NPTEL

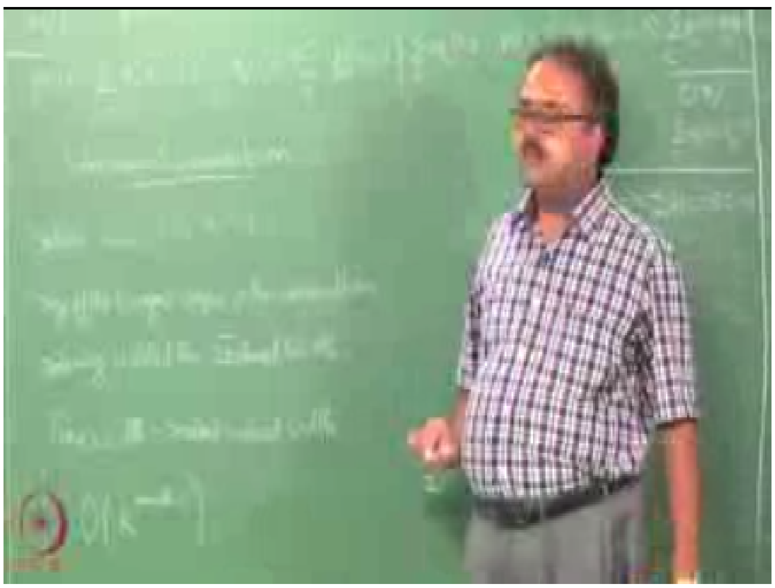
NPTEL ONLINE CERTIFICATION COURSE

Introduction to Machine Learning

Lecture-69
Belief Propagation

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

(Refer Slide Time: 00:16)



And we have a concept called the tree width of a graph, tree width of a graph which is the minimal induced with, the tree width of a graph is the minimal induced with so what do I mean by that so across all possible variable elimination orderings that you have right, so you can find out what is the induced with for every elimination ordering right, and there will be some ordering that gives you the smallest induced with that is your tree width such morals induced with.

So the complexity of doing variable elimination is actually the order of, but k is the number of values for each random variable, so in this case we assume k is 2 so it will be order of 2 power tree width right, so what would be the tree wit for a tree one or two depending on how you count tree width some people count tree width that is the size of the click -1 okay, in which case it will be 1 right, if you count as the tree with as the size of the click it will be 2.

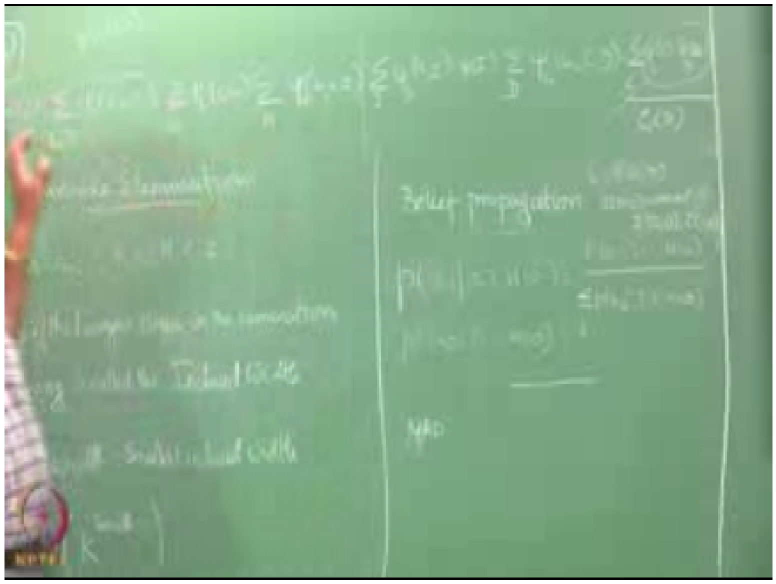
Because I can eliminate it from the leaf to the root right, at no point we will add a larger factor so every time I will be just collapsing one at a time right, no point will be adding a larger factor so if I eliminate it from the root to the leaf then that might be problems right, is some arbitrary ordering but the smallest thing is eliminate from the leaf all the way back to the root so every time you will be just removing one edge, right.

So if you think about it the smallest elimination ordering for us here right, also started off with kind of like the single node hanging of here right, so you have to eliminate C first if you eliminate C at the then you ended up adding some other nodes along the way right, so getting rid of here somehow coming from the outside inward right, or going from inside out was a bad idea so that is essentially well. So for trees variable elimination is great right, because it does things kind of in the best possible way you can expect it right.

But still there is a problem, what is the problem I asked you to find p(j) right, if I had asked you to find p(h) right, you basically have to redo this computation all over again right, so many of these tables that you computed internally could actually be reused that if I want you to find p(h) in fact up till this point everything can be reused, in fact some of this computation also could be reused right, in appropriately modified form right.

So but then you end up doing everything all over again right, if I asked you to do p(h) right, so there are more efficient techniques where you can keep cashing these things away right, the most popular of this is called.

(Refer Slide Time: 05:26)



Most popular of this is called belief propagation right, wherein you have some kind of a incremental way of computing these τ factors right, that you have by passing what are known as messages between the nodes okay, and the nice thing about belief propagation is it loves you to reuse a lot of the computation that you have already done for answering different marginal queries, okay any question so far. So I can answer any marginal queries you can see that right, so I basically have to some out all the other variables and have to find out now appropriate variable elimination ordering right, and then do it.

So the trick here is finding the right ordering, so I gave you the right ordering right, but in arbitrary graph finding the ordering is actually NP hard right, finding the right order is Np hard so you just have to do the best that you can and tree it is easy you can immediately see in trees you can go from leafs to the root but in an arbitrary structured graphs right it is hard to find out what is the right ordering, okay.

Great, so what about queries like, what is it probably that I probability that I let us say get a job, given that I am intelligent but I am not happy right, and also what is the probability that I do not get a job so this is essentially a conditional marginal line right, so I condition on some variables and I want you, I want to know the marginal right, so we know that everybody here is intelligent so once we figure this out if j=1 is lower than j=0 what should you do, be happy somebody said that yeah, so be happy yeah that is it, that will actually put you in we do not know we have to evaluate the marginal for that right, that does not guarantee your job but being happy at least well leaves you happy right.

No, that is one of the things forgetting a job right, you want to be happy so if you choose to be happy already it becomes irrelevant, anyway so what do you do for this right, so this is essentially I can just computer this as right, it is essentially ratio of two marginal’s but this is one marginal so this is another marginal what is this a marginal over, it is a marginal over I=1, H= the marginal probability of I=1, H=0 right, so this is a marginal probability of j=1, i=1, h=0.

So essentially I will have to eliminate all the other variables I will be left with one table and from that table I can read this value right, so if I eliminate all the other variables I will be left with one table that has j, i and h as the entries in it right, and I can just read off the entry corresponding to j=1, 1=1, h=0 right, and this one again is another marginal so once I know how to compute marginal’s I can also answer questions about conditionals, correct, great. So last thing then we will stop there will be end of graphical models for now right.

Suppose I am not interested in marginal distributions right, so this is another inference query I mentioned this in the last class right, the second kind of inference queries we would be interested in our map queries right, so why would we be interested in map queries? In fact many of the classification and other things we will be talking about we are only interested in map queries quite often right, so I will give you some image that I want you to label the image and I am interested only in the map estimate of the label right, I want you to give me a label I do not want the distribution over the labels I want you to give me a single label so in which case I need the map estimate.

So which label is the most probable according to the posterior right, so that is essentially what I need so for finding the map estimates so what do we do in this case, so what I am going to do is once I decide this kind of an ordering right, I am going to replace the sum here with a max, right so I am going to say look this is max over C right, of something right so this probability is I will compute then what you do is, when I do this max all over and I finish the computation right, I will get some probability right, the probability is the map probability, the probability of the most probable point.

Now how do I recover the most probable point at this computation is only for a probability right, so when I say do a max over C what does it mean for every value of D you will be entering one value right, you have a τ and D right, so $\tau 1D=0$ will be that probability right, for which is maximal whether you look at essentially you will have when you finish this computation right, so you will have some factor that is called a $\tau'(C,D)$ when you finish taking the product will have some $\tau'(C,D)$ right.

So $\tau 1(D)$ will essentially be something like this $\tau 1(D)=0$ will be max of $\tau'(0,0,1,0)$ right, this is what I mean by taking the max, so my $\tau 1D$ will be the max of okay when D=0 and C=0, to 0 c D=0 C=1 now these two entries whichever is the largest I will put that as τ 1(D)=0 likewise for $\tau 1(D)=1$ I will take $\tau'(0,1$ and $1,1)$ whichever is larger I will put that there, okay. Now I will go here repeat right.

Now I will be eliminating D here right, so for each value of $\tau 2(Gi)$ right, I will figure out which is maximum across D=0, D=1 right, I will put that in my τ 2 entry, so likewise I keep going until I finally get a product right, and now how will I recover the actual point now I have found out which is the most, what is the probability of the most probable configuration, how do I find out what is the most probable configuration.

I keep track of which one gave me the max right, so I keep track okay here for $\tau 1D=0$ did I get the max from C=0 or did I get the max from C=1 right, when I had D=1, did I get the maximum C=0 or did I get the maximum C=1, so I keep track of in every stage I keep track of which entry gave me the max right, and then once i finish the computation I just go back and read out the max entry so that essentially gives me the map the probability of the most probable point and also this most probable point, right.

So if you have a tie then you can choose 1 so it will give me at least one of the most probable points okay, so that is essentially how you do map estimates and yeah, so it is rather bad because it is exponential in the largest factor but it is useful for small graphical models, because most of the other methods have a significant overhead in setting up the entire process, right. Suppose you want to do believe propagation you will have to set up the data structures corresponding to the messages and it is a little bit of overhead is there in terms of computing, right.

If you have a very small graph like the earthquake graph okay, the earthquake graph you can graph you can do inference by inspection right, just look at the graph and do inference you do not have to even do any computation right, so some slightly larger graphs like this right, so where you have to actually do some computation you can do variable elimination it is fine. But when you start talking about running it on images right, so we told, I told you right we have like a lattice like structure one node for each pixel in image and then I want to run this on a 256x256 pixel image, right.

Then you really need some help right, and then such cases variable elimination is not the right thing to do right, because the first case the tree width can be large right, for that and so there are more efficient ways of doing it and even belief propagation right, in this kind of directed acyclic graphs right belief propagation is actually an exact algorithm right, even though it is pretty efficient it is an exact algorithm so can still be time consuming. So people look at answering queries in an approximate fashion, so I will not be able to tell you what the exact probability is, I will not be able to compute the exact that same map probability right, but I might be able to give you the map is the actual point that has the highest probability.

But if you ask me what is the highest probability I might not be able to give you that accurately, so those kinds of approximations we are willing to take so that we can do inference efficiently.

IIT Madras Production

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

www.nptel.ac.in