NPTEL

NPTEL ONINE CERTIFICATION COURSE

Introduction to Machine Learning

Lecture 53

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

Undirected Graphical Models-
Introduction & Factorization

Okay so we will be looking at so we look at me looking at graphical models right so we looked at belief networks right so it also said they will call Bayesian networks Bayesian belief networks etc, etc and then we looked at the concept of concept of yeah Cosell and then we looked at D separation right so we looked at what D separation is and then I asked you about thinking about D separation and we also discussed what is the question of inference.

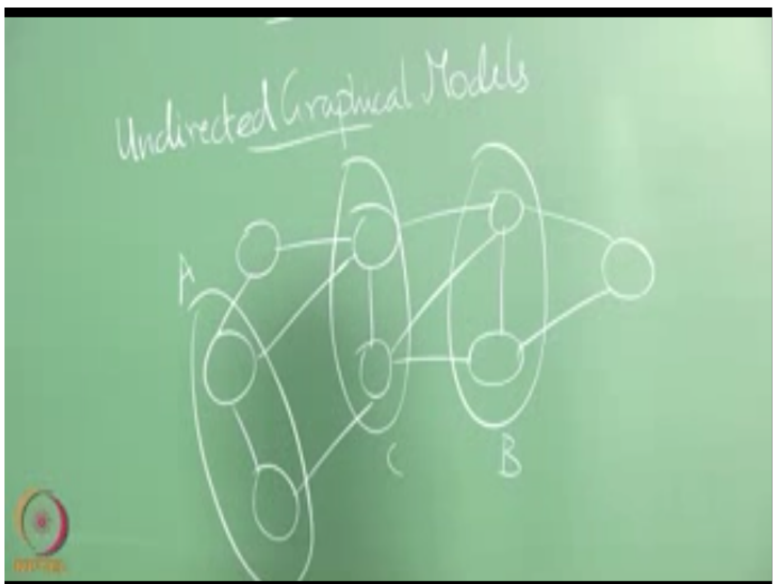
So what is the question of inference I will give you certain variables certain observations I can ask you questions about the conditional distributions given that the alarm rang what is the probability that there was a or given that may recall what is the probability that there was a earthquake right so those kinds of questions I can ask that if you think about this so what are the questions.

I am asking in that case they were all marginal right, right I have the Joint Distribution of Mary alarm John earthquake and burglary right this five variables so the whole system is specified by a a joint distribution over these five variables right but I was asking you a question but a marginal it what is the probability that earthquake happened okay that is a specific marginal I am asking a question about right.

So typically inference questions will be about marginal's or conditional marginal's right so conditional marginal's will be when I am given some observation given that may recall what is a probability of an earthquake so that is a conditional marginal right so if I just ask a question okay nothing happens what is the probability of a earthquake in California right so I want to ask that question I can still do that then I can take this whole entire network right marginalize out of everything I can tell you what is the probability of earthquake happening is not very interesting because it is already I give you a the marginal as one of the components in specifying the network right.

All I can do is just look up the earthquake probability distribution will be given you the marginal this one of the things that specify it right I can ask questions like okay what is the probability that Mary will call right I do not know anything else just ask the question hey what is the likelihood that Mary is going to call me today right so I can just marginalize over all other variables and I can give you that answer so this essentially queries or on some kind of marginal’s or conditional marginal’s it could be joint distributions as well it could be joined to marginal’s as well.

(Refer Slide Time: 03:10)



In the sense that what is the probability that Mary and John both will call me today is a joint distribution over a subset of the variable so it is some kind of a joint marginal rate so instead of looking at the full joint distribution so those are the kinds of queries I am asking right so we look at the directed networks so today we look at using undirected graphs look at undirected graphs and or like a very complex so I am going to call this set of nodes as a where these two nodes area these two nodes are bees just asking you this grouping the random variables together here right.

So as before each node is a random variable just like we had in the directed case each node is a random variable and the edge denotes some kind of dependence between the two random variables right so in the directed case you could confuse an edge for a causal relation right you could keep the edge as representing a causal relation right but here.

I removed the arrow direct there is no direction here so it is just some kind of dependence between two variables right you could say that but there is a slightly different semantics associated with the edges here right so see the edges encode some notion of conditional independence just like the edges therein encoded some notion of conditional independence they just here also encode some notion of conditional independence right so it is not like yeah.

So there is a subtle difference between the two right so I am not going to get into it because it is really subtle but the class of conditional independence is are different right there some that you can represent using directed networks there are some that you can represent using undirected it works most cases you can choose whichever you want but there are some cases which is more convenient to go one way or that right.

So I do not want to get into the discussion at least not in this class it is for the class next semester if people want to get in to do that right so if we look at this if I try to take any path from a node in a right if you try to take any path from a node in A to a node in B I have to pass through some node in C so any path from A to B I have to past through some node in C correct now then I would say that see the notes in see separate the notes in p from the nodes in a right last time.

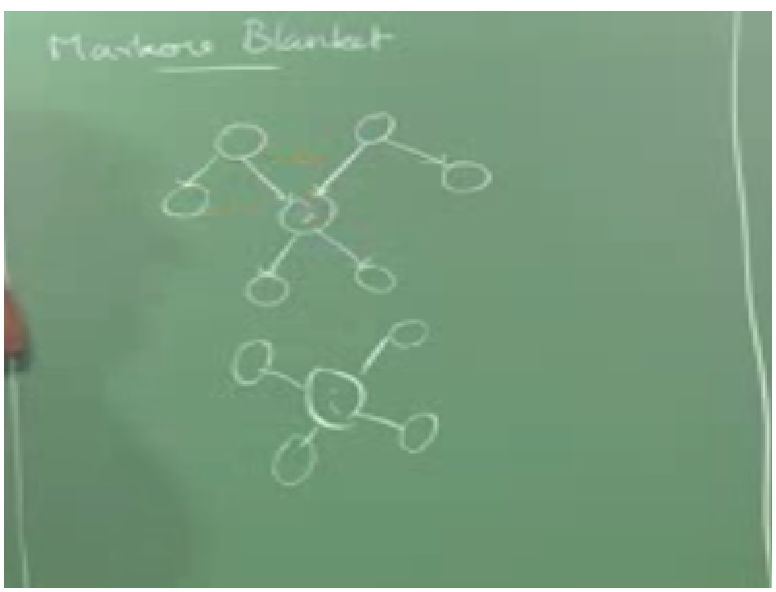
We had this notion of D separation like we have to write three different rules right D separation is nice for making up fun questions for exams right but, but it is a little confusing right so you have all these arrows are going to take care of the notion of separation here is very, very easy what it says is if there cannot be any path from A to B which goes through which does not go through see okay now like I put a double negative this every path from A to B goes through see right in which case then you say C separates AB okay.

The D is gone here right it is just simple separation that make sense so that is the simple enough let so this is what we started we started off with this notion of separation that is encoded in a directed model so in the undirected case I am telling you that separation is very simple like we had D separation we have separation here so next we have to think about something else right if you recall when we started the discussion of directed models.

We started off by looking at some kind of factorization of the Joint Distribution right so we had a very complex joint distribution over many, many variables so we started looking at some kind of factorization of it and from there we constructed the network right so the, the graphical model inherently implicitly encodes a factorization right it is so this all this separation business in fact ties indirectly to the factorization right.

So the D separation gives you the rules for the factorization also right so from a directed model it is very easy to write down the factorization you just look at the, the conditional distributions and you write down the factorization here we really do not have such kind of a one-way implication here and I am just saying something links the two variables together right so how do we go about doing the factorization right.

(Refer Slide Time: 09:34)



So there is something there is a concept called the Markova blanket and I forgot to mention when we did directed models right so a mark of blanket is essentially all the variables right that could potentially influence a given node right so given the node the parents can influence the node correct at anything else the children can influence the node right anything else siblings can influence the node right sibling yeah right.

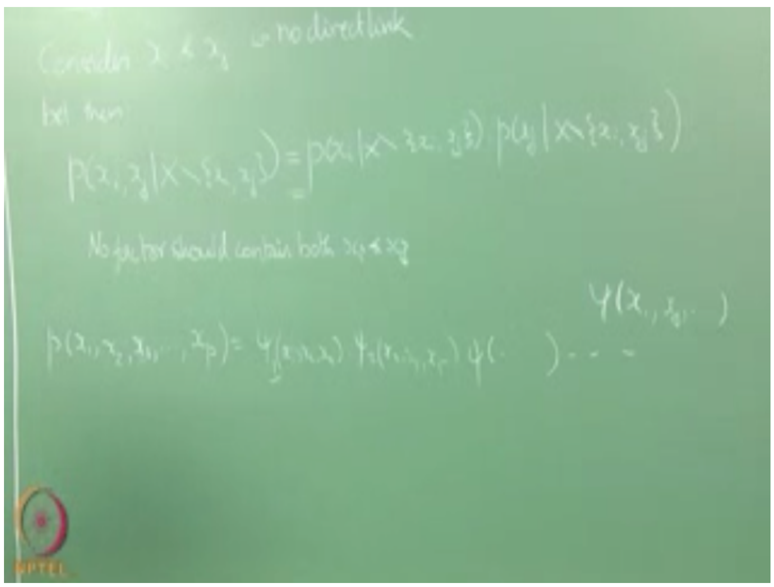
So that is essentially the Markova blanket over know right so I have a node I take node I right so this is essentially the right this is this is right so if I know if it do not know the parent then this node can influence this one right if I do not know the parent this can influence it is right here so this is essentially the Markova blanket of I right so what will be the Markova blanket is undirected case same let us be directly connected the neighbors of I in the undirected case is just the neighbors of I so these are connected to other nodes.

I do not care right so they cannot influence me except through this yeah I am looking at direct influence so that node cannot directly influenced me right so if I do not know this guy right so what is it thing so if I did yeah so let me think about it okay this one so if I know this node right these to get connected right if we know this node these to get connected right so, so essentially that is basically it so if I know this guy okay then I get connected to this guy so if I just condition if I say okay I know this I know this I know this I know this right okay so, so that is the Markova blanket of this I node I likewise here if I know these four guys then that is it basically nothing else will influence my right so it will cut off from everybody else write these four nodes will separate I from the rest of the network right and here.

I need to know all of these guys right before I am cut off from the rest of the network right so I you really do not need the Markova blanket right just an aside that I wanted to tell you know you need a markova blanket for other things but not for this lecture or for this course right so the Markova blanket is very important because you need it for doing inference later on right so I will clarify the thing but so essentially the idea behind the Markova blanket is once you know all of these nodes right.

You are completely separated from the rest of the graph yeah so if you know the children then these two get linked up right if I know the child but then the probability on I right so if you remember the third D separation rule we had right if I do not know the child then these two are independent right if I do not know the child in these two are independent information can go from one together but if you know the child then these two get link up.

(Refer Slide Time: 15:22)



So we will go back to the factorization consider two nodes x_i, x_j say sir there is no direct link between X_i and X_j right so something like this right so this could be X_i or could be exchanged so there is no direct link between them right let us say that the set X denotes the inverse of all variables right so it as X_1 to X_P so it is a text notes all the variables and if I condition the Joint Distribution of x_i, x_j on everything other than x_i, x_j right.

So what can you tell me about this distribution this should factor out right so people haven't encountered this before that is the difference okay so people are getting confused whether there is a sec difference right so this I am from x_i am reviewing x_i, x_j so conditioned on that right they should be independent right condition on the rest of the variables what is essentially I am saying is okay x_i and X_j I am conditioning it on everything else.

So every path between x_i, x_j has to go through well everything else right so I have conditioned not that I have to be independent so that is the basic assumption I have right the factorization I am assuming here is that the means or the rain conditional independence thing is if all path from A to B pass through see right then A is conditionally independent of B given C.

So here my C is everything other than x_i, x_j so this has to be true so if you think about it now when I write my factorization right when I am going to write me what I what I mean by writing my factorization I am going to take my probability of I am going to take this joint probability and are going to write it as some factor one on some set of variables right so like this I am going to write it out like this I this is what I mean by factorization right in the directed case.

I also did the factorization except that my SI had a very specific form my side would have been something like what is the probability of x_1 given x_2, x_3 right so that is a factor right and my side too could have been what is the probability of X_3 given X_4, X_5 that is another factor so like that I could have written out multiple, multiple factors in the directed case in the undirected case I have to figure out what these factors are going to be the directed case I already have figured it out under if it is have to figure out what this vectors are going to be right.

So we want this to hold right we want this conditional independence to hold so what can you say about the factors if there is a factor that directly connects X_1 X_I and X_J there is a factor right in all these factorization I have and some factor that has have some factor like that right so now there could be some assignments of this some, some kind of derivation of these factors which will connect X_I and X_J so I cannot say that X_I and X_J will be independent right regardless of what values I am assigning to the factors.

If there is a factor that actually has x_i, x_j in them right then I cannot be guaranteed of this independence for all assignments to those factors the only way I can ensure that this independence will hold is if I guarantee that x_i, x_j will not appear together in any factor so I cannot have a sigh function here in the set of factors that I write I cannot have a sigh function that has both x_i, x_j s_r givens is it clear the reasoning gear if x_i x_j appear in any one of this size together right.

Then I can assign some use to the size is that those two get connected right so to make sure that I can never do that that this conditional independence holds for all possible distributions I can write so I have to make sure that there is no factor no factor should contain x_i, x_j so I am going to take this intuition right and flip it around if there is no edge between X_I and X_J then no factor should contain x_i, x_j right if there is an edge between x_i, x_j that should be a factor that contains x_i, $x_{j.}$
And I am flipping it on if I want to encode the dependence between xi, xj there should be a factor that contains xi, xj suppose there is an edge between xi xj xk and xk xi so what would you want to do you can put in factors for every edge right alternatively a more compact way of doing this is to put in a factor for the, the click right so if you have some set of variables which are fully connected.

Then you will put in a factor for all of the variables together because there is this three-way dependence right so instead of putting in this three different factors for that you put in a single factor that looks at all of this together right so flip this around and then say that for I want me to repeat what is I represents because I did not I did not even say what it represents but I will tell you what it represents but the thing is it is just some factor right in directed models rights I was the conditional probability.

So essentially I am taking one complex function right I am taking this P of X_1 to a X_P right okay so I am taking this function that is a very complex function I am writing that out as a product of many smaller functions let each of this sigh is a one sense function x_1 equal to some person the next I else is a constant and then yeah sure yeah so that would be the case right so when I am conditioning on every other variable that means I am essentially assigning a value to every other variable.

So all the other factors will reduce to some fixed value right and then I will have only x_1 x_2 left right even then my choice for x_1 right would depend on the choice for choice for x_2 so that is one way of interpreting that yeah right so people got his way of looking at it he says it fix all the other values right so let us assume that sighs I one has x_1 and x_2 and x_1 x_2 do not appear in anything else just for simplicity sake right so fix all other values so everything else will reduce to a constant right.

So we say x_1 x_2 or constant and I am looking at the probability it now whatever value H is for x_1 will depend on X to write the probabilities are saying for a value of x_1 will depend on what is the value of taken for x_2 there is a way to make, make it independent even then right so I can just give the same value for same probability for x_1 equal to 0 when X_2 equal to 0 and when x_2 equal to 1 right.

So I can assign the same probability there is a very specific assignment that can make it look independent so something like this right, right so what is the probability that X_1 is 0 so I am conditioning it right so the conditional probability I can say that okay so this is 0 point to this is 0 point to this is 0.8 this is .8 right now x1 is independent of x2 even though I have a factor sigh x1 x2 one is independent of X to write but I have to be very careful about actually assigning the values to the factors right.

So what we are trying to do is give the factorization but regardless of what function we choose for sy1 right it will it will be it will end up being independent if a condition on everything else it will be independent so I can put in whatever random numbers I want there okay I still want x1 to be independent from x2 so that cannot be the case if there is a factor that has x1 x2 in it for each click you have a factor in this yeah you put in a factor so I mean you are designing how the factorization should be right.

So I am just saying for each click you, you include a factor yeah well the other way around so x2 is independent of x1 no, no it is correct write x1 is not even of x2 right so whether x2 is 0 or x2 is one right so it does not matter right so what is the probability of x1 equal to zero given X2 equal to 0 right so that is equal to the probability of x1 equal to zero given equal to 1,2 the probability of so it is independent no I said it is conditional my way no, no this is right I am writing the conditional.

I am only writing the probability of x1 given x2 I am not writing the Joint Distribution so each column should be a valid distribution lower is we know each column should be valid distribution do not need not be no need not be why I am writing the conditional r-8 I am writing the probability of x1 given x2 so why should the roast be able to join joint PDF should be well with its it is not John I understood in the conditional ordered in the joint yeah.

So whatever so I to get the joint I need to define a distribution over x2 right so which I am not then I just looked at one factor here so you could have a factor over x2 few one but I have not done that yeah what do you what are we so I am see for the directed graphs right so you can easily write down what the factorization is right so I have some probability the Joint Distribution remember all of these graphs represent joint distribution over P variables right p is hard.

Let us let us make it again okay I am going to make it so having probability as p and the dimensionless p is a little confusing for me so I have n variables whether it is a directed graph or whether it is undirected graph so what I am trying to represent is the probability destroying distribution over n variables right and the whole idea of going to this graphical models is that I do not want to define this whole N squared minus 1number of values for specifying my probability distribution right.

So I want to say that know it really in probability distribution is not so complex they are not n squared independent values in my distribution so there is some kind of factorization that will happen right so I am truly trying to figure out which are the independent values that I have to specify so that I can get the entire probability distribution again right so for that I am finding out what is the right way to factorize my probability distribution sure but can be more than one factorization by the way why what is the confusion right.

So if you think about even the directed models right so I you not down one factorization but you can always think of flipping the direction of the edges right and writing another factorization right so it is nothing to it there is nothing very sad across about this the reason we choose this factorization is because it is easy to handle these things so that is only reason right and so in the directed graph we knew how to find the factorization so we are we look at the conditional independence.

So in the undirected case right so we need to come up with some way of finding what the factorization this right so what I am saying is so what are the things the factors should not have that first if we characterize that then we can go and write down what should be the factorization what are the things the factors should not have is that if there is no edge between two variables they should not appear together in a factor right if they appear together in a factor then there is a way of assigning values to the fact right.

So here if you think about it this is actually a factor in a conditional in a directed graph this could potentially be a factor in a directed graph and so for example something like this so this could be it right but no, even though I have written this dependence factor I have assigned values here so that the dependence is not hold right but I can assign something else here right so then this equality is broken so if I have a factor that has both x1 and x2 in it when I write this factorization then it is possible for me to assign some values.

So that this independent does not hold so when we are talking about these factorizations what we are really looking for is a representation that regardless of the numerical values you eventually end up assigning to it preserve the independence relations that you are looking for they will introduce additional independence relations but at least the ones that your guarantee should be there in the probability so when the graph guarantee is something the graph structure guarantee some independence relation okay.

The factorization you give should guarantee that independence also so if I put in a factor that connects x1and x2 when there is no edge between x1and x2 I can no longer guarantee that so that is essentially what we are trying to look at right so any other questions I know it is always a little too key when you move to undirected graphical models right so director graphical modals are all very easy and everyone gets it the first time around so undirected graph demons are little confusing right.

And it is good too it is good to look at them from the very beginning because a lot of the techniques that we use for inference right or in some sense common to both directed and undirected models right so if you understand the, the two models as something common right having something common in between them then all the techniques that we look at afterwards I can just talk in one language I do not have to do inference techniques for directed models and then go back and do inference techniques for undirected models they are all the same right operationally.

There might be so like implementation wise you might do something a little bit more efficiently if you know that you are working only with directed models right or if you know if you are only working with the undirected model but the algorithm wise it is pretty much the pretty much the same okay so let us spend a little time and try to understand undirected model so any questions on this so far condition no we should not have a factor Z that connects x in exchange that have both xi, xj as an argument these are independent only in the case yeah.

But I can assign values or time seeing it I can always assign values to that functions I says that they become they become connected see the whole idea of giving a factorization is regardless of what you do with that side that you can o whatever you want with the site but I still want to guarantee the independence so as soon as I put in xi xj in the same function I give you the freedom to do something with that function to introduce the dependence so we do not want to give the freedoms to the process by which you are determining the values for the options.

So that is the restriction that we want to introduce you great so I am I am flipping this around since we say that if there is no edge there should not be a factor and I am saying that if there is an edge there should be a factor right in fact I want to go a little further and say that if there is a click right that should be a factor associated with the click so better let suppose I have a graph like that so I have four random variables right so what should be the factors here I should have something that connects1 2 3 1 2 3 4 is not it should have something that connects one two three four why is not a click right so 14 is not connected.

Therefore I do not want a factor that has 114 in it at the same time I cannot have a one two three four factor so it can be one two three two three four anything else right so you automatically made my point for me right all I need to consider are maximal cliques right if you think about it 1,2 is a click right 2,3 is at League 3,4 is a click everything is each edges by itself is a click right so you, you did not give me factors for the clicks you gave me factors for the maximal cliques right it turns out that if you introduce factors for the maximal cliques you essentially have the same representation power as having factors for all the cliques in the graph right so you can restrict it to so this case is that and that so we have only two factors okay is it here.

IIT Madras Production

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

www.nptel.ac.in