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-15

Welcome to NPTEL MOOCs course on machine learning and deep learning fundamentals and applications. In my last class, I discussed the fundamental concept of Parzen window. So in the Parzen window, I considered the volume is fixed. And we have counted the number of samples within this particular volume. And based on this we can determine the density. The density is Pn x that is the nth estimate of the density.

There are two issues. The one issue is if the volume is very, very large, that is the volume is Vn, if the volume is very, very large, then the estimate will suffer from less resolution. That is the first issue. The second issue is if the volume is very, very small, then the estimate will suffer from statistical variability.

So these two issues are very important. Now suppose n is very, very high, that is the number of samples are very, very high. Then in this case, the volume may be very, very small. That means within this small volume, we can expect that some of the samples will fall in this volume. So when n tends to infinity, the volume may be very, very small.

So that is another consideration. Now I will discuss the concept of convergence. We have determined the nth density that is the Pn x we have determined. Now what are the conditions for the convergence, that is the convergence from Pn x to Px, that is the actual density.

So there are two conditions.

One is the convergence of mean and another one is the convergence of variance. So these two conditions we have to consider for convergence of the nth estimate of the density to the actual density, the actuall density is Px. So let us discuss about these convergence conditions. One is the convergence of the mean and another one is the convergence of variance. So in my last class, I discussed the concept of Parzen window.

So in the Parzen window, we can determine the nth estimate of the density, that is the Pn x, that is actually the pdf, the probability density function at the point, the point is x, that is

$$\rho_n(\underline{x}) = \frac{1}{n}\sum_{i-1}^{n} \frac{1}{V_n}\varphi\left(\frac{\underline{X} - \underline{X_i}}{h_n}\right)$$

So in my last class, I discussed about this. So we considered a delta function, that is this delta nx actually the meaning is the delta nx is actually the impulse at the point, the point is x.

So it is the impulse.

So based on this condition, we can express the density, the density is equal to 1 by n summation i is equal to 1 to n delta n x minus xi. So this is the formula for the density, the estimated nth density. Now there are two conditions already I have explained if the volume Vn is very, very small, the volume Vn tends to 0, what will happen this estimated density, the nth density that is nothing but the summation of impulses at every sample points. So actually that corresponds to much statistical variability.

So this is the first case if Vn tends to 0, Vn is very, very small, then the estimated density that is the summation of impulses at every sample point.

And the second case is if this volume Vn is very high, it tends to infinity, very large. So this Pn x that is the estimated density, so it will be a flat PDF. So already I told you it is nothing but the superposition of n number of slowly varying functions. So that is why I am getting the flat PDF. So that corresponds to less resolution.

So these are the two important cases that volume is very small and volume is very high. So what we considered if n tends to infinity, then volume may be very, very small that we can consider because since n is very, very high the number of samples is very high, then we can expect that some of the samples will fall in this small volume, the volume is Vn. Now let us discuss about the conditions for the convergence. Let us consider this nth estimate of the density has mean, it has a mean, the mean is represented by Pn x bar and the variance sigma n square x. So we are considering the mean is Pn x bar and variance is sigma n square x.

So now this estimate that is the nth estimate converges to the actual density, actual density is Px. So what are the conditions? The conditions converges if limit n tends to infinity Pn

x that is the mean that will be equal to the actual density Px. So that means what is the condition? The number of samples should be very, very high. And if I consider n is very, very high, then corresponding to this condition n is very, very high and tends to infinity and corresponding to these conditions the variance will be also 0, the variance will be 0. And also the supremum phi u, phi u is the window function that should be less than infinity, the supremum of phi that should be less than infinity.

So what is actually the supremum? So do you know what is the meaning of supremum? The supremum here I am writing, supremum is actually it is a smallest number that is greater than or equal to every number in a set. So this is the definition of supremum. It is the smallest number that is greater than or equal to every number in a set. So that is the definition of the supremum.

So I repeat this, a supremum is the smallest upper bound on a set.

And also another condition is this limit Vn, n is very, very high and tends to infinity and corresponding to this, the volume may be very, very small. And another condition is limit n Vn because n is very, very high and tends to infinity. So this n Vn will be equal to infinity. So these are the conditions we are considering so that this nth estimate of the density converges to actual density, the actual density is Px.

So now let us discuss about these conditions.

So there are two conditions already I have explained. One is the convergence of the mean, another one is the convergence of the variance. So let us discuss about these conditions. So first is the convergence of mean. So this is the first condition, the convergence of the mean.

So we can determine the average density of Pnx that is nothing but the expected value of the estimated density. So if I take the average of this, I can determine the mean. So it is equal to 1 by n summation i is equal to 1 to n expected value that is the average value 1 by Vn pi x minus xi divided by hn. So this is the mean. And in this case what we have considered the delta function is this delta nx is nothing but 1 by Vn pi x by hn.

So this is the delta function that is the impulse function. And this expression I can write like

$$\underline{\rho_n}(\underline{x}) = \int \quad \delta_n(\underline{X} - \underline{V})\rho(\underline{V})dV$$

So this expression I can write like this and in terms of this delta function, so in terms of this delta function, I can write this delta nx minus V, V is the volume, the density in terms

of volume dV. So I can write like this. This is actually this rho V is nothing but this density x.

So if you see this expression, the expression is very important that is the mean we can determine like this. So mean of the estimated density x minus V rho V, so dV. So this is the expression. So this is one important expression. So this expression you need to understand.

So this is the average that is the mean we have determined. So what is the interpretation of this expression? The interpretation of this expression is if you see this expression minutely, you can see it is the convolution of actual or unknown density with the delta function. So this expression is nothing but the convolution of the actual or the unknown density. The density is PV with the delta function. So that means in this case the mean actually, this mean actually represents the blurred because we are taking the average.

So blurred version of the actual density, the blurred version of the actual density. So now let us discuss what is the conditions for the convergence. So let us move to the next slide. So I am writing this equation again and that is the mean value of the estimated density I can write like this. This is the delta function x minus V rho V dV.

So that is nothing but the convolution of the actual or the unknown density with the delta function. Now when VN tends to 0, that is the volume approaches 0, then delta N x minus V that is the delta function, the meaning will be this delta function will be centered at x. So if I consider VN tends to 0, then the delta function will be centered at x. So that means I can say then this mean approaches the actual density Px as N tends to infinity. N is very very high, then this volume VN may be very very small.

So one important thing is that is if the N is very very high, then we can consider a very small volume. That means we can expect that some of the samples will fall in this small volume. So you can see the conditions for the convergence, the volume should be very very small and corresponding to this, this is the condition for the convergence. The one important thing is N need not be very high.

So I can say N need not be infinity, but the VN tends to 0.

The VN should be very very small. Now let us discuss the second condition that is the convergence of the variance. So this variance, this is the variance of the estimated value of the density, that is the Nth estimate of the density. So that is obtained like this. So i is equal to 1 to N expected value. So from the definition of the variance we can determine like this expected value 1 by N VN pi x minus xi divided by hN minus 1 by N.

This is the mean, mean of the estimated density square. So you can see this is the expression for the variance. So this expression I can write like this. If I take N out from the summation sign, i is equal to N expected value 1 by N square VN square pi square x minus xi hN minus 1 by N square PN square x and I can write like this.

Now let us consider this term. So if I consider this term suppose, this term I can write like this, this star term I am showing. This is 1 by N square VN, I can write like this expected value 1 by VN pi x minus xi hN pi x minus xi hN like this. I can put like this, this position, this term I am showing like this here. So based on this I can expand that variance the sigma N square is equal to N summation i is equal to 1 to N.

So after expansion I am getting this one. So the next step is 1 by N VN, I am taking it out the integration 1 by VN, the pi is the window function x minus V hN pi x minus V hN PV dV minus 1 by N PN square x that is the mean square x. So in this case what I am considering this term, if I consider this term that I can consider as the supremum of pi, we can consider like this. So that means we are dropping in this case, dropping the second term. In this case how actually we obtain the dropping the second term and using the equation what equation? So equation already we know that is the mean of this estimated value of the density that is nothing but the delta N x minus V.

So we have derived this equation dV. So by using this we are getting this one, this expression. So now what is the conditions for the convergence? This variance sigma N square should be less than supremum of pi and the mean of PN x that is the mean value of the estimated density N VN. That means what we are considering taking the maximum value of pi. So that is the supremum we are considering. Now what condition we need? The variance should be 0, the variance tends to 0.

This is the required condition. So corresponding to this condition what will be the case? N VN should be equal to infinity. So this is the case. To get sigma N square tends to 0, the sigma N square that is the variance should be very very small. The N VN should be equal to infinity.

So move to the next slide. So what we have considered? We have considered that this variance should be very very small. This is the condition and corresponding to this condition N VN should be equal to infinity. That is the condition. And again the supremum of pi should be less than infinity. So pi U that is the window function should be equal to 0 when U J tends to infinity.

The supremum of pi should be infinity. So these are two conditions. And based on these two conditions you can see we have obtained the condition. The condition is the

convergence of the variance. So this is about the convergence. The one is the convergence of the mean and another one is the convergence of the variance corresponding to Parzen window.

Now let us discuss the second concept that is the K nearest neighbor technique. So what we have considered in the KN nearest neighbor technique? What we have fixed that KN is fixed and we are growing the volume so that it encloses KN number of samples. That means what we have to determine? Find a volume VN which encloses KN number of samples. So this KN is fixed, the number of sample is fixed and we have to grow the region. We have to increase the region so that this volume encloses the KN number of samples.

And based on this we can determine the density. So what is the estimate of the density? The Nth estimate of the density is KN divided by N divided by VN. So what is N? N is the total number of sample. KN is the number of samples within this volume. The volume is VN. So this is actually the probability that the sample falls in a region and the region has a volume, the volume is VN.

Now this KN we can select like this KN is equal to root N that is actually data dependent way. So KN is equal to root N and that is selected based on this data dependent way. So based on this we can determine the VN what is the volume? So KN divided by N divided by density. So from this expression we can determine the density.

$$V_n = \frac{1}{\sqrt{n}} \cdot \frac{1}{\rho_n(\underline{X})}$$

So if you see this expression the VN is equal to so this is one important equation. So VN is equal to 1 by root N into 1 by PN x. So what is the interpretation of this equation? If the density of the training samples is high around the point x then the region will be small and vice versa. So this is the concept of the density estimation by KN nearest neighbor technique. So what we have considered? So we are considering the point x and we are considering the regions centered at x.

So that means I can say each region is centered about the point x that is actually in R to the power d space the d dimensional space. So what we have to consider the size of the region is expanded until it encloses KN number of samples. So move to the next slide we have obtained the volume VN is equal to 1 by root N into 1 by PN x. So if you see the density of the training sample is very high around the point x the region will be very very small and vice versa. So now you can see the size of the region is expanded until it encloses

KN number of samples.

So where KN is a function of N that is already I have explained actually the KN is equal to root N. So then these samples are the KN nearest neighbor of the location x. So I can write like this. So these samples are then the KN nearest neighbor of the location x. So these samples are then the KN nearest neighbor of the location x.

Now let us discuss how this K nearest neighbor technique can be employed for classification. So in the base classification technique so what we considered suppose this probability of omega i that is the class given x is greater than the probability of omega j given x. So this is the condition we considered for classification. So that means corresponding to this condition I have to select the class the class is omega i this class we have to select. So now how to write this probability of omega i given x I can write like this this probability of x omega i divided by the probability of x I can write like this.

So this can be written like this Ki divided by NVN and this K by NVN. So that is equal to Ki divided by K. So this actually this K is nothing but summation of all the Ki's summation of all the Ki's. So this evidence Px the evidence Px so it is nothing but the summation of probability of x and given omega i.

So that is nothing but the summation of Ki divided by NVN. So this Ki actually represents these are the training samples or these are the samples corresponding to the class the class is omega i. So Ki is the samples corresponding to the class omega i and how many samples we are considering total number of sample is K. So for all the classes if I do the summation of this Ki's then I will be getting the total value the total value is K. So Ki is the sample corresponding to the class omega i. So if you see in the base classification we have considered this decision rule if the probability of omega i given x is greater than probability of omega j given x then in this case I have to consider the class omega i.

That means the meaning is x is assigned to the class omega i. So this is the case and by using the bayes rule we can write like this the probability of omega i given x is equal to probability of x given omega i divided by evidence. So we are not considering that this prior information. So probability of x given omega i that I can write like this Ki divided by nVn and this Px already I have shown it is nothing but the summation Ki divided by nVn. So that is nothing but the K nVn the Px I can determine like this. So putting this value I will be getting this one and I will be getting finally Ki divided by K.

So I am writing it again. So what we have determined the probability of omega i given x is nothing but Ki divided by K. So that means the meaning is the fraction of samples in the region R with level omega i. So that means the corresponding to the class omega i my

samples are Ki. So now if Ki is greater than Kj then based on this we can take a classification decision that means we have to select the class. If the Ki is greater than Kj then we can consider the x should be assigned to the class omega i.

So we can take a decision based on the voting process. So based on the voting process we can take a classification decision. So how to do that is voting the voting process that means we have to count the Ki's. The Ki is the number of samples corresponding to the class omega i and Kj is the number of samples corresponding to the class omega j. So if Ki is greater than Kj then we have to select the class omega i. So this is the classification rule based on the K nearest neighbor technique.

Now mainly it is a voting process. So let us discuss how actually we can do the classification decision by considering this voting process. In this figure you can see we are considering a new data point that is the blue color data point and I have two classes the class A and class B. So in the first figure you can see I am considering a new data point. After this what we have to consider we have to find the distance between the new data point and the corresponding samples of two classes. So based on the nearest neighbor distance I can assign this new data sample to a particular class.

So here you can see in the figure 2 the new data point is assigned to the class omega 1 that is the category 1 the class A because the distance is minimum with respect to the category A that is the class A omega i. So this class I can say it is omega i and this class I can say it is omega j two classes we are considering. So this point is now assigned to the class omega i because the distance from this new point to these samples of the class A is small as compared to the class B the category B. So let us see how actually we can do the voting.

So based on the voting we can take a classification decision. So here you can see in the first figure we are considering three classes one is the yellow one is the green and another one is the red. So three classes we are considering and these are the samples of the classes three classes and we are considering a data point the new data point is the gray color. So this is a gray color data point we are considering. So corresponding to this point this data point we are finding the distance between the samples.

So the first distance is 2.1 and another distance is 2.4 corresponding to the class the class is the yellow. After this corresponding to this green the one nearest distance is 3.1 and corresponding to that is red class I can say the orange class the distance is the 4.5 this is the nearest distance. So after computing these distances you can see the first distance, the nearest neighbor distance is 2.

1. So 2.1 is the first nearest neighbor distance corresponding to the class yellow. Again the second nearest neighbor distance is 2.4 corresponding to the class yellow. The third nearest neighbor is distance is 3.1 testifying to the class green and the fourth nearest neighbor distance is 4.

5 corresponding to the class red or the orange. So in this case, you can see how many votes I can give to the class yellow. So because you can see two times it is the nearest neighbor corresponding to the class yellow. So that means the yellow class it will get the vote of 2 and the green class will get the vote of 1 and the red class it will get the vote of 1. So that means that this new data point now will be assigned to the class the class is yellow.

So these three classes I can consider like this omega i, omega j and omega k. So that means this new data point is suppose x. So now x will be assigned to the class omega i because corresponding to this class omega i, I will be getting the maximum number of votes I will be getting because I have two votes corresponding to the class omega i and based on this new data point x is assigned to the class the class is omega i that is the yellow class. So this is the concept of the KN nearest neighbor algorithm for classification. So we have to find the distances.

So that is why it is computationally complex because we have to find all the distances. So that is why computationally it is more complex. So in this figure also the same thing I have explained. So you can see the new data point we are considering and we are considering two classes class A and class B. So that is suppose omega i and this is suppose omega j and after this we are finding the distance in the second figure and after this we are counting number of votes. So how many votes it is getting and based on this nearest neighbor I can decide the class the corresponding class.

So this is the concept of the KN nearest neighbor classification. So in this class I have explained two important concepts. One is the Parzen window technique and another one is the KN nearest neighbor technique. In the Parzen window technique we have considered the convergence of mean and the convergence of variance. These two important issues we have considered. That means the convergence of the estimated value of the density that is the Nth estimate of the density approaches the actual density.

Actual density is Px. So for this we have considered the two cases. One is the convergence of the mean another one is the convergence of the variance. After this I have discussed how to determine the density with the help of the KN nearest neighbor technique. In the KN nearest neighbor technique we are fixing the number of samples that is the KN we are fixing. And after this we are growing the region so that it encloses KN number of samples.

And based on this we can determine the density. After this I discuss the concept of classification with the help of the KN nearest neighbor technique. So it is mainly the concept of voting and that means we can find the nearest neighbor based on the distance calculation and based on this we can take a classification decision. This is about the Parzen window technique and the KN nearest neighbor technique. So let us stop here today. Thank you.