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-12

Lecture-46

Greetings everyone. Welcome to the MOOCs course of machine learning and deep learning fundamentals and applications. I am Bivek Goswami the teaching assistant for this course. I am a research scholar in the EEE department of IIT Guwahati under the supervision of Professor M. K. Bhuyan Today we will start our problem solving session on naive Bayes classifier linear regression for univariate and multivariate case and we will be looking into maximum likelihood estimation and support vector machine in this class.

So let us start today's session and solve the problems here. So the first question we are having is that find the linear regression equation for the following sets of data. So we are given a dependent variable and an independent variable. So x is our independent variable and y is the dependent variable and we know we can represent the equation of the dependent variable like this where a represents the slope of the line and b represent the intercept of the line.

Now there are formulas through which we can find the a and b. So for a, we know that we have the formula of where your n is the total number of samples. Similarly we have a formula for b. Now to calculate this a and b we have to first make a table for this xy x^2 and then find the summation. So we will find that out first.

So we will make a table for x and y. So the x values we have is 2, 4, 6, 8. So the x^2 values we will have is similarly the xy values we will have is 4, 28, 30 and 80. So we will find the sum of all of them. This will sum up to 20, this will sum up to 25, 120 and 144.

So this table we are having right now. Now in this formula we will put these values and we will see what a and b values we get. So putting the values in the formula. So the formula of a we are having is. So we know that we have the sum values for sum of x, this is summation of x, this is summation of y, this is summation of x^2 and this is summation of xy.

These values we are having. So if we put these values here and our N here will be 4 because we have 4 number of samples. So in this formula if we put the values, so we will put N equal to 4. We have seen the summation of xy is 144. So what values we will get is.

Similarly for B we will find out. So the equation of this line will be, so this will be the slope of the line and 1.5 will be the intercept of the line. So this will be basically our regression line that we will be getting here. So let us move to the next problem.

So this is basically a univariate case where we have only one independent variable or a feature and one dependent variable. Now we will be looking into a problem where we have two independent variable and one dependent variable. So let us say it is X1, it is X2 and this is Y. So this value we have to predict from the data of these two values, this and this, X1 and X2. So similarly we can represent this one like this where A and B will be the slopes and again C will be the intercept of the plane.

Here we want to have a line and we will have a plane because we have two independent variables and one dependent variable. So we will not be having a line and rather we will be having a plane. So let us see how we can solve this. So as we have seen we can represent it like this. So we can write like say Y can be written as AX + B and this can be represented like this.

If Y have values, so we will be having X1 and X2 values say X11, X12, X21, X31, X41 will be having a bias. This we have already seen in the previous videos. So I will tell you how can we find this term. So this can also be written like this. So if we include the bias also here, so we will be having a bias term say A0.

So these are A0 values. A1 will be the first intercept and second one will be the second intercept and we can find this like this. So let us see how we will do that. So we will be using a method known as pseudo inverse to find out. So in pseudo inverse how we can find out the coefficient is using this method.

So we have X transpose of X, we will find out the inverse of that, multiply with the X transpose again and then multiply this whole thing with Y. So for this A we will be having 3 values. We will be getting A0, A1 and A2. So our line will be represented like this, similar to this one. So let us start finding out.

So we have our X, see 1, 2, 3, 4, 4, 5, 8, 2 and this is the bias term that we have to include as I have already explained you. Now we have our Y here. So let us find out

$X^T X$. So if we multiply these two matrix we will be getting these values. Now what will be our next step? So we will find this inverse, inverse of this $X^T X$.

So if we find the inverse of this matrix the values we will be getting is this one. Please check the calculations by yourself once. So this is the inverse that we have got. Now again what we have to do? We have to find out this $(X^T X)^{-1} X^T$.

So we will find it out. Say $(X^T X)^{-1} X^T$. And we will obtain a matrix again which will be of this shape. So we will find out this $(X^T X)^{-1}$. Now again what is our next step to find this one? We will have the whole value of this multiplied with Y. So once we do that we will be getting this value.

So this will be our A0, this is our A1 and this is our A2. So we can represent this line like this. This with the regression plane that will fit this multivariate case. So this was the question that find the linear regression equation for the following data which we have obtained here. As we have obtained here the same linear regression variable but it was for a single variable and this is for a multivariable case.

So that is all we are dealing with linear regression right now. Next we will move to another problem of Bayesian classification. So here we will be looking into the Naive Bayes classifier. So let us look into the question. So it is told to estimate the conditional probabilities of each attribute.

So these are the attributes or the features and this is the value that needs to be predicted. So says this is a species. So M is one species and H is another species, has species class. So through the data in the table we have to predict for this new instance and before that we have to find the conditional probability of each of the attribute. So let us first find out the priors.

So what priors are we can see what is the class that we need to predict. So what is the probability of M species. So it will be we have total 8 number of samples of where 4 are coming to be M. So it will be 4 by 8. Similarly for each species we will be having.

So we have obtained our prior. Now let us start with color. So in making a table we have color. What are the colors available here? White and green right. And we have two species H and M. So let us see how many M are there total 4 of which how many are white so 1 and 2, 2 out of 4.

Similarly how many H are white so we have total 4 number of H and how many of them are green sorry white it is 3. So similarly for green also we can write and we can mix similar table for legs as well. So in legs also they can either have 2 leg or 3 leg. And again we have two species M and H. So how many M have 2 legs it is 1 out of 4 and how many H have 2 legs it is 4 out of 4 all the H have only 2 legs.

Similarly this we can make again we will make for smelly. You can pause the video and make this table yourself as well. And you can verify with my one. So I will tell you what now we have seen that we have found the conditional probabilities. But what is actually the conditional probability? So the conditional probability is given an attribute we have to look at the probability of an attribute given a species.

So this is what a conditional probability is. And you can see in this 4 table we have calculated this. So that is what you will see that every class every species when we sum up we will get 1. So this probabilities will sum up to 1 not this ones because these are not conditional probabilities. When we sum this up we will get 1 because we always know that the probability sums up to 1 and this is the conditional probability.

So even conditional probabilities for a same species will sum up to 1. So for this individual species say M or H the probability in every case is summing up to 1. So I hope that much is understood. So now we have found out the first one the conditional probability. Now in the second part of the questions we are asked to estimate the probability for this new instance.

So we know by Bayes theorem given the species is M and this new instance what we can write Similarly, So, you already know that this is the normalizing term and it would not actually affect your classification decision. This we have already seen in the previous classes. So for M given new instance we will just be left with this formula. Similarly Now what was your new instance that color is green have 2 lakhs height is tall and it is not smelly. So looking at this table we will find out we know the conditional probability we know the priors so we can easily find out that.

So let us see when the color is green and it is M when the color is green and it is M we have a probability of 2 by 4 this one. So it is 2 by 4 again legs 2 and it is M legs 2 it is M 1 by 4 again height tall and it is M it is 3 by 4 and smelly no and M it is 1 by 4 and we have to multiply with the prior of M. So the prior probability of M is 1 by 2. So if we calculate this, this value will come up to be similarly we will find for H. So when it is H and green so for H and green will have 1 by 4.

Similarly for legs 2 and green will have a probability of 4 by 4 for smelly and H for not smelly and H will have 3 by 4 and for tall and H will have 2 by 4. So coupling this probability together and multiplying with the prior we get the value of now we can see that probability of H given new instance is greater than probability of M giving given new instance. So in maximum likelihood estimation sorry in Bayesian classification we know that we will always favor the class which has the larger value of the posterior probability. So in this class this new instance will belong to the species class H.

So you understood this problem. Now look into another problem using maximum likelihood estimation. So what we have here is that an unfair coin is flipped 100 times and 61 times header observed. An unfair coin is flipped 100 times and 61 times heads are observed. So what is the maximum likelihood estimation when nothing is previously known about the coin.

So let us look into this problem. Looking into this that an unfair coin is flipped 100 times and 61 times head is observed we can easily identify that this is a binomial distribution where we will have the probability of occurrence as P and Q as the probability of non-occurrence or say failure which will be 1 by P. So this is the probability of success, this is the probability of failure. So we can write the PMF of failure say we tell the probability of success when head occurs and failure when head does not occur. So we know 61 times header occur and we do not know the success. So this is the PMF of binomial distribution that we know of.

This I hope everyone knows. So now to maximize this what is the first thing that comes in our mind that we will differentiate it with respect to P because we have to find the maximum this P for which the likelihood would be maximum. So this is the likelihood function that we have drawn and now what we will do is for this parameter P we will see what is the maximum criteria that we can obtain. So going by the product rule of differentiation we will find out that this will be the differentiation that we are getting. Now to find the maximum or the saddle point what we will do is we will equate it to 0.

So let us see what we get. So we know that this cannot be 0. So what we will be having is and if we solve this we will be getting three conditions for P. So P can take any one of these three values. Now we will find that for H equal to 61 and P equal to 0 what would be the likelihood. We already know this formula we have to just put the values of P equal to 0 61 by 100 and 1 in the formula.

So for P equal to 0 the likelihood would be 0. Again for P equal to 1 the likelihood would be 0 because this value will end up to be 0. So now what about for P equal to 61 by 100. Let us find out. So it will give us a positive value.

It does not matter what value it will give us. We know that it would not be equal to 0 ever. So it will give us a positive value. Since of these three this is the highest. So for P equal to 61 by 100 maximum likelihood is achieved and that was the question what is the maximum likelihood estimation and nothing is previously known. So we know for this parameter P will have the maximum likelihood.

So I hope this question is clear as well. Now we will be looking into a problem of from support vector machine. We know that support vector machine classifies samples into classes by drawing a decision boundary considering the support vectors. So we are given this four points belonging to positive class and this four points belonging to negative class. So positively and negatively level data are given. So let us denote this positive class by +1 and let us denote this negative class by -1.

So both belongs to R2. So it is a plane. So I have plotted these points here. From this visualization of the point we can easily guess this. This would be the support vectors here. Because beyond this point there are no, before this point there are no points which actually helps in the or affects the decision.

So this can be considered to be the support vectors. So our support vectors we will denote it by S1 would be another support vector S2 we will be denoting with S2 and S3 we will be denoting with. So from this one this plot we know that S1 belongs to plus 1 plus S2 and S3 belong to minus 1. So that much we are clear till now. Now so these are our support vectors S1, S2, S3. So these support vectors using this support vectors we can find the characteristic equation and the characteristic equation would be like this.

This is already been explained in the class how these characteristic equations are formed. So we know what S1 and S2 and S3 are putting those in this equation what we will get is so this is equation 1, equation 2, equation 3. So equation 1 I will show for one equation and then I will generalize the issue. So A1, what is S1? So this is basically a dot product that we are looking at.

What is S2? 3, 1. Solving this. So this is a dot product. So similarly for equation 2 we will be obtaining. So these three characteristic equation we have obtained. Now solve as we know these are basically series of linear equation and if we solve all of them we will be getting A1, A2 and A3 here. So our A1 will be so this will be our A1, A2 and A3 that will be getting solving the simultaneous equations. Now we know the weight vector of support vector machine because support vector machine can also be characterized like this.

This is the weight and this is the bias. So the weight vectors can be represented like so what we have here because we will be adding the bias term. So we are having 3. These are the bias terms that we have added. So this will sum up to we have added bias here.

So this will represent our B. So we will have B -2 and W will have as. So we can represent the equation like. Now since we have represented our equation like this so this bias term has a negative sign. So it will denote that if we draw an axis then it will be represented in the first quadrant and this W tells us about the line.

So we can see W is (1, 0). So the X coordinate is Y1 and Y coordinate is 0. So it tells us basically the line will be parallel to the Y axis because the Y coordinate is 0. Now we will see that we have 2 as the bias. So basically the line should be at this position. So this is our decision boundary that we have found out.

On our calculations of W as and our bias as – 2. So this is the decision boundary hyperplane that can classify these points into two classes negative and positive. So that was all about today's session. In today's session so basically what we have seen first we started with linear regression where we have seen linear regression for a univariate case. So it had only one dependent variable and only one independent variable. Then we have seen same linear regression for multivariate case where two independent variables were there and one dependent variable and we have seen how to solve for those cases as well.

Then we have solved a classification problem using a Naive Bayes classifier or a Bayesian classification technique where we have found out the conditional probabilities and for a new instance we have shown the class or which class it belongs to. Then we have again seen a problem related to maximum likelihood estimation. We have come to know what is a likelihood and how to estimate that and how for any feature maximum likelihood could be estimated. Again we have seen another classification problem using support vector machine where we have found out what are the support vectors for the data given.

We have found out the characteristic equation. Then we have found out the decision boundary or the hyper plane that is actually separating the classes. So that is all about today's session. More problems will be discussed on the other topics in the coming sessions. So that will be all. Thank you.