Artificial Intelligence
Prof. Mausam
Department of Computer Science and Engineering
Indian Institute of Technology - Delhi

Lecture-62
Bayesian Networks: Maximum a-Posterior Learning

Now I give you a coin and I tossed it 5 times and I get 5 heads. What do you think is the probability of heads of this coin? I give you one rupee coin. According to this; let us be intuitive, IA is a lot about intuition and inside and then modelling it mathematical. So I toss this coin. I show you the coin is a one rupee coin are tossed 5 times. I get 5 head how many if you think that probability of heads of this coin is 1?

How many of think probability of heads of this coin is 0.5 at least some hands are being raised other people not listening or uncare. Let us again ask this question. How many of you think the probability of heads 1? How many of think the probability of heads is 0.5? Ok for other people how many of think in the probability of heads somewhere between 0.5 and 1? Many people think very interesting.

How many of you think it is closer to 0.5 than to 1? How many of you think it is closer to 1 than to 0.5, wow there is split first time? Some people really believe that are one rupee coin where you can see both the sides and you are just MS. Dhoni and in always gets a right answer. The probability of head is closer to 1. So, I would have differed with the people who raised their hand last I understand the probability of heads you might think is between 0.5 and 1.

But years of understanding of tossing of coin have told us that it is usually very close to 0.5 if not exactly 0.5 and if you get 5 data points which have just had that will not move you so much that you move towards 1. I would have said that the probabilities probably 0.5 if not slightly more than 0.5. But 5 data points is two smaller numbers for me to make a strong claim the probability of heads is 1. And if I had done this, why would I do this? What am I say? What is the intuition that I am capturing?

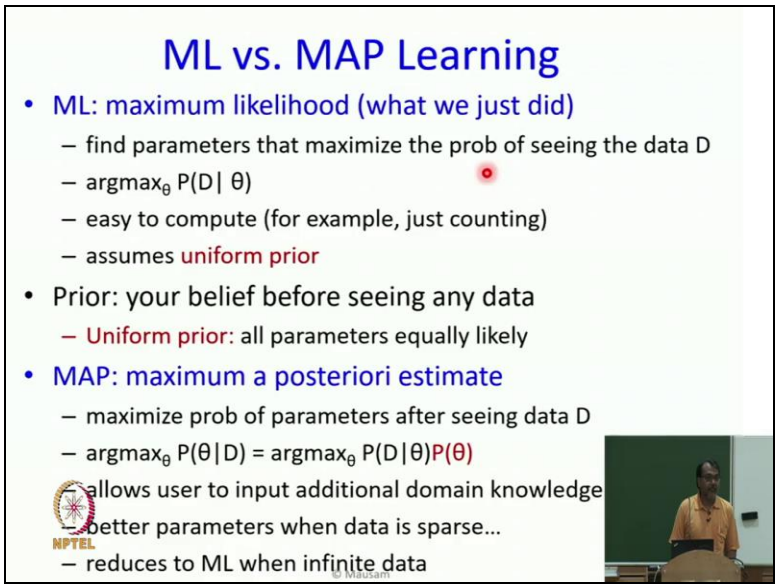
Exactly, what is your name? Yash, Yash says that generally a point of probability is 0.5 that is intuition I am capture. However in maximum likelihood estimation if I did this my probability will end up being 1 and if I did smoothing it will end up being 6 over 7 which is a fairly high number. In some sense maximum likelihood estimation is saying each parameter is equally likely all parameters are equally likely. I have no preference for one parameter versus another parameter.

But in the case of coin example what I am trying to say, is that the parameter which is close to 0.5 is much more likely than the parameter which is close to 1. Now let me flip this on this side. I toss this coin a million times and I get a million head. How many of you think the probability is 1? Come on not hard question. I toss this coin a million times. I can do this, right and I got head a million times. What is the probability that the coin of heads is 1? It is very close to 1 hour = 2or very close to 0.5? Very close to 1 it is Amitabh Bachchan's coin in Sholay.

Just saying, I do not know you are too young to even know Sholay or not I do not know. So a-priory I would assume that the probabilities is 0.5. If I give you no data what so ever I am going to assume the probability is 0.5 because I know that points are probability points. If you give me 1 data points it does not change me much. If you give me 5 data does not change me much and if give me a million data points obviously I will change.

But unfortunately maximum likelihood estimation will not have the characteristic. So, therefore the reason maximum likelihood estimation does not have this characteristics is maximum likelihood estimation says all parameters are equally likely 0.5, 1, 0, 0.9, 0.3 all parameters are equally likely where as I know that the parameter close to 0.5 is much more likely than the parameter that is close to 0 or 1.

(Refer Slide Time: 05:39)



And in order to model this we use what is called the MAP learning of maximum a-posteriori learning. In other words, it is something that uses base rule. That is why there a posterior learning. It is a posterior it has to be a prior and the prior is on the parameter here. So I am saying that instead that instead of optimising probability of be given theta and optimise probability of theta given D I want to find such parameters which has the highest probability.

This is given the data that I have seen, but I can always use base rule. So, when I use base rule I will get R max probability of D given theta which is the maximum likelihood objective function times probability of theta. And this property of theta is our prior on which parameter we prefer and which parameter we do not prefer. And here I will say the probability close to 0.5 is a parameter I will prefer and probability close to 1 is a parameter I will not prefer.

It allows the user to input additional domain knowledge about which parameters are more likely and which parameters are less likely. This is leads to much better parameters when the data is passed if I toss the coin only 5 times. Maximum likelihood shifted very close to 1 maximum a-posteriori will keep it very close to 0.5. Of course, it reduces to maximum likelihood parameter stops to have its weight once I get a large amount of data because then prior has little meaning and its all the likelihood that is sort of overpowering the objective function.

So, in the limit of infinite data MAP tends to ML. So, now and again, I am not getting the details, but if you want to estimate this you will take the ability of theta given D write down the mathematical expression and then using base rule and then take the derivative with respect to theta to get the best result. Is there a lot of stuff that I am hiding under rack that how do I even mention the probability of theta. How do I represent it? Because theta is a continuous parameter.

So then you are represented using some other distribution family of distributions like a Beta Function which is the probability distribution between 0 and 1 and I am not getting in to that. Now this is called the MAP learning.