Course Name: Machine Learning and Deep learning - Fundamentals and Applications

Professor Name: Prof. M. K. Bhuyan

Department Name: Electronics and Electrical Engineering

Institute Name: Indian Institute of Technology, Guwahati

Week-5

Lecture-23

Welcome to NPTEL online course on machine learning and deep learning fundamentals and applications. In this week, I explained the principles of logistic regression and decision trees. Today, I am going to explain the concept of one important machine learning model that is the hidden Markov model. Hidden Markov model is a graphical model, which can be used to predict a sequence of unknown variables from a set of observed variables. So for this, I will be considering Markov's scene property. So that I will be explaining today.

So let us discuss about the model. The model is the hidden Markov model. So first I will explain what is the Markov model. So for Markov model, we are considering a set of states.

So states are S_1, S_2. So we are considering n number of states. And after this another consideration, the process moves from one state to another generating a sequence of states. So what actually we are considering. So the process moves from one state to another.

And because of this moment, I am getting a sequence of states. So this sequence of states, I can write like this. So S_1, S_2. So these are the states Sk. So this is another consideration.

And we have to consider the Markov chain property. So what is the meaning of this? This mathematically I can show like this, $P(S_k|S_1, S_2, \ldots, S_{k-1}) = P(S_k|S_{k-1})$. That means the probability of each subsequent state depends only on what was the previous state. So you can see the probability of Sk, that is state is Sk, it depends on only S_1, S_2 up to S_{k-1}.

These are the previous states. So mathematically I can write like this, the probability of S_k, that means the state S_k depends on the previous state. The previous state is S_{k-1}. So that is the Markov scene property.

And to define the Markov model, we are considering following probabilities.

So to define the Markov model, we are considering some probabilities. So one probability that is very important, that is called the transition probability, that is $a_{ij} = P(S_i|S_j)$. That means in this case what we are considering the transition from one state to another and corresponding to this we are considering one probability and that probability is the transition probability a_{ij}. So probability of S_i given S_j, that is the transition probability. And also because in this Markov model we are considering number of states.

So that is why we are associating some probabilities with the states and that is nothing but the initial probabilities. So initial probabilities of the states, initial probabilities that is defined by π_i and that is nothing but the probability of a particular state. Because in the case of the Markov model, we are considering number of states and also we are considering the transition from one state to another state. So that is the concept of the Markov model. So we have number of states like S_1, S_2, S_n and we are considering the Markov scene property that already I have explained that is the probability of each subsequent state depends only on what was the previous state that is the Markov scene property.

And after this we consider some probabilities one is the transition probability. So transition from one state to another that is the probability of S_i given S_j that is the transition probability from S_i to S_j and also we are considering the initial probabilities associated with each and every states of the model. So that is the π_i. So that is the initial probabilities. So corresponding to this concept I can give one example of the Markov model.

So let us move to the next slide. So I can give one example suppose I am considering a state this is suppose rain and another state we are considering suppose dry. So in this example I am considering two states and now I am showing the transition from one state to another. So this is the transition from rain to dry and this is the transition from dry to rain and also we can consider self transition. So that means the self transition is like this.

So self transition also we are considering and we are considering the probabilities. So probability of transition from the state rain to the dry suppose I am considering 0.7 the probability of transition from dry to rain suppose we are considering this probability as 0.2 and the self transition the probability is 0.3 suppose and this self transition the probability is 0.7.

So in this model we are considering two states. So two states we are considering and these two states are nothing but the rain and dry and we are considering transition probabilities. So what are the transition probabilities? So suppose the probability of rain given rain that is actually the self transition. So that probability is 0.3 that is I have shown in the figure.

So another probability is the probability of dry given rain. So what is the probability you can see from the figure. So it is 0.7 and another probability the probability of rain given dry that means the probability of obtaining rain given dry condition is 0.2 and also another probability I have to define the probability of dry given dry that is nothing but the self transition.

So it is 0.8 sorry this should be 0.8, we are considering this self transition probability as 0.8 corresponding to dry and the self transition probability corresponding to rain is 0.3 and you can see the probability of dry given rain is 0.7 and probability of rain given dry is 0.2.

So these are the transition probabilities we are considering and also the initial probabilities also we are considering Initial probabilities. So what are the initial probabilities suppose we can consider suppose say the probability of rain is equal to 0.4 and the probability of dry that is 0.6. So this initial probabilities also I am assuming probability of rain is 0.4 and probability of dry is 0.6. So this is one example of the Markov model. So now how to calculate sequence probabilities. So let us discuss about how to calculate sequence probabilities.

So move to the next slide there is the calculation of sequence probability. So for this we are considering the Markov's sin property. So probability of a state sequence can be found by the following formula. So suppose I want to determine the probability of a state the sequence is S_1 S_2 S_k so this sequence we are considering so that you can determine like this probability of S_k depends on the previous states. So previous state is S_1 S_2 up to S_{k-1} so $P(S_1, S_2, \dots, S_k) = P(S_k | S_1, S_2, .., S_{k-1}) P(S_1, S_2, .., S_{k-1})$

So this is the joint probability. So this I can write like this $P(S_k | S_{k-1})$ so this is the Markov's sin property. So $P(S_1, S_2, .., S_{k-1})$ so like this we can consider and finally you can say this $P(S_k | S_{k-1}) P(S_{k-1} | S_{k-2})$. So we are applying the same rule $P(S_k | S_{k-1}) P(S_{k-1} | S_{k-2})$. So this is the same rule we are applying corresponding to the Markov model till S_2 $P(S_2 | S_1)$ and finally $P(S_1)$.

$$P(S_1, S_2, \dots, S_k) = P(S_k | S_{k-1}) P(S_{k-1} | S_{k-2}) \dots P(S_2 | S_1) P(S_1)$$

So we are applying the Markov's sin property. So corresponding to this expansion I am showing the calculation of the sequence probability. So if I considered a previous example suppose I want to calculate the probability of a sequence of states of the previous example. So I want to determine or I want to calculate a probability of a sequence of states in my previous example. So suppose the states are like this dry rain and rain.

So corresponding to this sequence I want to determine the probability. So what is the probability in this case the probability I can determine the probability $P(\{Dry, Dry, Rain, Rain\})$. So this sequence we are considering we are determining the probability and that is actually equal to if I apply the same rule $P(Rain|Rain)P(Rain|Dry)P(Dry|Dry)P(Dry)$. So we can expand like this so we can determine the probability of the sequence the sequence is $P(\{Dry, Dry, Rain, Rain\})$. So how to get this probability so by considering this Markov's sin property I can do this expansion and based on this formula I can determine $P(\{Dry, Dry, Rain, Rain\})$.

So already I have defined this probability so it is $P(Rain|Rain) = 0.3$ into $P(Rain|Dry) = 0.2$ the $P(Dry|Dry) = 0.8$ and $P(Dry) = 0.6$. So this will be approximately equal to 0.288 so this probability I can determine so that means the probability of obtaining this particular sequence so we can obtain this probability. So this is the fundamental concept of the Markov model so now I will discuss this concept in more detail so what is the Markov model. So let us move to the next slide So already I told you in the Hidden Markov model I have number of states so that means set of states so these sets are like this S_1, S_2 so sequence of states like this and the process moves from one state to another generating a sequence of states. So process moves from one state to another generating a sequence of states the sequence of states through which the model passes are hidden and they cannot be observed so that is why it is called hidden.

So I am repeating this the sequence of states through which the model passes are hidden and it cannot be observed. So that is why we are considering the term the hidden the third point I considered the Markov chain property so that already I have explained so the probability of S_k it depends on the previous states so previous state is S_1 S_2 up to S_{k-1} so I can write that probability of the state S_k depends on the previous state that is the Markov property. And already I told you the states are not visible states are not visible so that is why the term is the Hidden. But each state randomly generates one of M observations or I can say visible states so visible states are suppose V_1, V_2 up to V_m. So here you can understand what are the main concept of the Hidden Markov model so the first point is I have number of states that is S_1, S_2 up to S_n

so process moves from one state to another generating a sequence of states.

So that means the sequence of states I can write S_1, S_2, S_k so like this I am getting the sequence of states and after this the next point is we are considering the Markov chain property so the state S_k depends on the previous state S_{k-1}. So probability of S_k depends on the previous state that is the S_{k-1} so mathematically $P(S_k|S_{k-1})$ so that we are considering. And the states are not visible so that is why the term is the Hidden but each state generate randomly one of M observation or the visible states so I am getting the observation symbols the symbols are V_1, V_2 up to V_m. So these are the components of the Heiden Markov model so now mathematically how to define the Hidden Markov model let us move to the next slide. So how to define mathematically the Hidden Markov model HMM.

So already I told you that we have to consider some probabilities so we have to define some probabilities. So what are the probabilities I have to define? So for defining the probabilities we are considering the matrix of transition probabilities. Because there is a transition from one state to another so we have to consider transition probabilities. So we are considering the matrix, $A = [a_{ij}]$ and actually the $a_{ij} = P(S_i|S_j)$.

So transition from the state S_i to S_j so that we are considering that transition and corresponding to that transition we are considering a probability.

So I am getting a matrix and that is the matrix of transition probabilities. After this the next important point is we are considering another matrix of observation probabilities. So for this we are considering the matrix the matrix is B and b_i the observation symbols is V_m the observation symbol I am considering as V_m. So $B = [b_i(V_m)]$ so where $b_i(V_m) = P(V_m|S_i)$. The state is not visible but the observation symbols are visible.

So that is the probability of V_m given S_i that probability we are considering and that is the observation probabilities. The second matrix we are considering to consider the observation probabilities. The first matrix I consider for considering the transition probabilities and after this I am considering a vector, a vector we are considering for initial probabilities the vector of initial probabilities. So this π is a vector so elements are π_i so where this $\pi = \pi_i$ so $\pi_i = P(S_i)$.

So these probabilities we are defining one is the transition probability one is the observation probabilities and also we are considering the initial probabilities.

So based on this I can define the Hidden Markov model so my model the HMM model I can define like this the model is M and the parameters are A, A means the transition probabilities we are considering B is the observation probabilities and π is the initial probabilities. So these are the components of the model $M = (A, B, \pi)$. So I can represent a particular Hidden Markov model based on these parameters A B π. So, corresponding to this concept let us move to the previous example that I considered earlier. So we consider this Hidden Markov model so let us move to the next slide.

The example of HMM so based on my previous example I am showing the Hidden Markov model. So one state already I have shown the states are rain and the dry and suppose I am considering this low state the high state and outcome is rain and another outcome is the dry. We have considered self transition so what is the probability of this so probability is suppose 0.3 this self transition probability high suppose 0.8 and transition from low to high that transition probability also we considered so it is 0.7 suppose and transition from high to low we considered suppose so we are considering as 0.2 the transition. So after this the outcome is whether the prediction of rain or dry so from this we can do some predictions we can do predictions and the probability suppose 0.6 this probability is 0.6 and this probability is 0.4 and corresponding to dry also we can do the prediction. So this probability is 0.6 and suppose this probability is 0.4 so we can do the prediction and I can get the information about the rain whether rain will be there or the dryness so this we are considering.

So these two variables we are considering low and high and based on all these probabilities the transition from low to high high to low we can do the prediction of rain or dryness.

So this is one example of the Hidden Markov model. In this case I have two hidden states that is the low and high that corresponds to atmospheric pressure. So I am writing this important points here so I have two hidden states two hidden states so what are the states one is the low another one is the high and these are atmospheric pressure. So these are two hidden states and what are the observation states so observation states so again I have two observation states rain and dry. So now we are defining the transition probabilities so what are the transition probabilities $P(Low|Low) = 0.3$, $P(High|Low) = 0.7$. So another probability is $P(High|High) = 0.8$ so this is the self transition and another probability is $P(Low|High) = 0.2$. So these are actually these probabilities are the transition probabilities. So these are transition probabilities also we are considering the observation probabilities so what are my observation probabilities $P(Rain|Low)$ the atmospheric pressure is low so what is the probability of obtaining rain given low is the atmospheric pressure so that is 0.6, $P(Dry|Low) = 0.4$, $P(Rain|High) = 0.4$ and $P(Dry|High) = 0.6$. So these are observation probabilities we are also considering initial probability so initial probability we are also considering so we are also considering initial probabilities.

So what are the initial probabilities probability of low suppose 0.4 and probability of high is suppose 0.6 so these initial probabilities also we are considering. So in this example you can see we are considering the transition probabilities observation probabilities and also we are considering the initial probabilities suppose based on this problem I want to determine the probability of a sequence of observation and suppose I want to determine the sequence {Dry, Rain}. So if I consider the probability of obtaining first dry and after this the rain so that sequence probability I can determine.

So move to the next slide calculation of sequence probability. Suppose in this example I want to determine the probability of the sequence of observation the sequence is suppose Dry and Rain so this probability I want to determine. So we have to consider all possible hidden state sequence we have to consider so how to determine this probability of Dry and the Rain this sequence we are considering. So all possible hidden state sequence we have to consider so $P(\{Dry, Rain\}, \{Low, Low\}) + P(\{Dry, Rain\}, \{Low, High\}) + P(\{Dry, Rain\}, \{High, Low\}) + P(\{Dry, Rain\}, \{High, High\})$

So like this we can expand so if I consider the first term so this first term if I consider this first term I can write like this $P(\{Dry, Rain\}|\{Low, Low\})$, the low atmospheric pressure and the low atmospheric pressure. So if I consider this the first term the first term I can determine $P(\{Dry, Rain\}|\{Low, Low\})P(\{Low, Low\})$ so this is equal to $P(Dry|Low)P(Rain|Low)P(Low)P(Low|Low)$, $P(Dry|Low) = 0.4$, if you see the previous slide into $P(Rain|Low) = 0.6$ if you see the previous slide and what is the probability of low, $P(Low) = 0.4$ that is the initial probability and $P(Low|Low) = 0.3$.

So ultimately this probability of the sequence will be 0.0192 so that is the $P(\{Dry, Rain\})$ that probability I can determine that is the $P(\{Dry, Rain\}|\{Low, Low\})$. So the first term I can determine like this and the rest of the terms also I can determine like this so this is the calculation for the first term only so rest of the terms also I can determine like this. Now briefly I will explain different types of hidden Markov model structures. So let us move to the next slide so the types of types of the hidden Markov model structures so one model is number 1 ergodic model. So in this case I can show it pictorially so I can show it pictorially suppose I have a state 1 another state is 2 another state is 3 another state is 4.

So four states considering and I have to show the transition between the states so from 1 to 2 there is a transition from 2 to 1 there is a transition from 1 to 3 there is a transition from 3 to 1 there is a transition and from 3 to 4 there is a transition from 4 to 3 there is a transition from 2 to 4 there is a transition from to 2 there is a transition and similarly if I want to consider the transition from 2 to 3 this is the transition from 3 to 2 also I can consider transition from 1 to 4 also I can consider transition and from 4 to 1 also I can consider transition. So, all the transitions I am considering and also I have to consider the self transition also I can consider it is a self transition. So, the self transitions also we are considering. So, I am showing all the possible connections. So, in the agro-dig model every aij that is the transition probability is positive and every transition is possible every transition is possible.

So, this is about the ergodic model. So, another one is the bakis model another model I can consider that is the bakis or it is also called a left to right model. So, I can show it pictorially. So, suppose I have number of states 1 2 3 4. So, these are the states. So, I am showing the transition from the state 1 to 2 from 2 to 3 from 3 to 4.

So, these are the transitions. So, there may be transition from 2 to 4 1 to 3 or maybe 1 to 4. So, all these transitions I am showing and also I have to consider self transitions. So, this is the self transition corresponding to state 1 this is the self transition corresponding to the state 3 self transition corresponding to the state 4. So, in the bakis model this aij is equal to 0 that is j less than i. So, what is the meaning of this the transition probability is 0 corresponding to j less than i that means the meaning is cannot go backwards.

So, only we are showing the forward transitions. So, backward transitions are not possible. So, we are only showing the forward transitions and backward transitions are not possible. So, generally this texture is good for temporal structure.

So, maybe we can consider speech signal. So, for temporal pattern recognition. So, this model we can use. So, for example, in the speech recognition, and this is a popular model that is the bakis model. So, for recognizing temporal patterns, we can consider this model. So, now let us discuss about the main issues in the hidden Markov models. So, in my next slide, I will be explaining what are the issues in the hidden Markov model.

So, the main issues in hidden Markov model. So, these are very important issues. So, what we need to consider the first point is number 1 the evaluation problem. So, what is this problem already I have explained about the model of the hidden Markov model. So, already I have explained the model corresponding to HMM. So, the model is represented by like this the model is M, $M = (A, B, \pi)$.

So, these are the components of the model A, B, π. So, the model is given. So, given the model. So, the model is given the model is A, B, π and what are the things that given the model and the observation sequence. So, that is also available the model is also available the observation sequence is available. So, observation sequence is suppose $O = O_1, O_2, \ldots, O_K$. So, corresponding to this case, I have to calculate the probability that the model M has generated the sequence O.

So, I can write calculate or I have to determine the probability calculate the probability that model M has generated the sequence. So, generated the sequence O. So, this is the first problem that is the evaluation problem. So, the model is given the model is M and already I told you I have 3 components A, B, π of the hidden Markov model and observation sequence is also given that observation sequence is O the sequence is O_1, O_2 upto O_K.

So, we have to determine the probability that the model M has generated the sequence O. So, this problem can be considered by one important algorithm. The name of this algorithm is the forward backward algorithm forward backward algorithm. This is the first problem the evaluation problem. So, let us discuss about the second problem.

So, what is the second problem number 2 problem and that is called a decoding problem. So, in this case also the model is given given the model model is the hidden Markov model a b pi this model is given and observation sequence is also given observation sequence is available. So, sequence is O_1, O_2 up to O_K. So, what I need to calculate the most likely sequence of hidden state the hidden state is represented by S_i that produced produce these observations observation sequence O. So, that means in this case what is the second problem the second problem is the decoding problem.

So, the model is given and observation sequence is also given. So, we have to determine the most likely sequence of hidden state the that is S_i that produce this observation sequence this is a very important problem. And this problem can be considered by one algorithm there is a very popular algorithm and that is called the Viterbi algorithm. So, Viterbi algorithm the concept of the forward backward algorithm and the Viterbi algorithm I will not be able to explain in this course because of the time constraints of the 12 weeks course this is a 12 weeks course and because of the time constraints I will not be able to explain these 2 important concept one is the forward backward algorithm and another one is the Viterbi algorithm. So, this concept you can read from some research papers.

So, one paper by Rabiner the hidden Markov model by Rabiner. So, that paper also you can see to understand these 2 important concepts one is the forward backward algorithm and another one is the Viterbi algorithm. So, the third problem is the learning problem. So, let us move to the next slide learning problem. So, that is the third problem.

So, what is this problem? Given some training observation sequence. So, given some training observation sequences. So, suppose the observation sequence $O = O_1, O_2, ..., O_K$. So, these are the observations and general structure of the HMM. So, what is available in the structure of the HMM? So, maybe we can consider number of hidden the number of hidden and the visible states number of hidden and visible states number of hidden and the visible states.

So, that is the structure of the HMM. So, we have to determine HMM parameters. So, that means I have to determine the model. So, the parameters I have to determine the parameters are a b and pi we have to determine that base fit training data. So, we consider the observation sequence O.

So, observations are O_1, O_2 up to O_K. So, this is nothing, but the sequence of observations. So, this O_K the observation symbols are V_1, V_2 these are the observation symbols. So, the third problem is quite interesting because in this problem what is given some training observation sequences are given. So, training observation sequences are given and also I have to give the general structure of the hidden Markov model that is the number of hidden state and number of visible states that I have to give. So, general structure of the HMM I have to give and also I have to give the number of training samples that is the training observation sequence I have to give and corresponding to this I have to determine the HMM parameters.

That means I have to determine the HMM model is represented by A, B, π. So, that parameters these three parameters I have to determine that base fit the training data because I have the training data. So, that is why it is a learning problem that is the training of the HMM. So, this problem can be considered by one popular algorithm and this algorithm is the Bamwels algorithm. So, we can consider this algorithm for learning problem.

So, I am considering three main issues in the HMM the first one is the evaluation problem. So, in the evaluation problem what we are considering the model is given and the observation sequence is given and we have to consider or we have to calculate the probability that the model M has generated the observation sequence O that is the evaluation problem and for this we can consider the forward backward algorithm. For decoding problem what we are considering we are considering the Viterbi algorithm. So, the model is given and the observation sequence is given and we have to calculate the most likely sequence of hidden states that produce these observations that is the decoding problem and one popular algorithm is the Viterbi algorithm we can apply for this and finally, the third problem is the learning problem. So, in this case we have the training observation sequences and we have the general structure of the HMM. So, we have to determine HMM parameters that means we have to determine the model and this is the learning problem the training of the HMM.

So, for this we can consider the popular algorithm is the Bamwels algorithm. So, already I told you because of the time constraints of this 12 weeks course I will not be able to explain the concept of the forward backward algorithm Viterbi algorithm and the Bamwels algorithm. So, you may see some research papers to understand these concepts. So, this is the basic understanding of the hidden Markov model. So, in my next slide briefly I will show one example of computer vision. So, here you can see in this case I am showing gesture recognition problem that is actually the sign language recognition problem or I can say that gesture recognition problem.

So, I have all these gestures performed by this person. So, you can see the meaning of these gestures the first one is the share the second one is the he has lost it open the door he has forgotten it listen to it throw it away. So, these are the meanings of these end gestures. So, if these gestures are performed by different users or different persons there may be some temporal variations the spatio temporal variations even the same gesture is performed by the same user in 2 different times there will be spatio temporal variations. So, that is why if I consider the hidden Markov model I have to train the hidden Markov model because there will be spatio temporal variations. And suppose corresponding to a particular gesture suppose in this case suppose if I consider this gesture you can see the number of states I can consider suppose there will be some initial state S_1 suppose another state is S_2 another state is S_3 if it moves from S_1 to S_2 if it moves from S_2 to S_3 So, that means, a particular gesture is performed the particular gesture is recognized.

So, we have to recognize a particular gesture. So, we have to determine the transition probabilities all these probabilities we have to determine that already I have explained. So, you can see we have number of states. So, we have the initial state S_1 and also we can consider the initial the probability and also the self transition also we have to consider the self transition also we are considering because you can see in a particular state the hand is waiting in a particular state for some time. So, that means, I have to consider self transition and also we have to consider the transition from one state to another. So, if I consider S_1 S_2 S_3 and all the transitions then this particular gesture the hand gesture can be recognized and we have to consider the spatio temporal variations the variation of the gestures in space and also in time.

So, already I told you if the gesture is performed by different persons there will be spatio temporal variations and even the same gesture is performed by the same person in 2 different times then also the spatio temporal variations will be there. So, that is why we have to consider the training the training of the HMM. Similarly, a corresponding to this gesture also maybe we can consider some model HMM model.

So, maybe $S_1, S_2, S_3, S_4, S_5, S_6$. So, we have to consider these transitions also. So, these transitions we have to consider and based on this state based on this HMM I can recognize this gesture the hand gesture. So, you can see the use of the HMM the application of the HMM in gesture recognition that is a very important computer vision problem. So, I am not explaining in detail, but briefly I am explaining how you can apply the hidden Markov model in recognizing gestures or recognizing sign language you can see in this example. So, this is one application.

So, similarly in a speech also you can see there are many applications of hidden Markov model. In this class I briefly explained the concept of the hidden Markov model. I explained the concept of the transition probabilities and the self transitions and based on these probabilities I have defined the hidden Markov model. So, the model is defined by 3 parameters A, B and π. So, hidden Markov model is defined by 3 parameters A, B and π.

And after this I discuss 3 important problems of the hidden Markov model. The first one is the evaluation problem. So, that can be considered by forward backward algorithm. The second problem is the decoding problem. So, that can be considered by the Viterbi algorithm. And the third one is the most important problem that is the training of the HMM.

So, that can be considered by Baum-Wehel training algorithm. So, this is about the HMM. So, briefly I explained this concept. So, let me stop here today. Thank you.