NPTEL

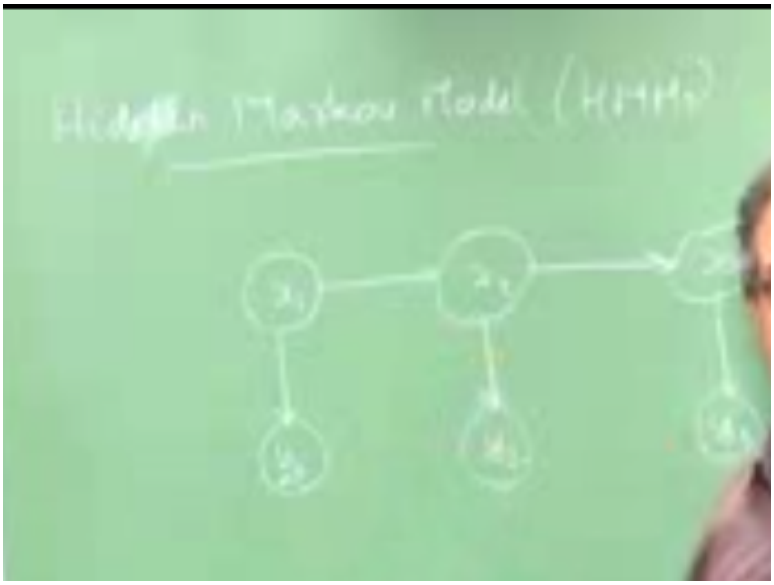
NPTEL ONLINE CERTIFICATION COURSE

Introduction to Machine Learning

Lecture-67
Hidden Markov Models

Prof: Balaraman Ravindran
Computer Science and Engineering
Indian Institute of Technology Madras

(Refer Slide Time: 00:16)



Right so people might have come across HMMs in other contexts like and you might be wondering what does he mean by saying it is a graphical model but it is a graphical model right so here I will stick with a concrete example which I am going to have a sequence right so I am going to have a sequence of random variables right I am going to assume that it is Markov in the traditional sense.

So what would that mean okay so it is left to right okay so that is the earliest point in time okay so there is no time it is a sequence right this is the first element in the sequence okay the second element in the sequence the third element in the sequence and because it is Markov so this is this

will be dependent only on the previous one right this will be dependent only on the previous one so knowing this value makes this independent of anything that came before right.

So this is essentially the graphical model version of Markov right just have a chain directed graph so it is this is sometimes called a left-to-right model it is improved after that model, so this is this is the Markov part right I am also going to assume that I have a set of observations that I make right this could be like the pixels we are talking about right so I am having labels for the images right labels for each pixel on the actual pixel value right.

So likewise in the hidden Markov model so I will have random variables which could be labels right so these are labels right and these are the entities that are being labeled so give an example of such a situation they find a media is revealed tricky because you also have the spatial dimension to it right so text is why i asked how many of your of NLP in and out of x for text right so I might want to do say part-of-speech tagging.

And I want to take each word again I want to say in apart of speech to the world right all I want to assign each word I want to take each word and I want to say that whether the word is part of a place or not part of a face right for example I would need to figure out the united states of America is actually a single thing and all the united states of America is a single phrase right so sometimes I want to capture things like that from the text right.

Is that so is that some way of doing this automatically right so people have actually come up with ways of doing there are many things there are tasks in the NLP called chunking so people have come across chunky, no, okay. So chunking essentially says that I am going to take a piece of text right and break it up into some meaningful chunks it could be noun phrases verb phrases whatever right.

So but some chunks right so that is called chunking right and there is another task called shallow parsing right so shallow parsing essentially breaks it up into phrases right but does not look at the structure of the phrases I just want to give the phrasal structure of the center so there are many tasks which people do this right so people use hidden Markov models a lot in text right and the other place where people use hidden Markov models a lot is in speech right.

Because speech is inherently a like a forward process right so the speech is inherently a forward process so people use that a lot in speech and of course you can use it in videos except that you

are individual random variable will now become something that covers the entire spatial dimensions this becomes a little more complex hidden Markov model right so great so I am going to call these after that as excess.

Well these are words right this could be words and these could be whatever label you want to send to them you can say that okay so is it part of a chunk or not part of the check what is it this is part of a name of not part of a name so I could do all kinds of labels I want on this because assumption we are making to make the model easy, okay. So yeah we can relax this assumption there are models that the more complex model seduces specifically the conditional random field that I mentioned when I was talking about Markov random piece.

It relaxes the assumption of this dependence right so this Markov assumption is still there right but this dependence is relaxed so that is typically what you are looking for I can look at any word in the sentence and then I can label the seventh word so that is that assumption is this is like right now I have to look at only the seventh point and the sixth label I cannot look at the sixth word I had to look at the sixth label and the seventh word right and that is all the information I have.

So one thing to note here is a hidden Markov model essentially says that your XI give rise to the way right does not go the other way in fact this gives this gives also gives erased or other problems later on so the exes are essentially you have labels so you do not know the labels you don't see the labels but whatever you do not see right that part of it is Markov okay and whatever you see is or the labels okay and they are just given influence only by the I am sorry.

Whatever you see are the words then they are influenced only whether labels RSA assumption you are making so very strong independence assumptions right but it turns out that it works like naive Bays works right so this also works in many situations right but then of course it also does not work in a lot of situations so people have come up with other models I just wanted to introduce you to HMMs.

Now I do not see X 1, X 2, X 3 and I only get to see y 1 y 2 y 3 and X 1 X 2 X 3 I will have to guess I mean you can also think of this as a hidden Markov random feel if you want but they do not use the terminology but hidden Markov models is something that is used quite often and there are inference techniques which people have specially honed for HMMs right but it turns

out that the same things work on many of the graphical models that you will see you will see in practice right.

So they have all kinds of things they have an algorithm called the Viterbi algorithm which essentially tells you what is the probability of the labels given the text no it does not give you the probability of the labels given the text it gives you the map estimate and it gives you the most probable label sequence given the text but then I can use it on any kind of this kind of an inference process right.

So I can use a Viterbi algorithm for any kind of a map inference process is instead of looking at the probability distribution over x and if I want to answer the query which is the most probable configuration over exercise this is the map estimated we looked at that right maximum likelihood map and then the full Bayesian inference which case me the full distribution right so far we have been talking about knowing the full distribution but you also want to do the map query right.

So that is essentially what better we will give you give it right so we are talking about estimating a potential between x and y aright so we are talking about estimating a potential between xi and yi so the question is the question is ψ (xi yi) independent of I with regardless of where in the pixel the label occurs right sorry we're in the picture the label occurs right it is the relationship between the pixel the label the same right.

When I say that the ψ is independent of I that is what it means right that need not be the case right suppose on the edge if a particular pixel value occurs it might have a higher probability of being a background right I mean it is very you typically not going to frame a picture so that the foreground goes to the very edge rate but then that might happen but the problem is suppose I have a 256 x 256 image then I have to learn a classifier for every one of those positions or to learn a classifier that gives me the probability of X I given yI for every one of those positions.

And that is a very large problem right it is a very hard problem so what we typically assume that ψ is a independent of ψ (X I YI) is independent of I so I am going to estimate the same model across the entire image right so of course you can try to be more clever I mean if you know some things about the about the about the task we can be trying to be more clever and say that okay I am going to break it up into four classes of ψ okay for xi and yi.

I will break it into four classes I will use one class here one class here rather than the bulk of the image and so on so forth you can do things like that but yeah but typically without any prior knowledge you just assume that it is the same all along all three this has to be different right and their faults are the same all through then you are basically we can use one pixel and predict the whole image the label for the whole amazing it is yeah.

That can also be disabled at it do we end up doing that you could do that also yeah say if xi + 1 right whatever that could also be taken as being independent of I you lose more modeling freedom that way but you could do it that way as well yeah depends on depends on how much effort you want to spend in doing the building the model yeah.

IIT Madras Production

Funded by
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in