Transcriber Name: Saji Paul
Statistical Learning for Reliability Analysis
Prof. Monalisa Sarma
Subir Chowdhury School of Quality and Reliability
Indian Institute of Technology, Kharagpur

Lecture - 50
Bayes Classification (Part - I)

(refer time: 00:37)

Hi guys, so today we will talk of Bayesian classifications and we have last class I have already given a basic introduction of statistical based classification method. So, Bayesian classification falls under statistical based classification matter. So, here we will see we will first introduce you to Bayesian classification technique what is that principles and properties we will see our Bayesian classifier.

And further in Bayesian classifier less theory of probability is very much in use we used theory probability took classify using Bayesian technique. So, the already we do we have already learned probability theory but still it is a long time that is in the beginning of this class. So, I thought I will just recapitulate a portion of this theory of probability that will quickly see go into this. So, that is all we will be discussing in this lecture. So, now coming to the introduction.
(refer time: 01:27)

So, as I already this is this figure already, I have shown you in my last lecture. So, this different classification techniques so we will see this is the Bayesian classifier it falls under statistical base matter.
(refer time: 01:41)

So, Bayesian classifier is a concept if it works like a duck, walks like a duck then it is probably a duck. As I told you in my last lecture, we try to classify the dark classify any object based on the relevancy of the some known objects known of some known object means some object with known class level. We try to predict the class of an object based on the relevancy of some objects with known class level.

So, here it is if it works like a duck, quacks like a duck then it is probably a duck. So, this is the concept of a Bayesian classifier.
(refer time: 02:22)

So, just that is its principle it is a statistical classifier it performs probabilistic prediction. That is it is predict class membership problem is a probabilistically it classifies probability of this objects this observation belonging to this classes 0.9. Probability of this object belonging to this classifier is 0.8 so that way it gives a probabilistic prediction of the class membership. Its foundation is on bayes theorem.

I do not know if you can remember bayes theorem it is a very important and interesting theorem in probability theory anyway I will be discussing I will be having a quick recap of that. So, the classes are mutually explosive and exhaustive that is obvious in all the classification techniques the classes are mutually exclusive and exhaustive. And it has this Bayesian classifier it has certain assumptions two assumption mainly.

What are these? The attributes first assumption is the attributes are independent given the classes. So, for a class the attributes that we will consider all the attributes are independent, independent means suppose for any class suppose we have attribute a b c and for that say suppose this class is say X so these are the attribute values I should not say here X anyway forget about this. These are the attributes a b and c are the attribute.
So, these attributes are independent meaning that it is not that if I get an attribute value a 1 that will somehow affect my getting an attribute values b 1, b 1 for attribute b or getting a b 1 value for attribute b it will has somehow have an effect on getting a c 1 value in attribute c it is no not that it is very much independent. That is an assumption in real life actually that is not very much what to say it is not applicable very much this (04:17) independence of attributes.

So, but then the base classifier it considers this as an assumptions the attributes are independent given the class. And this will be more clear when we see some examples and then the attributes have equal contribution to the outcome. So, each attributes these are the attributes a b and c this all these three attributes will have equal contributions. Suppose we have for this for a particular value of a b 1 c 1 we have classified it to be X 1 for a particular value of a b 2 c 2 is classified it to be X 2.

So, this all has equal contribution attribute a attribute b attribute c all three have equal contribution for identifying to a class. It is not that if attribute b will overpower attribute c not like that so, because of this assumption this is called a naive, naive classifier because this assumption are very much unrealistic actually. So, that is why it is called a naive classifier basically because of this assumption.

However, though these assumptions are very much unrealistic but empirically it is proven to be very useful and the empirical it is found that it really gives very good results. Though the assumptions are very naive and it scales very well. When I talk it scales very well means it gives it does not need very much computation it is not very computational intensive even if we use very much a high dimensional data or high volume data.

And both the cases high volume higher dimensional it is not very computationally intensive that is why it is a bayes classifier has also advantage of the scalability.
(refer time: 06:03)

So, a consider an example. So, from this example you will be able to understand better what do I mean by independence what do I mean by equal contribution first example. So, let us consider a set of observations which are recorded in a database there are some observation are there which is recorded in a file whatever we say. So, what data we have in a file? Regarding the arrival of air planes in the route from any airport to New Delhi under certain conditions.

But we have in this file what data we have any arrival of flight to New Delhi airport from maybe from any airport and under certain conditions. Conditions may be weather condition may be day whether it is a weekday whether it is a Sunday holiday whatever it is then condition may be the weather condition, condition may be the season whatever it is under certain condition we have tried to keep we have maintained the data of all the flight arrivals to Delhi airport.
(refer time: 07:10)

So, see here these are the observation which I am talking about. So, like these are the different observations that we have noted that. If it is a weekday means Monday to Friday and the season is spring and when there is no fog and no rain then the flight arrives on time. Then similarly weekday winter no fog slight rain then flies arrives on time. Again, here if we say if it is a holiday winter and high fog slight rain the flight arrives late.

Again, here we see if it is a Saturday spring high fog heavy rain flight cancel. So, these other data which we have kept in a file. So, this is the class level these are the attributes day season for rain these are the attributes feature set whatever you call it attribute set or feature set this is. This is the class level. Now these attributes day season for grain all are independent does the what do I mean by independence being a weak day it does not under any way influence the season obviously.

And being a season it does not anyway influence whether they will be fog or not but that is not true actually that is why I told this assumption that (()) (08:28) assumption is not very realistic. So, season and fog and rain it will it usually there is it we cannot tell this is independent there are some dependency to it. But in Bayesian classifier we consider the (()) (08:38) independent. So, an equal contribution means that the flight arrives on time it has these weekdays this day season for rain all has equal contribution.

It is not that if the day is a weekday, then definitely or irrespective of this condition usually with the flight arrives on time there is no it is not that this weekday will overpower these attributes no all has equal contribution, that is the assumption. As I told you again this is not very realistic, this is how we start with Bayesian classified now. Let us see how we will classify this.
(refer time: 09:17)

So, in this database there are four attributes. What are these attribute? Day whether it is a weekday or it is a Saturday or Sunday or a holiday then season spring season winter season summer season but rainy season than fog with a (()) (09:33) very high fog less fog no fog than rain (09:36) rain heavy rain light rain. So, these are the different attributes and we have total 20 tuples of data 20 observations.

Now different categories are on time late very late cancel these are the different categories. Now we have to form a model were given these values given the values of a weekday. What is the day? It is a sorry given the value of the day given the value of the season given the level of the fog and amount of rain and we should be able to predict whether it is the flight is on time late very later cancel.

So, given this is the knowledge of the data and classes we have to find a most likely classification for any other unseen instance. For example, if it is a weekday winter high fog no rain then what will be its class level. Classification techniques eventually to map this tuple into an accurate class. So, that is the classification task. You will have to it has to map this technique to a accurate class. What are the different classes? Classes are on time late very late and cancelled.
(refer time: 10:56)

So, Bayesian classification methodology first is in many applications the relationship between the attribute set and the class level is non-deterministic which I already mentioned in statistical based method this is one of the characteristics. The relationship between the attribute set and the variable is non-deterministic. We cannot say this is there this attribute is there that means definitely it will fall in this category in this class.

So, that deterministic is not there it is non-deterministic like from this example this is a very good example you cannot say there is a weak day that means deterministically we cannot say since it is a weekday and since the spring season the updated class it will the flight will always be on time we cannot deterministically say. So, this is one of the features of the statistical classification technique.

In other words, a test cannot be classified to a class level with certainty. Since there is no deterministic it is not deterministic, we cannot certainly certain and we cannot predict it with certainty. This will be this class we cannot predict with certainty in such a situation the classification can be achieved probabilistically. The Bayesian classifier is an approach for modelling probabilistic relationship between the attribute set and the class variables.

It is an approach for modelling probabilistic relationship there is a probabilistic relationship between the attribute set and the class variable. More precisely Bayesian classifier use base theorem of probability for classification. In remember bayes theorem do not worry if you do not remember I will have a quick recap. So, before going to discuss Bayesian classifier we should have a quick look at the probability theory and bayes theorem. Because without knowing bayes theorem we cannot proceed with Bayesian classifier.
(refer time: 12:50)

So, quick recap of probability, so if there are n elementary events associated with a random experiment n elementary event associated within random experiment and m of n of them are favourable to an event A, then the probability of happening or occurrence of A is m / n. The probability of happening or occurrence of A is m / n. Now it is irrelevant but do you remember now, what is odd?

We have discussed odd if I would have find out what is the odd of 8 and what would have been the odd of A? Odd of A is m / n - m anyway this is not needed here or this odd concept does not come here but still I have just I just wanted to know whether they forgotten or not anyway. So, this is the concept of probability. Probability of occurrence of A is if m of n are favourable to an event A is the probability of A is m / n.

Some small example what is the probability rolling an even number on a die? Yet a random X when E is rolling a die, A is getting an even number, favourable outcomes for the even number when we rolling A die how many will be even numbers what are the favourable outcomes rolling a die even number is 2, 4 and 6 so there are total three favourable outcomes. When we roll a die how many total outcomes six total outcomes.

So, probability have given getting an even number so favourable outcome is 3 total is 6 so probability of getting an even number is 3 / 6 that is equals to half simple concept of probability.
(refer time: 14:37)

So, we have also seen what is probability of A U B. Probability of A U B is probability of A + probability of B - probability of A ∩ B. Why we do that? Because this is probability of A this is probability B suppose this is A this is B and if we try to find out A U B then if we take probability of A means these whole things get covered and probability of B means these whole things get cover probability of A + probability of B this thing sets we have contacted twice.

So, since we have counted it twice, we will have to subtract it so that is minus of A ∩ B this portion is minus of A ∩ B and now suppose if A ∩ B is 0 so if my A is like this B is like this. This is A this is B then this A ∩ B is 0 then in that case probability have A U B is probability of A + probability of B and under if what case if this will be when both are mutually exclusive.

So, if A and B are mutually exclusive then it will be something of this case this sort. So, then probability if A ∩ B = 0 then probability of A U B will be simple probability of A + probability of B it is not it. So, now if A and B are independent events suppose this is not mutually exclusive suppose this is probability of A U B as this is the formula for us. Now we know that probability of A and probability B are independent events.

So, probability of A and B are independent events then what happens. Probability A ∩ B is nothing but probability of A into probability of B, what is the probability of raining today and probability of me coming to the class. So, this is an example of dependent probability because if it strains, I may not come to the class. So, there is a very bad example I have given. Probability of any independent events now I am not I cannot think of any so immediately anyway.

If A and B are independent events probability of A ∩ B is probability of A and probability of B. Probability of me going to market (16:46) and probability of someone acts also going to market both are independent event. So, in that case probability A ∩ B is probability of A into probability of B. In that case what will be a probability of A U B probability of A + value B - probability of A into probability of B.

We have not come up dependent event dependent will come. Thus, for mutually exclusive event probability of A U B will be probability of A + probability of B mutually exclusive events when this is the case when both they have events cannot happen at the same time.
(refer time: 17:28)

Now if the events are dependent, then the probability is expressed by conditional probability. Probability of A given B given that B has already happened what is the probability that A will happen given that B has already happened what is the probability that A will happen. Now if both A and B are independent if both A and B are independent the probability of X going to market and probability of Y going to market this is very much independent.

So, given that X has gone to market what is the probability that Y will go to market. These are very much independent. So, what is the probability of Y will go to market is probability of Y? So, here probability of A given B = probability of A if it is an independent event. Now if it is a dependent event, dependent event suppose I cannot stand Y at all. So, if Y goes to market if somehow, I come to know that Y goes to market I will try not to go to market.

So, then it is a dependent, what is the probability of A given that probability of B as have happened. So, this is dependent event. So, the events are dependent and the probability is expressed by conditional probability. The probability that A occurs given that B is given that B has occurred is denoted by this. Suppose A and B are two events associated a random experiment the probability of A under a condition of B has already occurred and probability B is not equals to 0.

Conditional probability also we have to see the probability of B is not equals to zero that is given by this, what is this probability of A given B number of events in B which are favourable to A number of events in B which are favourable to A that is nothing but $A \cap B$. Number of events B which are favourable to A this is B this is A number of events in B which are favourable to A is this portion, this B this events it is favourable to A as well.

So, this is $A \cap B$ so, that is probability of $A \cap$ divided by number of events in B so, $P(A \mid B)$ is $P(A \cap B) / P(B)$ this is how we find when these events are dependent.
(refer time: 19:44)

Similarly, so this is we from this P of A given B is P of $A \cap B$ / P of B so, how we can write P of $A \cap B$ P of $A \cap B$ is nothing but P of A given B into probability of B so, that is what this I can write in this way given probability of A is not equals to 0. This I can also write it this way because probability of $A \cap B$ = probability of $B \cap A$ so, I can also write it this way.

Similarly for three events this already we have done all these things just a quick recap probability of $A \cap B \cap C$ probability of A into probability of give B given that A is occurred into probability of C given that A and B has occurred. And if it is independent, it is nothing but the multiplication of all the probabilities. So, if events are mutually exclusive if one happens data cannot happen so probability of A given B is ill then it will be equal to 0.

Toss up a coin if head occur still cannot occurs what is the probability that a given a head occurs what is the probability that a tail will occur that is zero mutually exclusive. So, probability of A given B if A and B are independent probability have X going to market and probability Y going to market both does not know each other. So, it is independent so probability of A give probability that probability have Y going to market given that X is going to market is called your Y going to market.

It is there is no dependency on X so, similarly probability of A given B into probability of B I can

also write probability of B given A into probability of A from this to expression. And this already we have seen.
(refer time: 22:01)

So, basically probability of A given B this already we have seen and probability B ∩ A we can write it in this way. So, this probability A ∩ B I can write it in this way basically this is nothing but the bayes theorem only. We will see there is nothing but the bayes theorem so what we have just done some twisting if initially our probability of A given B is this and we know probability of B ∩ A is probability of A ∩ B is nothing but probability of B ∩ A.

And probability of B ∩ A I can write it in this term then probability of A ∩ A given B, I got this.
(refer time: 23:02)

So, this is remember what is a total probability there is more space here let me use this here so, total when I tell it total probability I have already explained so I am still I am repeating suppose I want to suppose this occurrence this occurrence is suppose this is A. I want to find out probability of this occurrence, this occurrence this round portion I do not know what is the probability of this occurrence.

But I know this occurrence has contribution of three different other events. So, this is on this is B 1 this is B 2 this is B 3 this B 1 has contribution to occurrence of A, B 2 has contribution to occurrence of A, B 3 has contribution to occurrence of A when I know this three then with this knowledge I can I am able to find out the probability of occurrence of A this is nothing but the total probability.

This total probability, probability of occurrence of this total A (()) (23:58) I find out with the knowledge of this disjoint events these are the disjoints events. How do I find it? This probability of B 1 ∩ A + probability of B 2 ∩ this is B 2 probability of B 2 ∩ A probability of B 3 ∩ A that gives me my total A so, when what when I know probability of B 1 ∩ A I know the formula for that I know probability B 2 ∩ A.

I know the formula for that B 3 ∩ A I know the formula of order when I know that then I will be then I found out basically the formula for the total theorem total probability.
(refer time: 24:39)

So, this A is formed by this the events B 1, B 2, B k constitutes a partition of the sample space S these are the partition of a sample this is whole is a sample space S such that probability of B i is not equals to zero then for any even A of S this whole is a sample space any event A I do not know what is the probability of occurrence of A but I know the occurrence of all these events all these are all disjoints events.

Then I can find out this probability of occurrence of A how with the ∑ all this disjoints events. So, that is probability of A is this C ∑ probability of B i ∩ A that means probability of B 1 ∩ A + probability of B 2 ∩ A + probability of B 3 ∩ A plus what how many other interesting disjoints events are there. So, this is my A and I know what is my probability of B 1 ∩ A.

I know what is that this already I have seen when I have sin what is probability of A ∩ B is nothing but problem A given B into probability of B I have seen this so, this simply we will use this, what is probability of B i ∩ A? Probability of A given B i into probability of B i this is nothing but my total probability this is A. These terms I have broken into this term the probability of A given B i into probability of B i.

Why I have written in distance? Because I know probability of B i I cannot write in terms of probability of A because I do not know probability of A I know only probability of A so, I will because if I take this form that B i ∩ A is same as probability of A ∩ B i then this would have been my different end, I would have needed a term here probability of A but I do not know what is probability of A so I will not be using this.

So, this is my this expression I have written it in this form probability of A given B i into probability of B i so this is my probability of A it is total probability. This is theorem total probability now from here comes the bayes theorem. What is bayes theorem here? Now given that I A has happened given that A is happen now I know what is A given that A is happen what is the probability that it is from B 3 what is the probability that it belongs to B 3 class.

Now I can see the class concept that is coming given that these are attributes what is that it belongs to a B 3 class. Given this these attributes what is the probability that it belongs to B 5 class. These are the attribute, what is the probability that it belongs to B 4 class my bayes what is my bayes theorem, bayes theorem is that given that this A has occurred given that my A has occurred what is the probability that is due to the event B 3.

Event that probability that event A has occurred what is the probability that it is due to event B 3 what is the probability that it is due to event B 1 that is bayes theorem and this bayes theorem we can very easily use in classification how given that this are my attribute because this is A as occurred means these are my attribute (28:36) basically. What is the probability that it belongs to this class because A is the combination of all these classes.

The whatever attributes we have all the classes hidden here so this particular set of attributes I

got what is the probability that it belongs to this class, what is the probability that it belongs to B 1 class. That is how we have used here bayes theorem.
(refer time: 29:02)

So, by the laws of total probability this is what already what we have discussed.
(refer time: 29:13)

Now this is the bayes theorem if E 1 to E n be n mutually exclusive and exhaustive event associated random exponent. If A is in any event which occurs with E 1 or E 2 or E 3 then what is the probability of E i given A has occurred. What is the probability that it belongs to class A I given that A has occurred given that these attributes are there given this is my observation. So, this is nothing but this is the denominator is nothing but the total probability total probability of A.

Denominator is the total probability of A, what is the numerator? Numerator is the probability of that class into probability of A given probability of E i. What is the probability of A given that E i has occurred into probability of E i divided by the total probability will give me what is the probability of E i given A. So, one more thing here I forgotten now this thing is the this thing is called a posterior probability.

What is the probability of A given that E i has occurred? Because I want to find out probability of E i given A. Now this probability of A given as I this is the posterior probability and this is called a prior probability.
(refer time: 30:59)

See here the table shows that event A has two outcomes so event Y has two outcomes A and B basically we can say that it is these are the different classes. Namely A and B which is dependent on another event X with various outcome x 1, x 2, x 3 maybe this may be the X is one attribute it has just one attribute set and this is the class level. Suppose we do not have any information of an event A then for a given sample space we can calculate probability of Y = A.

How many occurrence of A by total occurrence the five version this is the probability of Y = A this is called prior probability this is called probability of Y is equal to this is called prior probability. Here in this case this is called prior probability. So, now suppose we want to calculate what is the probability of X given Y = A this we will be able to calculate is not it what is the probability of X given Y = A given Y is what is the probability of X = x 2 from here we saw that from for Y = A x 2 how many is there this is here this is another one.

So, it is 2 / 5 total Y occurrence is 5 so this 2 / 5. So, this is called posterior probability we will be using this posterior probability and prior probability concept while we will be solving Bayesian classifier.

(refer time: 32:33)

In the next lecture we will continue our discussion on page and classify. So, in this lecture we learned a scope and application of agent classification the properties and its applicability we have also done a quick recapitulation of Bayes theorem and the next lecture we will go to the details of Bayesian classifications.
(refer time: 32:50)

Thank you, guys.