothing but this transition probability multiplied by this multiplied by b jk V T and this case b jk is nothing but b j or b 1 and my symbol is v 3.

So, this will be b j 3 or b 1 3. So, this probability multiplied by a 1 1 multiplied by b 1 3 plus this probability 0.01 multiplied by v 2 1, sorry a 2 1 multiplied by v 1 3, similarly, from here. So, if I add all these defined terms, what I get is 0.0052. If you compute in the same manner, here it will be 0.0343; here it will be point 0.0057. So, over here, what I get is alpha 2 at time step two for all j, this is for j equal to 0, j equal to 1, j equal to 2, j equal to 3. This gives alpha 1 for all j. In the same manner, if you compute at t equal to 3, so here I have alpha j 3, for j equal to 0, I get this. For j equal to 1, I get this. For j equal to 2, I get this. For j equal to 3, I get this.

So, here again, this term will be equal to 0. This figure will be point 0.0019. This will be 0.0008 and here it will be 0.0012. Coming over here alpha j 4, if you compute, this will be 0.00138, this will be 0, this will be 0, and this will be 0. You find that now if you look at this algorithm, the forward algorithm that we had written, the final probability that the machine theta has generated this sequence V T is given by alpha 0 t that was my algorithm and here alpha 0 t is nothing but 0.00138.

(Refer Slide Time: 28:03)



So, the probability that our machine that you have considered here has generated the given V 4 that is this, this sequence for the given machine or the hidden Markov model theta, where Markov model theta is defined by these parameters is equal to 0.00138. So, this is the probability that our given model theta has generated that given sequence and here at every state, the term the figure within the circle indicates the probability that the model will be in omega 2, so here in omega 2, at times step 2, after generating v 1 and v 3 as for this sequence.

So, it has generated v 1 and v 3. After generating these first two symbols, the probability that the machine will be in step omega 2 is given by this. Similarly, after generating these two symbols v 1 and v 3, the probability that the machine will be in state omega 1 is given by this, which is 0.0052. So, here what you are getting is the probability that the machine has generated this V 4 in the model; theta has generated this V 4. This generation may be by any of the parts because when computing these probabilities at every time step, I have considered that the contribution from all the parts.

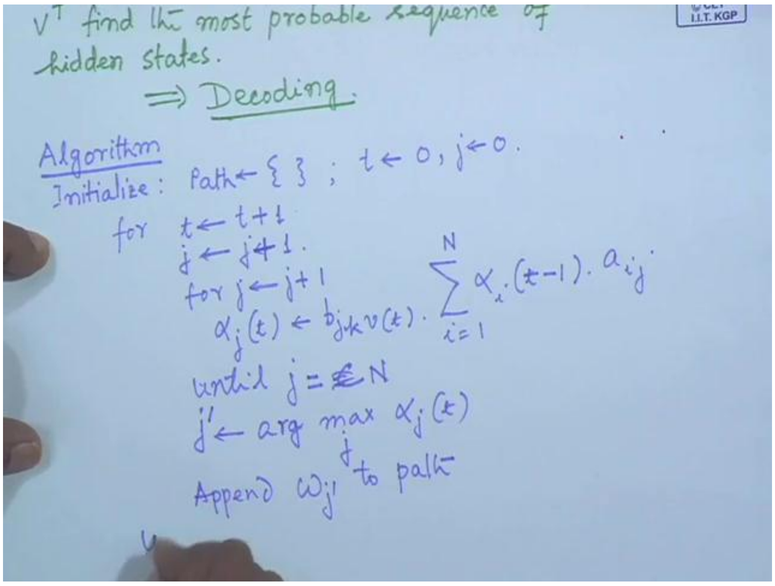
So, this is the probability that this V 4 has been generated by machine theta by this model theta and while generation of this v four, the model can make use any of this parts. So, this is what we said is what evaluation problem. The second problem that we have said is the decoding problem and what is we have said in the decoding problem is that what is the most likely sequence, or most probable sequence of hidden states through which the machine have transmitted while generating that V T.

If I try to find out this most probable sequence of hidden states through which the machine has made the transitions while generating V T, I can simply say that at every step, I can only consider that state which is most probable because you come over here, the most probable state where the machine can be, in which the machine can be at t equal to 1 is 0.2 because the probability that the machine will be in state omega 3 after generating the first symbol is 0.2, whereas the same probabilities for the other states is less than 0.2.

So, I can say that in time step one, this is the most probable state. Similarly, in time step two at t equal to 2, the most probable state in which the machine can exist is omega 2 because here the probability that the machine will be in state omega 2 after generating first two symbols is 0.0343 or less, for omega 1, it is 0.0052, for omega 3 it is 0.0057, for omega 0, it is 0. Obviously, here we cannot reach. So, this is the next most probable state after generating the first two symbols. Similarly, over here, the next probable state after generating the first three symbols is this and obviously the final state in the sequence is the absorbing state or the final state which is omega 0. So, once I have this trellis diagram, the decoding problem is very simple. I know my initial state was omega 1.

So, the sequence of states through which the machine transits while generating this sequence V 4 is equal to v 1, v 3, v 2 and v 0 is the sequence of states omega 3, omega 2, omega 1 and omega 0. So, this is the most probable sequence of states and that is what is my decoding problem? Once from the trellis diagram, I have this simple method to find out, to solve the decoding problem, I can also write an algorithm for this decoding problem and this algorithm is very simple that straight away comes from this diagram.

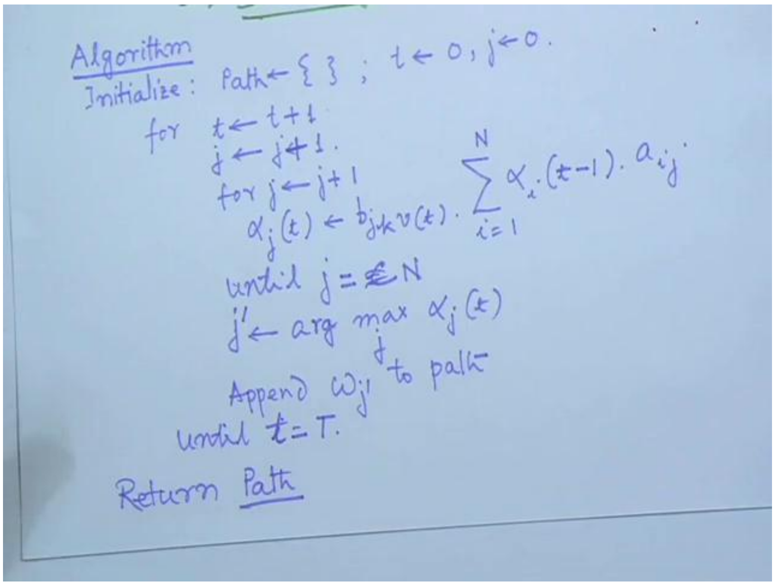
(Refer Slide Time: 33:52)



So, the decoding problem is we said that given V T; find the most probable sequence of hidden states. So, this is what my decoding problem that was the second issue in our hidden Markov model and obviously, I can write an algorithm for doing this. So, in algorithm, basically what I am trying to do is I am trying to find out the path, the most probable path, which is nothing but the sequence of the hidden states through which the model will make transitions.

So, I will have as before an initializing step. In initialization, I set path to be an empty set and I initialize t to 0. Then I will have a number of iterations. So, I set go for iterations with increment in t and increment in an index j. Then, for j to j plus 1, I put alpha j t will get b j k v t into sum of alpha i t minus 1 times a ij, i going from 1 to N as N is the number of hidden states that I have and this will continue until j becomes equal to c let us say the number of not c, j is equal to N where N is the number of hidden states. Once this condition is ditched, then what I have to do is I have to put j at which is nothing but I have to find out the state, which is having the maximum probability. So, this argument maximum alpha j t or this maximum has to be computed over j.

(Refer Slide Time: 37:29)

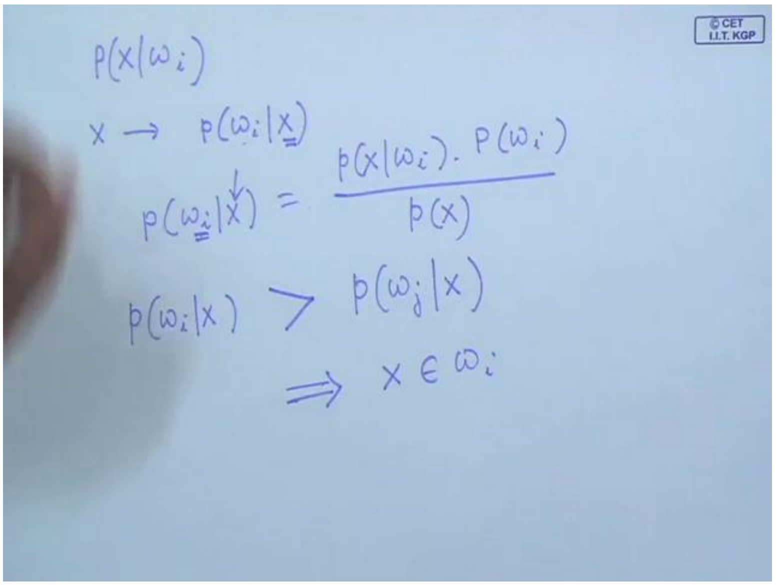


Then, append omega j dash to path and you have to repeat this until t becomes equal to capital T, which is the length of the sequence and at the end of the algorithm, what we have to do is we have to return the path. So, if you run this algorithm for this given problem, we will find that we will come out with this path only. So, this is what is known as the decoding problem that is I find out the most probable sequence of hidden states through which the machine makes a transition while generating the sequence of the visible symbols v t.

At the end, what we get is what is the probability that the given sequence of the symbols have been generated by the given machine theta, but theta is actually specified by the state of states, the state of visible states and the state of hidden states along with two transition probability tables, probability transition tables. One is the transition probability from the hidden state to the hidden state and the other one is the transition probability from the hidden state to the visible state.

This is not really a transition. What we can call is that this is probability of emission of with visible states from different hidden states. So, these four quantities define my model theta and given this model theta and given sequence of visible symbols, I want to find out what is the probability that the model theta has generated this sequence of visible symbols v t. Now, my final goal is I want to classify this sequence of these symbols. So, now you can recollect that when you talked about the bays' rule, we have said that we have class conditional probability functions…

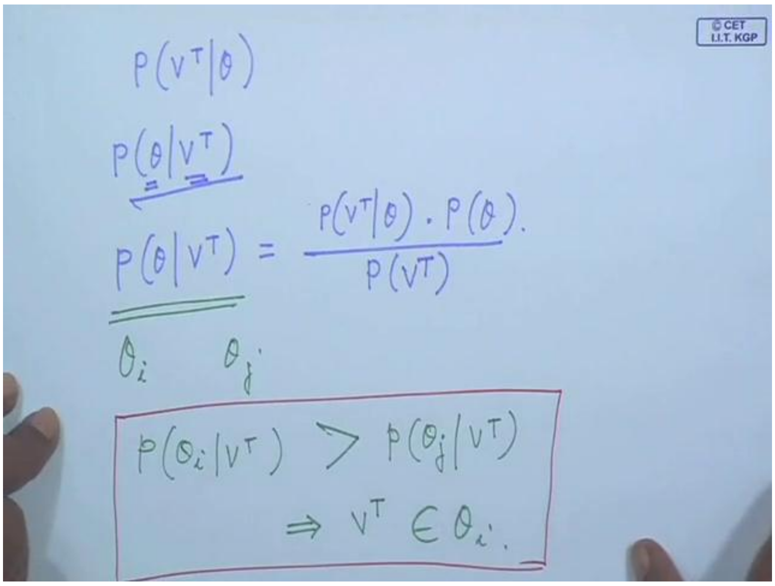
(Refer Slide Time: 40:09)



That is what we had is something like P of X given omega i. This was class conditionally, conditional probability function. What can be estimated now from as unknown sample say X to classify this in one of the classes, what we needed to compute is P of omega i given X and that particular i for which P of omega i given X is maximum, this unknown sample was classified to that particular class. We have said in that case, that what base theory says is, it computes this is this P of omega i given X is called a posterior probability, P of P of omega i given X is called posterior probability. And P of X given omega i is given, is called a priory probability is the class conditional probability and along with that, we can have a priory probability, what is the probability of the occurrence of the particular class.

We can combine this class conditional probability density function with the priory probability by using the bays' rule, bays' theory which states that P of omega i given X is nothing but P of X given omega i into priory probability P of omega i upon P of X. So, if I have two classes and the same unknown sample X, I will compute P of omega i given X. I will also compute for another class omega j, P of omega j given X. So, these are two posterior probabilities. If P of omega i given X is greater than P of omega j given X, then we conclude that X belongs to class omega i. So, this is what we have done with bays' classification.

(Refer Slide Time: 42:22)



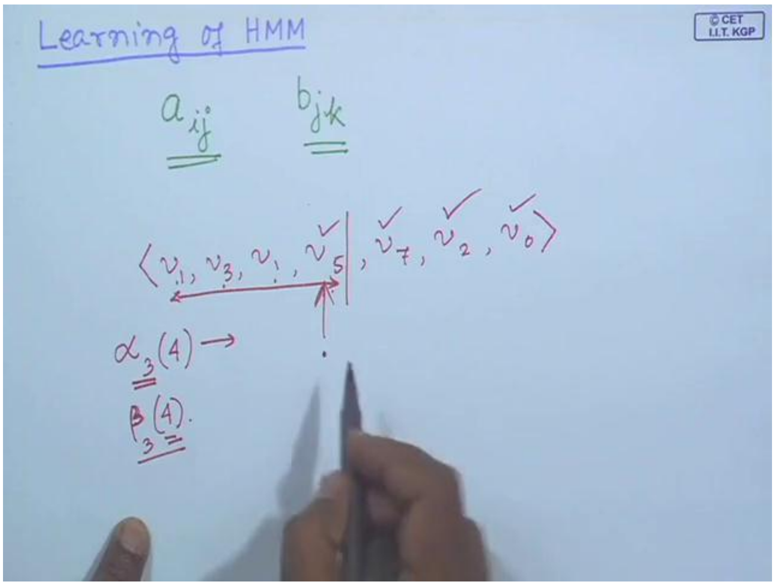
In the same manner, in this particular example or in this particular case of hidden Markov model with the sequence of visible symbols, what we are finding out is P of V T given theta that is what is the probability that V T has been generated by theta, but what we need for classification purpose is P of theta given V T. That is given this visible sequence, what is the probability that this sequence belongs to a class theta.

So, here again, we can apply this bays' rule. So, this bays' rule can simply be applied as P of theta given V T to be taken as P of V T given theta which is nothing but this quantity multiplied by P of theta, which is the priory probability upon probability of this sequence V T irrespective of the models. Once we have this quantity P of theta given V T, then I can use this quantity for classification. So, if I have two models, one model is theta I, the other model is theta j. I can compute this term for theta i, I can compute this term for theta j, I can compute this P theta i, P theta j which are the priory probabilities. Then I have this posteriori probability P of theta i given V T; I also have a posteriori probability P of theta j given V T. If P of theta i given V T is greater than P of theta j given V T, theta j given V T, then obviously my interpretation will be that V T belongs to class theta i.

So, this is how I can classify a sequence, a time sequence of visible symbols. So, this will be my classification. So, out of the three major issues that we have said that three central issues in a hidden Markov model, what are the issues we have said? We have said that first issue is the evaluation, where we try to compute what is a probability that a model has generated in a given visible sequence. The second issue was decoding issue where we have found out, what is the most probable sequence of hidden states through which, the model has made transitions to generate the given sequence of visible symbols.

The third central issue, which is a very important issue, is learning of the hidden Markov model, training of the hidden Markov model. Again, as we discussed before, when we talked about the supervised learning and unsupervised learning while when you try to train the hidden Markov model, you make use of a number of sequences of visible symbols and you already know to which, what is that visible symbol. As using a number of known visible sequences, you try to train the hidden Markov model. So, the training or learning process of hidden Markov model is actually a supervised learning.

(Refer Slide Time: 46:31)



So, now what we try to discuss is learning of hidden Markov model, which as we have said it is very important issue. So, before learning what we said is the hidden Markov model is coarsely specified that is I have a very coarse representation of the hidden Markov model in terms of I know what is the number of hidden states of the hidden Markov model. I know what are the visible symbols, which are generated by the hidden Markov model?

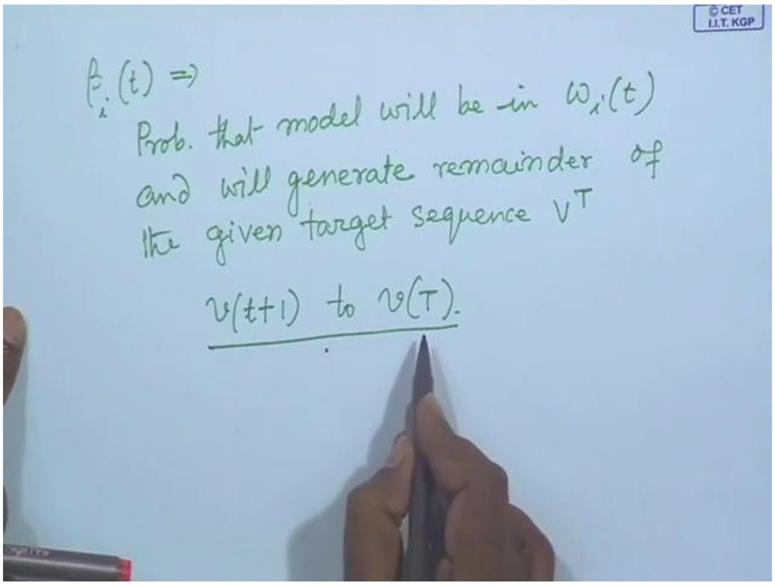
So, these are the two quantities, which are known that is the hidden states and the visible states or visible symbols. So, by learning of hidden Markov model or by training a hidden Markov model, what we usually mean is that we have to estimate the transition probabilities a ij and b jk. So, I know what are the hidden states? I know, what are the visible states? We have to estimate is the transition probabilities a ij and b jk.

So, these are the two parameters of the hidden Markov model, which needs to be estimated through the learning process. For the estimation of this a ij and b jk, as we said this is the supervised learning or supervised training process. So, we make use of us number of known sequences, the sequences for which we known that to which of the class those sequences belong. So, using this, we go for the training of the hidden Markov model.

Now, you remember when I talked about this forward algorithm, this one, that time we said for training of the Markov model or for learning, we use a similar algorithm, which is the backward algorithm. So, in case of forward algorithm, we have said that this alpha j t, this actually tells what is the probability that the model will be in state omega j at time state t after generating first t number of visible symbols in the given sequence of visible symbols v t, When I go for backward algorithm, the backward algorithm says that what is the probability that the machine or the model will be in state omega j at time instant t and will generate the remaining part of the sequence of visible symbols? So, what I mean to say is that if I have a state of visible symbols, which are given like this v 1, v 3, v 1, v 5, v 7, v 2, v 0, suppose this is the sequence of visible symbols.

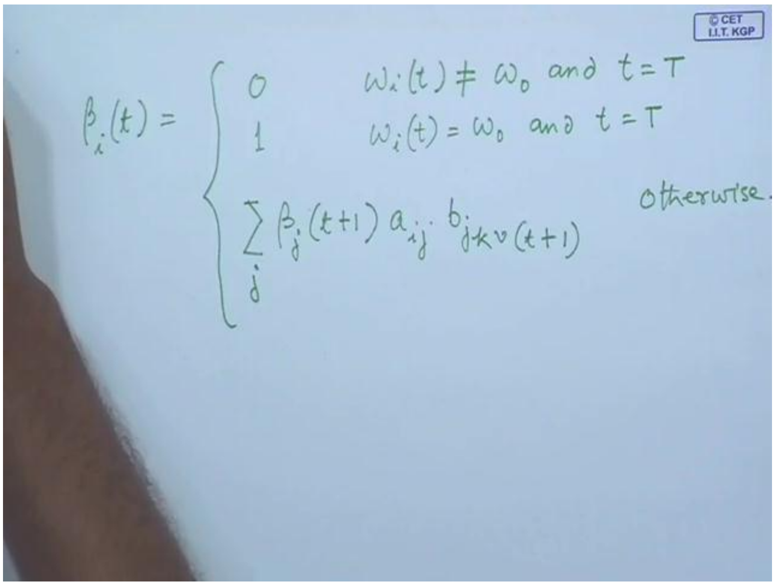
So, alpha say 3 4, this gives me what is the probability that the machine will be in state omega 3 after generating v 1, v 3, v 1 and v 4 up to this. So, the machine has generated these symbols in the given sequence and then what is the probability that it will be in the hidden state alpha 3 in the hidden state omega 3, whereas if I write say beta 3 4, what it will say is what is the probability that the machine will be in state omega 3. It will generate the remaining four symbols of the given sequence that means it will generate v 7, it will generate v 2, it will generate v 1 and of course, as I said 4, so v 5. So, what is the probability that in this time, the machine will be in state omega 3 and it will generate v 7, v 2, v 1.

(Refer Slide Time: 51:44)



So, in other words what I can say is that this beta j or say beta i t, this actually represents the probability that model. So, this beta i t gives you what is the probability that the model will be in state of omega i in time state t and it will generate remainder of the given target V T that means it will generate all the visible symbols v t plus 1 to v T. So, all the symbols from v t plus 1 to v T that will be generated and the machine is in state omega i at time state t. So, that is what is beta i t. So, in the forward algorithm, we find out that what is the probability machine is in the state omega i after generating first t number of symbols and beta i t in the backward algorithm that tells you what is the probability that the machine is in state of omega i and it will generate the remaining part of the symbols of the target sequence. So, accordingly I can write the definition beta i t.

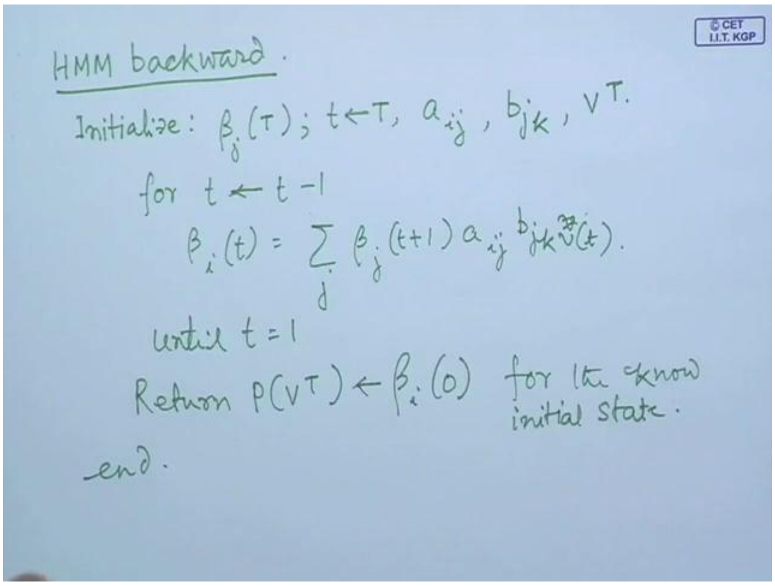
(Refer Slide Time: 53:48)



I can define this beta i t. So, this beta i t will be equal to 0 if omega it is not equal to omega 0 and t is equal to the last symbol in the last time state that is t is equal to t because as we said that the last state hidden state has to be omega 0 in which the machine generates the only visible symbol, which has to be 0. So, if this condition is not true, then beta i t will be equal to 0. Beta i t will be equal to 1 if omega it is the final state or absorbing state of omega naught and t is equal to capital T. In all other cases, this beta i t will be beta j t plus 1 into a ij into b j k v t plus 1 otherwise. The summation has to be taken over all j.

So, this is the definition of beta j t. Now, if you look at the same diagram that we have used earlier, we can explain that part actually means. Coming over here, in case of forward algorithm, we have taken all possible transitions from the states in time say t minus 1. In case of backward algorithm, in the other, on the other hand, what we will do is for every step in for every state in step t, I will find that what will be possible transitions to different states in step t plus 1 because all these transitions are supporting the existence of this particular probability of existence of this particular state. So, that is what is done in the backward algorithm and for that, this beta i t has been defined.

(Refer Slide Time: 56:26)



So, as we defined this, we can also write an algorithm, this backward algorithm which is, which we will term as HMM backward. So, earlier we have written HMM forward algorithm. Now, I write HMM backward. So, here again we have to initialize, we have to have an initialization state says for I out beta j t, t T, a ij b jk and V T. So, all these are initialized. Then we have to have for loop for t to t minus 1, beta i t will be simply that expression beta j t plus 1, a ij, b jk v t, take the summation over j. This has to be as we are moving backwards, so I have to move from capital T to 1. So, until t is equal to 1 and at the end, you return P V T, which is beta i 0 for the known initial state and end of the algorithm.

So, we can estimate the probability that the machine will be in state say omega i or omega j at a time step t after generating the first t number of sequence or it will be in the same state, same state at time step t and will generate the remaining symbols from the given sequence by this backward algorithm. So, I will stop here today. In the next class, I will use both this forward algorithm and the backward algorithm for estimation of the model parameters a ij and b jk.

Thank you.