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-3

Lecture-13

Welcome to NPTEL MOOCs course on machine learning and deep learning fundamentals and applications. In my last class, I discussed the concept of parameter estimation. I explained two techniques, two popular techniques, one is the maximum likelihood estimation and another one is the Bayesian estimation. Today I am going to discuss the concept of non-parametric techniques, that is how to estimate density. So in the non-parametric techniques, I can determine density, the density of $P(X/\omega_i)$, that is the class conditional density I can determine or directly I can determine the posterior density, that is the $P(\omega_i/X)$,. In the non-parametric techniques, I will be explaining two popular techniques, one is the Parzen window technique and another one is the k nearest neighbor technique.

So today I will explain the basic concept of non-parametric techniques. So let us see what is the non-parametric technique for density estimation. So today I am going to discuss the concept of non-parametric method and that is actually the density estimation. So that means in case of the non-parametric methods, we can determine the density, but in case of the parametric estimation technique that already I have explained, the density from is available, but we have to determine the values of the parameters.

In case of the non-parametric methods, we can estimate the density. So that means what is the density? So directly we can estimate density of this, the class conditional density that is the likelihood, or directly we can determine the density of the probability of omega j given X. So directly we can determine this. This is the concept of the density estimation. That means the density from is not known, but we can estimate the density.

So how to determine the density? So let us explain this concept. So what is the probability that a vector, vector is the Feature vector, X falls in the region R. So region R is considered. And what is the probability that a vector X falls in a region R? So that probability I can write like this. The probability is the region is R.

So the X bar and suppose this is equation number one. So what is the meaning of this equation? So this probability P is the smooth, or I can say it may be averaged, averaged version of the density function. So this probability is the average version of the density function. So based on this I am writing the equation, equation number one.

So we are considering n number of samples, n samples we are considering, n number of samples we are considering.

The samples are suppose X1, X2, these are the samples and these samples are in the region, particular region suppose R. So suppose we are considering any arbitrary region, the region is R and in this region I am considering these are the samples X1, X2. So these are the samples in the within the region R. So these are the samples in the region R. So now I want to determine the probability that K points falls in the region R.

$$P = \int \rho(x')dx'$$

$$P = \int \rho(x')dx' \approx \rho(\underline{x})V$$

So total number of samples we are considering n and that means in this case what we are considering, so only K number of points out of n, so total number of points is n. So the probability that K number of points fall in the region, the region is R. So that I am determining. So this is actually the binomial distribution.

This is a binomial distribution we are considering. So I can say this is the equation number 2. So let us move to the next slide. So from the previous slide we obtained that the probability that K number of points fall in a region R is given by, this is the binomial distribution nK, probability of K, 1 minus P, n minus K. So this was the equation number 2.

$$P_K = (n\ K\ )P^K((1 - P)^{n-K})$$

$$E(K) = nP$$

That is the binomial distribution for K points very sharply present in the mean. So we are considering the binomial distribution and in this case we are getting the K number of peaks about the mean. So this expected value of K we can determine like this from the equation number 3.

Now let us consider the maximum likelihood estimation, ML estimation, maximum likelihood estimation of P is equal to theta. So theta already I told you it is a parameter vector. So maximum likelihood estimation of P is equal to theta. So what we have to consider? We have to maximize the likelihood function. The likelihood function is probability of K given theta.

$$P(\underline{x}) \approx \frac{\frac{K}{n}}{V}$$

So maximum likelihood estimation of P is equal to theta we can determine. And what we are considering? We are maximizing the likelihood function to maximize the probability of K given theta and we are estimating this value.

So estimated value of theta, the theta hat is equal to K divided by n that is equal to P. So this K by n because we are getting the ratio K by n that is a good estimate for the probability P and the probability density function, probability density function, this P. So you can see that what is the probability? The probability we are getting that K divided by n that is the probability. So what is n? n is the total number of samples. What is K? So number of sample points falling within the region, the region is R.

So n is the total number of samples. I am repeating this. So total number of sample we are considering is equal to n and K is nothing but the number of sample points within the region R. So that is the K. So if I compute the K divided by n that is nothing but the probability, the probability we are getting.

So this density is continuous and the region R is very small, R is so small so that this the density does not vary significantly within it. So that means we are considering the region is very very small and so the density does not vary significantly within the region. So that is why I can write the probability P and region R is considered that is the X dash that is approximately equal to because this density is almost constant and I am getting the volume. So what you can see in this equation, you can see this density is constant. So I am taking it out from the integration and if you see the remaining one that is the region R we are considering and dx we are considering that is nothing but the volume.

So I can write this X, X is a sample point, X is a point within R within the region R and V is the volume enclosed by R enclosed by the region R. So we are considering V is the volume enclosed by the region R. So in the equation number 4 that already I have explained that the density I can consider as constant because the region is very small. So the density does not vary significantly within the region. So I can consider it as a constant.

So this is a constant and after this if you see this integration so dx dash that is nothing but the volume. So this is the volume. So we are getting the equation number 4. So move to the next slide. So now we can combine the combining equation 1, 3 and 4.

Combining equation 1, I think I can write equation 1 again 3 and 4. So what is the equation number 1? So already I have defined the equation number 1. What is the equation number 1? The equation number 1 is I can write this is a probability region R. This was the equation number 1. After this the equation number 3, the equation number 3 was that expected value of K that is equal to NP that is equation number 3.

So this was the equation number 1. This is the equation number 3. And what is the equation number 4? The equation number 4 is equal to region it is considered so dx dash and that is approximately equal to the density into volume. So this is equation number 4. So considering 1, 3 and 4 so all these equations we are considering.

So we can determine the density. The density is nothing but or I can say it is approximately equal to K divided by N and divided by V. So we have this. So we are getting this expression. This is the expression for the density.

Now this is the important expression for the density.

The density is nothing but the K divided by N that is the K by N ratio divided by volume. Now let us consider the equation number 4. What is the justification of the equation number 4? The equation number 4 already we have explained that equation number 4 the region we have considered is a very small region and that is why the density does not change significantly. So we obtain this expression.

So this expression we obtain in equation number 4.

This is the equation number 4. So what is the justification of the equation number 4? This density is continuous and the region already I told you the region is very very small. The region R or I can say the region R is so small that this density does not vary significantly. Within the region R within the region R. So that is the concept already I have explained. So now because of this condition this density is constant

and so it is a constant so it is not a part of the summation.

So we can take it out from the integration. So it is not a part of the summation. So this expression 4 I can write like this. So this I can consider like this. This density and suppose we are considering the function Mu R.

So what is Mu R now in this case? This Mu R it may be a surface in the Euclidean space R square. So it may be a surface in the Euclidean space R square the two dimensional space. So if I consider a three dimensional space it will be a volume in the Euclidean space R cube that is the three dimensional space. And if I consider the high dimensional it will be a hyper volume in the Euclidean space R to the power n. So high dimensional space we are considering and corresponding to this Mu R will be a hyper volume for the Euclidean space R to the power n.

So we have these cases. So if I consider two dimensional case then it will be a surface. If I consider a three dimensional space then it will be a volume and if I consider a high dimensional space then it will be a hyper volume. So moving to the next slide. So this density is constant.

So that is I can write this density at X is equal to density at X dash that is constant.

So I can write like this because the density is constant because the region is very very small and this is volume. If I consider that Euclidean space R cube R to the power 3. So three dimensional case it will be a volume. So that is why from this you can see the density is given by or approximately it is given by the ratio K by n divided by V.

So this is the estimate for the density. So this is the explanation of the equation number four and from this we can determine the density. So that already I have defined. So this is the expression for the unknown density. Now let us see the conditions for the convergence. So what are the conditions for the convergence because I have to estimate the density.

So one important point is the conditions for convergence. So this density I can obtain it is obtained only if this volume V approaches zero. So density is obtained and that is the density is obtained only the volume is very very small. So that is the condition. So that means mathematically I can write the limit the density and volume tends to zero and suppose volume is very very small.

So it cannot enclose any number of samples. And then in this case if I consider n is fixed suppose if I consider if n is fixed this condition we are considering if n is fixed is an important condition then if the volume is very very small then what will happen it cannot enclose any number of samples. So that is why k is equal to zero. So corresponding to this the density will be the estimate of the density will be zero. So I am repeating this case what we are considering this the number of sample that is n is fixed.

If n is fixed and the volume is very very small that condition already we have explained the volume should be very very small.

So that is why it cannot enclose any number of samples. So that is why the k is equal to zero. So k is equal to zero means the meaning is the k is equal to zero meaning that is no samples are included in region R. So this case I can consider as the uninteresting case because in this case the density is equal to zero. So I can say it is an uninteresting case uninteresting case.

So the no samples are included in the region so that is why k is equal to zero and the estimate of the density is equal to zero. So let us move to the next slide. So another condition is suppose suppose again if I consider limit V tends to zero the volume is very very small but k is not equal to zero the density the estimate of the density will be infinite. That means we are considering n is fixed n is fixed the number of samples are fixed. In this case also we are considering n is fixed corresponding to this k is not equal to zero that means corresponding to this the estimate diverges.

So corresponding to this I can write estimate estimate of the density estimate diverges. So I can say it is not nothing but the uninteresting case. So if I consider the k is not equal to zero then also the density will be infinite. So it will be very high. So these two conditions I am considering so that means if I consider n is fixed and volume is very very small.

So for these two cases k is equal to zero and k is not equal to zero you can see in one case I am getting zero the estimate of the density and in another case when k is not equal to zero the estimate of the density is infinite very high. So that is why to consider these two cases what we have to consider one condition is volume V needs to volume needs to approach zero. So that means the volume should be very very small. And what is actually the variance the variance is the ratio this ratio of k divided by n this is called the variance is the ratio k by n. And now we are considering unlimited number of samples many many samples infinite number of samples unlimited number of samples.

So that means n is very very high unlimited number of samples. So earlier we considered the n is fixed. So corresponding to n is fixed we have these two conditions in one condition the density the estimated density is equal to zero and in another case that density is very very high infinity. So these are uninteresting case. So that is why what we are considering the volume should be very very small and we are considering unlimited number of samples unlimited number of samples we are considering and the variance is the ratio of k divided by n.

So what we are doing now form I am forming sequence of regions sequence of regions the regions are R1 suppose some of the regions R1 R2 like this containing containing the samples that sample is suppose x. So we are considering the large number of samples we are considering and we are considering the sequence of regions and containing the sample the sample is x that is a Feature vector. So in earlier case what we have considered we are considering a region region is suppose the volume is V and n is the total number of samples. So these are sample points. So out of all these sample points we are considering that k number of sample points within this particular volume that we have considered.

So n is basically the n is the total number of samples and V is a volume I am considering and k is the number of samples within this particular volume that we have considered. Now in this case we are considering the sequence of regions R1, R2 like this containing the samples. So the first region it contains only one sample.

Second region contains two samples and so on.

So like this we are considering number of regions. The first region contains one sample. The second region contains two number of samples and like this we are considering the number of regions. So let us move to the next slide.

So now we are considering Vn. Vn is the volume of the region Rn. So we are considering the regions R1, R2 like this. So corresponding to a region Rn the volume is Vn and corresponding to this Kn number of samples falling within the region Rn. So what we are considering the Vn is the volume corresponding to the region Rn and Kn is the number of samples falling in the region Rn. So that is we are considering. So I am repeating this corresponding to the region Rn the volume is Vn and

if I consider three dimensional case it will be a simple volume.

If I consider the high dimensional case then in this case it will be hyper volume. So Kn number of samples within the region Rn. So that we are considering. So this density Nth density that is the be the Nth estimate for the actual density.

Actual density is this. So we are estimating that is the Nth estimate for the actual density. So from the previous equation so from the previous equation this Nth estimate of the density actual density will be Kn divided by N and divided by Vn. So this already I have explained. So this is the estimate for the density and now we are considering the Nth estimate of the actual density.

So this is the estimate for the density. Now we have to see the conditions so that this Nth estimate of the density converges to the actual density. So that means I can write necessary conditions should apply should apply so that the Nth estimate of the density. So it is the Nth estimate of the density converges to the actual density actual density is this the actual density. So we have to see what are the conditions so that the Nth estimate of the density converges to the actual density that is the actual density is this. So now what are the conditions the condition is number one condition limit Vn is equal to 0 and tends to infinity that means what is the meaning of this.

So we are considering large number of samples samples and tends to infinity large number of samples and the volume is very very small. So if I consider the large number of samples the volume may be very very small. So that is the first condition for the convergence. The second condition is limit KN N tends to infinity and that is equal to infinity. So the meaning is if the volume is very very small there is a possibility that this small volume even can enclose the infinite number of samples because we are considering infinite number of samples.

So the volume is very very small. So this is a volume is very very small then this small volume can also include some of the samples the high number of samples because we are considering a very large number of samples. So large number of samples we are considering these are the samples so very large number of samples we are considering and if the volume is very very small even in this case it can enclose the high number of samples points. So I am repeating this the volume is very very small and number of samples the n is very very high. So even in this case also this small volume can enclose many many samples the infinite number of samples that is the explanation of the equation number 2 the number 2 equation means the condition number 2 and number 3 the Kn is very very high. So I can say the limit Kn divided by n n tends to infinity n is very very high then it will be approaching to 0 because n is very high as compared to Kn.

So that is why this limit Kn divided by n that should be equal to 0. So these are the important conditions for the convergence that is the convergence of the nth density to the actual density. So I am repeating these conditions are very important the first condition you can see I am considering the number of samples are very very high that is n is number of samples are very very high and in this case the volume may be very very small the Vn is equal to 0 the volume may be very very small that is the condition number 1. Condition number 2 N is very very high the number of samples are very very high then in this case also this small volume I can expect that it can enclose many many samples the infinite number of samples. So that is why I am writing limit n tends to infinity Kn is equal to infinity that means even this small volume can enclose the many many samples the Kn number of samples that is the condition number 2. In condition number 3 the n is very very high and tends to infinity the variance the variance is nothing but the Kn divided by n that will be equal to 0 that approaches 0 because as compared to Kn the n is very very high.

So that is why the variance is the ratio Kn divided by n it approaches 0. So these are the important conditions for the convergence. Now let us consider two approaches for estimation of the density. So already I told you two popular techniques one is called the

Parzen window technique and another one is the k nearest neighbor techniques.

So now move to the next slide. So what are the two approaches. So two approaches are there two popular approaches. One is the Parzen window technique Parzen window. So in this approach what is the case shrink what is the Parzen window shrink and initial region where the volume Vn is equal to 1 by root n. So n is already I told you n is nothing but the number of samples and Vn is the volume corresponding to the region Rn. So in this case what we are considering that is the volume is fixed the volume is fixed we are fixing the volume.

So suppose this is the volume the volume Vn corresponding to the region Rn and these are the sample that is a Kn number of samples we are considering. So Kn number of samples and what is n n means the total number of samples. So these are the samples. So n means the total number of samples. So the volume is fixed and we are determining the ratio because for the estimation of the density you can see we have to determine Kn divided by n divided by Vn.

So this ratio Kn divided by n we can determine and from this expression the expression is this expression. So from this expression I can determine the density. The volume is fixed what I have to determine I have to count the number of sample points within this particular volume that means I have to count Kn. So count Kn. So how many samples within this particular volume we have to determine and based on this we can determine the density and this density we can determine.

So this is the fundamental concept of the Parzen window technique. Number 2 technique number 2 that is called a Kn technique. Kn nearest neighbor technique. So that also I will be explaining the Kn nearest neighbor Kn technique. So in this case the Kn is fixed that is Kn is suppose root over n that is actually data dependent because n is the number of samples.

So it is I can say it is data dependent. So what I have to consider specify Kn as a function of function of n. So already I have explained that is Kn is equal to root n. Now what I have to consider now in this case I am fixing that Kn. Kn is fixed I can say that out of this that means the Kn is fixed.

The volume Vn is grown the volume is increased until it encloses Kn neighbors of the X. X is the sample. So this is actually the concept is it is called this concept is Kn nearest neighbor estimation method. So in my next slide I will be explaining what is the Kn method. So in the Kn method the Kn nearest neighbor method what we are considering Kn method. The Kn is fixed this is fixed and we are defining Kn is suppose root n that is actually the data dependent.

So for example in this case suppose Kn is suppose 5. So we have to grow the region or we have to increase the region so that it encloses the 5 number of samples. So suppose we are considering a region this region is suppose R this region is R1. So this region R1 encloses only the one number of sample. After this I am growing the region that means I am increasing the region.

So suppose the region is now R2 this is the R2. So it encloses suppose the two number of samples. So like this I have to grow the region and like this we have to grow the region and corresponding to this suppose the region is suppose R5. So this region R5 it encloses the region R5 encloses the 5 number of samples. This 5 is actually what is the 5. 5 is nothing but Kn is equal to 5 we are considering Kn is equal to 5.

So this region 5 considered or this region 5 encloses the 5 number of samples. So we have to increase the region we have to grow the region. So corresponding to this region R5 we can determine we know the volume the volume is Vn. So from this information we can determine the density the density is nothing but the nth density is nothing but the Kn divided by n divided by Vn. So you can see in the Knn method what we are considering the Kn is fixed and what we are determining we are determining the volume.

So that it encloses the Kn number of samples. So we are growing the region so that it encloses the Kn number of samples and corresponding to this I can determine the volume and from that volume information and from this Kn information and we know what is n the total number of samples we can determine the density. So you can estimate the density. So this is the Kn nearest neighbor technique. In case of the Parzen window technique what we are considering the volume is fixed in the Parzen window technique already I have explained.

So Parzen window technique so the volume we have considered a fixed volume is also data dependent. So I can say it is fixed and what we have to determine we have to determine how many samples within this particular volume. So we have to determine the Kn and based on this we can determine the density we can determine. So you can see the fundamental difference between the Kn nearest neighbor technique and the Parzen window technique. So in this class I explain the fundamental concept of the non-parametric estimation. So in the non-parametric estimation I have explained two approaches one is the Parzen window technique and another one is the Kn nearest neighbor technique and what is the fundamental difference between these two techniques.

In case of the Parzen window technique the volume is fixed and we have to determine the number of samples within this particular volume that is the Parzen window technique. In case of the Kn nearest neighbor technique what we are considering the Kn is fixed that is the data dependent. So for example we considered Kn may be equal to root over n that is the data dependent the Kn is fixed and we have to find the volume that means we have to grow the region so that it encloses the Kn number of samples. And corresponding to this we can determine the volume of the region and from that information we can determine the density.

So these are two techniques two approaches for estimating the density. So in my next class I will be explaining the concept of the Parzen window. So let me stop here today. Thank you.