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-2

Lecture-10

Welcome to NPTEL MOOCs course on machine learning and deep learning fundamentals and applications. I have been discussing about the concept of Bayesian decision theory. Today I will be discussing another topic. The topic is Bayesian belief network. This is a graphical model. With the help of this graphical model, I can take a classification decision.

In the graph, I have nodes. Actually, it is the random variable I can represent by nodes. And also in the graph, I have edges. So this edges actually represents the dependencies.

So in case of the Bayesian belief network, the conditional dependencies can be represented by directed graphs. And based on this belief network, I can take many classification decisions. So let us discuss about the Bayesian belief network. So the introduction, a Bayesian network is a probabilistic graphical model. So it is a graphical model and it is a probabilistic model and which represents a set of variables and their conditional dependencies.

So we will consider the conditional dependencies of the variables using a directed acyclic graph. So this is the definition of the Bayesian network. And the dependency of variables can be represented efficiently by the Bayesian belief network or I can say the belief network or maybe the Bayesian network. Also, the Bayesian network allows us to represent a joint probability density that is the probability of X, Y, Z that is the joint probability density that can be represented by considering the dependency relationships. So this concept I am going to explain what is this concept and you can see this conditional dependencies can be represented by directed graphs, the acyclic graph.

So a belief network is usually a directed acyclic graph. Each node represents one of the system variables. So nodes I can consider some random variables and each variable can assume certain states. So some probability value I can assign corresponding to these variables. So each variable can assume certain states.

So what is actually this belief network? This is actually I can say this Bayesian network. It is a compact yet expressive representation. And the second point is efficient reasoning procedures. So that Bayesian network is the compact but yet expressive representation and efficient reasoning procedures.

So this Bayesian network identifies a joint distribution in a structured form.

So that I will explain and it also represents independence via a directed graph. So this represents dependence and independence via a directed graph. So in the graph actually we have nodes. So in the nodes I have the random variables and I have the edges in the graph. So this edges shows direct dependence.

So structure of the graph, structure of the graph that is actually the conditional independence relations. The structure of the graph actually it shows the conditional independence relations. And one important point is the graph is acyclic.

This is an important point. The graph is acyclic.

This is an important point. The meaning is no directed cycles. So two structures or maybe I can say two components of the graph. One is the graph structure that is mainly based on the conditional independence assumptions.

So the graph structure is mainly based on this condition that the conditional independence assumptions and in the graph actually we are considering some numerical probabilities.

So that I will explain later on what is the numerical probabilities corresponding to this graph. So these are the components of the Bayesian network. One is the graph structure. Actually it represents the conditional independence or conditional dependence.

So it is the assumption is the conditional independence assumptions and

also the numerical probabilities are assigned in the graph.

So let us discuss what is the general form of the Bayesian network. The general form of the Bayesian network is the probability $X1, X2, \ldots, Xn$. So this is actually what is the probability. This is the full joint distribution. So this is represented like this.

This is the product for all i's. So i is equal to 1 to n and Xi, Xi is the variable and the parents of Xi. So this is the approximation. This is the graph structured approximation. So in this case what we are considering the conditional independence assumption.

If I consider the conditional independence then I can consider the representation. So we are considering the full joint distribution. After this I am getting the graph structures approximation. So suppose I can give one example.

Suppose I have three variables A, B and C.

So we are considering joint distribution probability of A, B, C. That I can represent like this. Probability of C given A, B that is the conditional probability. Probability of C given A, B and also I can write like this probability of A, probability of B.

So this joint distribution can be written like this.

And corresponding to this equation, I can say equation or the these probabilities I can represent in the graph. So one node is A, one node is B. These are independent node and if you see C, the node C depends on A and B. So I can show by the directed graph.

So the node C depends on A and B.

So I can represent like this. So that is nothing but directed edges, directed edges. So these are the directed edges. If you see these are the directed edges. That represents direct dependence.

And if I consider suppose no edges or I can write absence of an edge, absence of an edge that is the conditional independence. So here you can see if I consider A and B, if I consider A and B, there are no edges and actually this A and B they are independent. A and B are independent. But if I consider C, C depends on A and also it depends on B.

So move to the next slide. Suppose I want to draw the belief network. So suppose belief network is given, belief network is given. The network is something like this. A, three nodes are there, A, B and C. So it is a three way Bayesian networks.

So corresponding to this I can write the joint probabilities or joint distribution like this, probability of A, B, C that is probability of A, probability of B and probability of C. This is actually the absolute independence. This is the absolute independence. Now suppose I can give another example. There is probability A, B, C is equal to suppose probability of B given A, probability of C given A, probability of A.

Suppose I am showing like this. So suppose the variable A, it represents the result and this B and C, I can say symptoms given A. So A is the disease and B and C are symptoms given A. So we can model B and C as conditionally independent symptom given A. So you can see this can be represented like this. A is the disease and B and C, they are symptoms.

So B and C, I can write like this, B and C are the symptoms. So here you can see the B and C are conditionally independent. So I can write B and C are conditionally independent. So in this example what we are considering, A is a disease and we model B and C as conditionally independent symptoms given A.

So you can see here in the graph, the B and C are conditionally independent.

So this is the representation I am considering. Now move to the next example. Suppose it is given like this, probability of A, B, C that is equal to probability of C given A and B. So how to represent this one? So this is represented like this.

C depends on A and B, but A and B, they are independent.

So A and B are independent. Suppose let us consider another graph A, B and C and you can see the dependency. The dependency is represented like this. This is the dependency. So corresponding to this, this probability of A, B, C, I can write like this. Probability of C, C depends on B and B depends on A and this probability of A, it is independent.

So you can see the C depends on the previous state or previous variable. The variable is B and B also depends on A. So this type of dependence is called the Markov dependence.

Markov dependence. So this is one example. So I can give some of the structures like this. Suppose if I draw a structure, another structure I can draw like this. Suppose D is a node, C is a node and suppose A is a node. This is another example.

I may have this type of graphs. So it is B suppose and suppose it is E. So I have five nodes. So corresponding to this, you can see the dependence. You can see the dependence. The D depends on A and also I am considering this is, so D depends on C also because already I told you I have to assign the probabilities corresponding to these variables.

So probability of D given A and C. So this probability we can consider and this C, you can see if I consider this, this is the directed graph. So corresponding to B, the probability is suppose probability B and corresponding to A, suppose the probability is A is a probability and you can see that E also depends on D. So this probability I can write the probability of E given D and this C depends on B. So I can write the probability of C given B. So I am assigning probabilities to all these variables.

So this is one example of the Bayesian belief network. So we have to consider this type of networks. Now let us consider one problem. So how actually we can take a classification decision based on the belief network.

So let us move to the next slide. The slide is how to determine or how to take a decision based on the belief network and this is the belief network for fish. So here you can see I am considering these variables A, B, X, C and D. So A represents season. So season of year I am considering. So A1 may be winter, A2 is spring, A3 is summer, A4 is autumn.

So these are the states corresponding to the variable, the variable is A. So corresponding to this B, the location, I have two states, one is B1 and another one is B2. So B1 is the North Atlantic and B2 is South Atlantic. And similarly if I consider X, X is the fish. So two types of fishes we are considering, one is the salmon, another one is the sea bass.

And also the weightness, the lightness of the fish that variable we are considering as C. So it may be light, it may be medium or it may be dark. That is the lightness of the fish. And also the thickness of the fish corresponding to this we are considering the variable, the variable is D.

So D1 is white and D2 is thin. So corresponding to all these variables you can see the probabilities we are assigning. So corresponding to suppose the A, the season, the probability of A1 is 0.25, probability of A2 is 0.

25, probability of A3 is 0.25 and probability of A4 is 0.25. So we are considering all these probabilities corresponding to the state, the state is A, that variable. Similarly, corresponding to B, I have the probabilities, the probability of B1 is 0.6, probability of B2 is 0.

4. And after this if you see, I am considering the dependence. So X depends on A and B. This variable X depends on A and B. So that is represented by the probability of X given A, B. And corresponding to this you can see this table.

In the table we have considered these probabilities. In the first column you can see the probability of X1 given that Ai Bi or Ai Bj. And similarly, in the second column we are considering the probability of X2 given A and B. This we are considering. And corresponding to this all the states we are considering first A1, B1, A2, B2, like this we are considering.

So all these probabilities we are defining like this. These are the probabilities in the table. These are the conditional probabilities. Similarly, corresponding to the variable C, the C depends on X. So we are writing probability of C given X. And corresponding to this you can see I am defining the probabilities, the probability of C1 given X1, probability of C2 given this X, X may be X1 or X2, and probability of C3 given X, X may be X1 or X2.

And corresponding to this I have all the probabilities are defined.

So it is 0.6, 0.2, 0.2, 0.2, 0.3, 0.5. And similarly, corresponding to the variable D, I have two states, one is D1, another one is D2. So now I am defining the probability that is the conditional probability, the probability of D given X. So this D1 given X1 like this, D2 given X1. So all these probabilities we are defining, these probabilities we are defining, we have these probabilities. So because corresponding to D, I have two states D1 and D2, and X we have two states also X1 and X2.

So X1 corresponds to the fish, the fish is 7, X2 corresponds to the another fish, the name of the fish is C bus. So this is the belief network for fish. So which fish we have to select, that is we have to go for a decision making. So which fish we have to select based on these conditions. And after this, this link joining these two nodes is directional and represents a causal influence.

So already I have explained what is the causal influence, X depends on A or I can say A influences X. Here you can see in this graph, X depends on A, that means A influences X directly. And you can see A influences C indirectly through X. So if you see the C and another one is A, C depends on X direct dependence. But if I see this graph, A influences C indirectly through X, that means C also depends on A, C also depends on A via X.

So A influences X directly and A influences C indirectly through X. So these directional graphs we are considering to show the dependence because we are considering the dependence of the variables. So these directed graphs we are considering to show the dependence. And each variable is associated with a set of weights, which represents prior or conditional probabilities. So the probabilities may be discrete or continuous. So already I have shown these probabilities we are considering corresponding to the state A, I have the probability the probability of A.

So I have probability of A1, A2, A3, A4. So all these four probabilities Similarly, probability value of B corresponding to the variable B we are considering two probabilities probability of B1 and probability of B2. And after this you can see we are considering the probabilities the conditional probabilities we are considering the probability of X given A, B that is also we are considering, probability of C given X that is the conditional probability we are considering, probability of D given X that is the conditional probability we are considering. So all the possible probabilities we are considering here. And also you know that the probabilities the sum is equal to 1.

So if I see this row, the 0.25, 0.25, 0.25, 0.25, so ultimately it will be 1, because some sum of all the probabilities should be 1.

So similarly, if you see here this line, so 0.2, 0.3 and 0.

5, so it is 1. So in all the cases like this 0.3, 0.7 it is 1, 0.6 and 0.4 it is 1. So sum of the probabilities corresponding to a particular condition it is equal to 1. So that we are considering. Now we are considering the same rule that is the same rule in the probability, you know the same rule in the probability. The joint probability of a set of variables, the set of variables are X1, X2, X3 like this.

So we are considering these variables and we are considering the joint probability. So this is nothing but a joint probability. So by same rules this joint probability you know how to write. So we can write like this probability of X1 given X2, X3 up to Xn.

Next the probability of X2 given X3, X4 up to Xn. And finally the probability of Xn. So we can represent like this. And we are considering the conditional independence. The conditional independence relationship encoded in the Bayesian network state that a node Xi is conditionally independent of its ancestor given its parents, the parents are suppose pi i. So based on this if I consider the conditional independence that can be represented like this. So the joint probability is represented by this one based on the conditional independence relationship.

So the probability of Xi given pi i. So pi i is nothing but the parents, the parents of Xi. So i is from 1 to n. So suppose I have n number of parents. So that is you can see by considering this one, I am representing the conditional independence.

And this relationship is very important because the same rule is very important. So if you study the probability course, so you can get this same rule. So with the help of the same rule, you can see how I am representing joint probability. So this is about the joint probability and the concept of the conditional independence. So now come to this problem.

The problem is the classification of the fish by the belief network. So we can compute the probability of any configuration of the variable in the joint density distribution. So suppose this is the problem. The probability of catching a medium lightness thin sea bass from the North Atlantic in summer. So corresponding to this what is the probability I can determine. So what is the probability of a3? a3 means that means we are considering the probability of catching medium lightness.

So which one is the variable corresponding to the lightness? The variable is C thin. Thin is the variable D. So this is D2 and sea bass we are considering the variable X the sea bass is X2 sea bus from the North Atlantic. North Atlantic is the variable we have to consider B.

So North Atlantic is B1. North Atlantic is B1. B2 is the South Atlantic. So we are considering all these variables. So probability of A3 that is A3 means it should appear in the in the summer. So summer is A3.

So summer we are considering. So A3 is considered that is a summer we are considering. What is B1? B1 is not Atlantic. What is X2? X2 the name of the fish is sea bass.

What is C3? C3 is the dark. So C3 is dark. And what is D2? D2 is the thin. So we are determining this probability. And you can see this probability of A3 that is independent. Probability of B1 that is also independent.

And X2 depends on A3 and B1. The probability of X2 given A3 and B1. So that is why we are considering this probability. That is the conditional probability. And probability of C3 given X2.

So because C3 depends on X2. So we are writing probability of C3 given X2. And probability of D2 given X2. So D depends on X2.

So that is the direct dependence. So D2 depends on X2. So all these probabilities I am defining. And in the table or in this list we have all these probabilities. So these probabilities I am just taking from the table. And if you see what is the probability of A3? The probability of A3 is nothing but 0.

25. This is the probability. What is the probability of B1? The probability of B1 is 0.6. So all these probabilities I am taking from this table. And ultimately after multiplication I am getting this probability. The probability is 0.012.

So that means corresponding to this problem, corresponding to this problem my answer is the probability is 0.012. That probability of catching a medium lightness, thin seabass from the North Atlantic in summer. So that probability we can compute. So this is a very simple understanding, the simple concept that is the concept of Bayesian belief network.

Now let us consider this example suppose. Suppose I have the nodes A, B, C and D. These are the nodes. And you can see I am showing the directed graphs. These are the directed graphs. So this probability of A is given, probability of B given A.

So that is also the conditional probability. Probability of C given B and probability of D given C. These are given. So in this case we have to determine the probability of D. So we have to determine this, the probability of D. So the probability of D, P D is nothing but summation over A, B and C.

So we have to consider this joint probability A, B, C and D. This joint probability can be represented like this by considering the same rule. The probability of A, probability of B given A, probability of C given B and probability of D given C. It can be represented like this or this can be written like this. Probability of D given C I can take outside here and suppose another summation over B, summation over B, probability of C given B.

So in determining probability of D, we are considering D given C. So for this we are considering all the Cs, the summation over all Cs we are considering. Similarly if I want to determine probability of C given B, we have to consider all Bs. So that is a summation over all Bs we are considering.

And the summation A, probability of B given A, probability of A. So we can write like this. So if I do this computation here, so this part, from this part to this part if you see, this part is nothing but probability of B. And if you see from this part to this part, this part to this part, this is nothing but probability of C. And if I consider the complete expression, the complete expression is this is the probability of D. So here you can see in this case, we have all these values, the probability of D given C, probability of C given B, probability of B given A, probability of A, all these are available.

So that means we can determine the probability at D, the probability at D we can determine. So determine the probability at D. So this is one example. Come to the next example. Suppose the belief network is given, the belief network is something like this.

Suppose E and suppose it is F, suppose it is G and this is H. So these probabilities are given, the probability of E is given and you can show or you can see here that F depends on E. So this is the probability of F given E and also here the probability of G given E that we are considering and you can see the dependence, the dependence like this, these are the dependence. So considering this graph, I have to determine the probability at H. So I have to determine the probability at H, that probability we have to determine. So how to determine this probability? You can see this probability I can determine, the probability at H I can determine that is the summation over E, F and G.

So probability of E, F, G, H I can write like this. So this can be written like this E, F and G, the probability of E because it is independent, E does not depend on any other nodes, the probability of F given E. So F depends on E, the probability of G given E, so G depends on E and probability of H over F, G. So these are the cases, so which can be written like this. It is F, G, probability of H, F, G, summation E, probability of E, probability of F given E and probability of G given E.

So we can write like this. So already we have all the probabilities defined in the graph and based on this we can determine the probability at H, the pH we can determine. So this probability we can determine. This is about these two examples we are considering how to determine the probability at a particular node we can determine. So come to this classification problem.

So the classify fish given that the fish is light, so light is defined by C1 and was caught in South Atlantic. So the variable B2 we have to consider and no evidence about what time of the year the fish was caught nor its thickness. So based on this, what is the problem? So already I have shown this belief network already I have shown that is already I have explained, but now problem is slightly complicated. The complicated is no evidence about what time of the year the fish was caught nor its thickness. So that the thickness information is not available and also that season information that what time of the year the fish was caught that means this information is also not available.

The P A probability of A is not available and probability of D given X that is also not available. So we have to determine or we have to classify a fish given that the fish is light. So light means we have to consider the variable the variable is C. So this probability we have to consider probability of C given X we have to consider and was caught in South Atlantic that means we have to consider the variable B this probability of B we have to consider, but no evidence about what time of the year that means the season information is not available. So we have to we have to we should not consider this variable probability of that information is not available and also that the thickness of the fish that information is also not available.

So corresponding to this how to determine this one. So we have to determine the probability of X 1 given C 1 comma B 2 this you have to determine this is the problem. So what is the problem. So we have to determine or we have to classify a fish given that the fish is light. So light means C 1 and caught in the South Atlantic that is B 2 and no evidence about what time of the year of the fish was caught nor is thickness.

So that information is not available. So we have to determine this probability. So that is equal to probability of X 1 comma C 1 comma B 2 divided by probability of C 1 B 2 I can write like this these probabilities. So from the probability theory I can write like this and after this we are considering this one. So alpha is the proportionality constant the summation over A and D. So probability of X 1 A B 2 C 1 and D.

So in this case we are writing A is a vector because A may have all the values maybe it is A 1 A 2 A 3 A 4. So that is why we are considering A as a vector and similarly D also we are considering as a vector because we do not have the information about the thickness of the fish. So that means that D has these two components one is D 1 and another one is D 2. So D 1 is the wide and D 2 is the thin. So that is why A and D we are considering as a vector.

But we have this information X 1 we have to consider B 2 we have to consider C 1 we have to consider but I do not have the information of A. So that is why we have to consider all the cases A 1 A 2 A 3 A 4. So A is represented as a vector and D also I have to consider as a vector because I have two information one is D 1 another one is D 2. So alpha is the proportionality constant so it is alpha summation over A and D A and D is the vector A and D the probability of A probability of B 2 probability of X 1 A B 2 A is a vector already I told you probability of C 1 given X 1 and probability of D D is a vector so it is X 1. So we are considering this one it is alpha the probability of B 2 probability of C 1 given X 1 into summation over A.

So probability of A probability of X 1 A B 2 and the summation over D probability of D given X 1. So we can write like this. So corresponding to this or maybe equal to this I can write the probability of B 2 probability of C 1 given X 1 into probability of now we have to consider this all the components of the A the vector A the components of the vector A is A 1 A 2 A 3 A 4. So that means the corresponding to the variable A I have four states that the states are A 1 A 2 A 3 A 4 these are the four states. So probability of A 1 and probability of X 1 A 1 B 2 plus probability of A 2 probability of X 1 A 2 B 2 plus probability of A 3 probability of X 1 A 3 B 2 and finally we have to consider A 4 also. So all the states we have to consider A 1 A 2 A 3 A 4 into so this is multiplied and into here this D we have to consider D 1 X 1 because in this case D has two states D 1 and D 2 so this we have to consider.

So here you can see if I consider these two probabilities then values would be equal to 1. So from the table also you can determine this one.

So here if I put all the probabilities values available in the table so alpha 0.

4 0.6 after this 0.25 into 0.7 plus 0.25 so these probabilities are available in the belief network 0.8 plus 0.25 into 0.1 plus 0.25 into 0.3 so this is a bracket close into the last one is because this probability is 1 so it is 1 so is equal to 1. So ultimately we have this value so alpha 0.

114 so I will be getting this. So we can compute the probability of X 1 given C 1 comma

B 2 we can determine. So move to the next slide. So in this case what we have to determine the probability of X 1 given C 1 B 2 that is equal to alpha 0.114 so alpha is the proportionality constant. Similarly we can determine the probability of X 2 given C 1 B 2 that is also equal to alpha 0.066 that also you can determine. So you see we are determining these probabilities so this probability of X 1 C 1 B 2 already we have determined and now we have determined the probability of X 2 given C 1 B 2 that should be equal to 1.

So we have to normalize the probabilities and based on this we have obtained and the value of alpha is equal to 1 by 0.18 that is the constant of proportionality. So alpha we have determined like this because we have to normalize probabilities but not needed necessarily so it is equal to 1. So that means this probability of X 1 given C 1 B 2 is obtained 0.

73 and probability of X 1 given C 1 B 2 probability of X 2 this should be X 2 probability X 2 C 1 B 2 this is equal to 0.27. So these two probabilities we can determine. So the problem already you know what is the problem the problem is classify a fish given that the fish is light, light means C 1 and was caught in South Atlantic that means B 2 and no evidence about what time of the year the fish was caught nor it thickness. So these two information it is not available and based on this we have determined the probabilities. So this is the example so the classification of the fish so this calculation already I have shown here this is the calculation so we can determine this alpha and finally this normalized probabilities we are getting this probability already I have explained this probability is obtained 0.

73 and 0.27. Now this is these are the examples another example for the Bayesian belief network I can give one how to draw the Bayesian belief network. So this is one example the example is the fire diagnosis. So this is a problem the problem is whether there is a fire in a building. So the conditions are receive a noisy report about whether everyone is leaving the building so receive a noisy report about whether everyone is leaving the building because when there is a fire there should be some report and all the people should leave the building. Next one is if everyone is leaving this may have been caused by a fire alarm this is the another condition and if there is a fire alarm it may have it may have been caused by caused by a fire it may be caused by fire or by a tempering.

So all these conditions we are considering and maybe another consideration we can consider if there is a fire there may be smoke. So these are the cases so these are the problem of fire diagnosis and we are considering these conditions. So based on these inputs I can formulate the problem the problem formulation is number one tempering is true when the alarm has been with so tempering is true when the alarm has been tempered with so this one condition we are considering number two condition is the fire is true when there is a fire. So this is the second condition so move to the next slide number three alarm is true when there is an alarm when there is an alarm number four condition we can consider smoke is true when there is a smoke when there is smoke number five living is true that is the living from the building living is true if there are lots of people living the building. So number five is considered and number six report is true because we have to report about the incidents the fire the report is true if the sensor reports that lots of people are living the building.

So these are the things these are the consideration we are considering based on the problem. So what are the variables the variables are fire so what are the variables we can consider the variable is fire tempering alarm smoke living and the report so these are the variables we can consider so fire tempering alarm smoke living and a report. So based on this we can draw the Bayesian belief network so the network is something like this the tempering is I am writing tempering here tempering the fire alarm smoke when there is a fire there may be smoke living and a report. So when there is a tempering the alarm may be on so I am putting this condition when there is a fire there may be alarm when there is a fire the possibility of smoke when the alarm rings people are leaving the building and after this if all the peoples are living then we have a report. So like this we can construct the belief network so based on this I can determine the probabilities so how to determine the probabilities the probability of T, T means the tempering F means the fire F means the fire A means the alarm S is the smoke L is the living R is the report. So we can write like this probability of T tempering is independent into probability of fire that is also independent but this alarm probability of alarm alarm depends on two variables one is the tempering another one is fire into the probability of the smoke smoke depends on the fire so S given F and probability of L living depends on alarm and finally the probability of the report generation based on the information of living.

So I can determine these probabilities so this joint probability I can determine by considering this so this is one example so how to construct a Bayesian belief network. So up till now I discussed the concept of the Bayesian belief network so how to construct a belief network based on the problem I have shown it is a directed graph and you can see we can determine the probabilities this joint probabilities can be determined by considering the conditional probabilities. So for this we are considering the assumption the assumption is the conditional independence and one important point is the same rule of the probability so with the help of the same rule we can decompose the joint probabilities into the conditional probabilities.

So this is the fundamental concept of the Bayesian belief network with the help of this graphical model I can do classifications. So let me stop here today. Thank you.