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-14

Welcome to NPTEL MOOCs course on machine learning and deep learning fundamentals and applications. In my last class, I discussed the fundamental concept of non-parametric techniques. So, directly I can estimate the density, the density is the posterior density, the probability of omega i given x. And in that discussion, I discussed two approaches, one is the Parzen window technique and another one is the k nearest neighbor technique. In the Parzen window technique, what I considered the volume is fixed and I have to determine the number of samples within this particular volume.

That is the concept of the Parzen window technique.

In the k nearest neighbor technique, the kn, kn is fixed and I have to grow the region until it encloses the kn number of samples. So, you can see the difference between the Parzen window technique and the k nearest neighbor technique. In the Parzen window technique, the volume is fixed, I am counting the number of samples that is kn within this particular volume. And in case of the kn nearest neighbor technique, kn is fixed.

So, I have to increase the region, I have to grow the region until it encloses the kn number of samples. That is the difference between the Parzen window technique and the k nearest neighbor technique. Now, today I will discuss the concept of the Parzen window. So, let us start this class and the concept of the Parzen window. So, in my last class, I discussed the concept of the non-parametric techniques or I can write non-parametric approaches.

So, in the non-parametric approaches, what I have determined the probability, this is the recap of the previous class. So, probability that a vector, is the feature vector, the vector x will fall in a region R. So, what is the probability? So, probability we derive like this. So, the region is R and this is px that is the density dx. So, now, what we considered n is the total number of samples, the total number of samples is n.

So, k means the how many samples within this particular region. So, we have determined that probability of k, what is k? The number of samples within this particular region, the probability of k is nothing but we are considering the binomial distribution p to the power k 1 minus p n minus k. So, that is the probability. So, what is k? The k is nothing but the expected value of k that expected value of k. So, that is equal to np.

So, last class I discussed about this. $P = \frac{K}{n}$

So, suppose this is equation number 1. And this probability is nothing but this from the previous equation I can write. So, if I integrate the density, then I will be getting the probability that is the average version of the density.

So, the region is very, very small, then what I can consider this density px, I can consider as a constant. So, that means I can consider as flat pdf. So, the region is very, very small. So, within this small region, I can consider the density is almost constant

. So, for the flat pdf, this density term, I can take it out from the integration and it will be equal to V because if I integrate dx and if I consider region R, then I will be getting the volume.

$$P = \rho(\underline{x})V$$

So, suppose this is equation number 2. So, from equation number 1 and 2, what I obtain the density px is given by k n divided by V. So, I obtained this. So, what we obtained this density I can determine that is the ratio k divided by n that is the variance.

So, last class I discussed about this.

So, this is a k divided by n is the variance. So, this density is actually this density. This is a class conditional density. And what is the volume V? The volume is the hyper volume of the region.

So, region R is considered and corresponding to this the hyper volume of the region is V.

Now, we discussed about the conditions for the convergence. So, what are the conditions? The n should be very, very high. So, based on this condition, you can see what we have obtained the limit n tends to infinity that means the number of samples are very, very high. Then the volume may approaches 0, the volume may be very, very small.

And since we are considering the n is very, very high.

So, that is why k is also very, very high, the k tends to infinity. And the ratio that is the variance that k divided by n, it approaches 0. So, these are the conditions we have explained in the last class. So, the number of samples are very, very high, corresponding to this the volume may be very, very small. And the k is also very high because the n tends to infinity and the variance that k divided by n that tends to 0 because n is very, very high as compared to k.

And if I consider the region is suppose R n. So, corresponding to this region, the volume is V suppose V n. So, the nth estimate of the density is given by that is k n divided by n divided by V n. So, the concept is like this that if I consider this is the region, the region is suppose R n and within this region, we are considering the k n number of samples are available. k n number of samples are available.

And what are the what is the total number of sample, the total number of samples is n, n is the total number of sample. And what is V n? V n is nothing but the volume corresponding to the region R n. So, V n is the volume corresponding to the region R n. So, this is k n. So, you can see and this is k n that k n number of samples we are considering n is the total number of samples and region R n is considered and corresponding to this region the volume is V n.

So, from this, we can determine the nth estimate of the density. So, that is k n divided by n divided by V n. So, there are two approaches already I discussed about the density estimation. The first is the Parzen window technique. So, in the Parzen window technique, what we are considering V n is fixed, the volume is fixed.

And we have to determine the k, we have to determine k and we have to determine. So, how many samples within this particular volume. The second approach is the k nearest neighbor approach. So, what we have considered k is fixed, suppose k is equal to 7. So, region R 1, it contains k is equal to 1 that is one number of samples.

After this I have to grow the region, region growing. So, region R 2, it contains k is equal to 2, the two number of samples. Like this I have to increase the region and suppose region R n, it contains k is equal to k n, k n number of samples. So, it encloses the k n number of samples, corresponding to this region R n, the volume is V n. So, from this information, I can determine the density at x.

So, nth density at x, so k n divided by n divided by V n. So, nth density I can determine. In my last class, I discussed these two approaches. Now, let us discuss about the concept of Parzen window in more detail. So, what is the Parzen window technique? So, for Parzen window, what we are considering? We are considering a hypercube.

So, the one side is H n. So, what we are considering? We are considering a hypercube, each side is H n. So, each side is H n. So, corresponding to this, what is the volume of this hypercube? The volume is H n to the power d, because we are considering the d dimensional hypercube. So, d dimensional hypercube, we are considering. So, what we have to determine? The number of samples falling in the hypercube.

So, for this, what we are considering? We are considering one window function. So, the window function is something like this, pi u is equal to 1, u j is less than 1 by 2 and otherwise it is 0. So, this window function is nothing but unit hypercube centered at origin. So, this is the unit hypercube centered at origin. So, this u is I can write like this, u is nothing but u 1, u 2, u d, because we are considering the d dimensional.

And another assumption we are considering, the pi u that is the window function cannot be negative. That is the meaning is pi u should be always greater than 0. So, this is the condition and also another condition is the integration of pi u, if I integrate the window function, then it will be 1. So, these two conditions we are considering.

So, again I am repeating this, what I am considering, I have to determine the number of samples falling in the hypercube.

So, for this we are considering one window function, the window function is this pi u is equal to 1, if u j is less than equal to 1 by 2. So, I can write if and otherwise it is equal to 0. So, that is the pi u that is the window function, this is nothing but the unit hypercube centered at the origin and we are considering the d dimensional vector. So, pi u cannot be negative. So, that is the pi u is greater than equal to 0.

And also if I integrate pi u with respect to u, then it will be equal to 1. So, these are the conditions. So, let us move to the next slide. So, now we are determining the how many samples within the volume VN that is the hyper volume is VN.

So, how many samples within this particular volume.

So, that is determined by this is equal to KN, KN is the number of samples within this particular VN that is the summation i is equal to 1 to N pi x minus x i. So, x i is the incoming vector or the sample, we can determine the how many samples within this particular volume. So, that can be actually determined like this. Suppose this point I can consider as x and suppose this point we are considering the incoming vector or the new vector is x i. So, if I see this difference between these two, the difference between these two is nothing but x minus x i.

So, this is suppose the age of the hypercube because we know the length of the one side of the hypercube is h N. So, if you see, so this x minus x i should be within the hypercube. It should not cross the hypercube. So, that means corresponding to this case, x minus x i divided by h N should be less than 1 by 2, less than 1 by 2. Then only it will be within the hypercube and the side of the hypercube is h N.

So, because you can see twice x minus x i is actually nothing but the h N, it is is equal to h N. So, based on this condition actually we are getting this condition. So, x minus x i divided by h N should be less than equal to 1 by 2. So, this pi x minus x i divided by h N is equal to equal to unity, it is equal to unity if the input vector, the incoming vector x i falls within the within the hyper volume V N centered at origin. So, you can see here the pi x minus x i divided by h N is equal to unity if the incoming sample x i falls within the hyper volume V N centered at origin.

So, based on this actually we are determining this based on this we are determining this the number of samples within this particular volume. So, what we can determine from the previous formula, we can determine the density directly we can determine the density.

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

So, it is nothing but a k N divided by N and divided by V N. So, summation i is equal to 1 to N 1 by V N pi x minus x i divided by h N we can determine this one.

So, this is the expression for the density. Now, let us see whether it is a valid PDF or not. So, for this what we are doing now P N x d x that is equal to 1 by N same expression I am writing 1 by N i is equal to 0 to N and just I am integrating if I integrate a PDF then it should be equal to 1. So, that is why the integration is considered. So, I am integrating that density with respect to x then if I integrate this one if I integrate this one then this value should be equal to 1. So, I have to test if it is equal to 1 then it will be a valid PDF that means our estimated density is perfect.

So, now for this and determining this so we are considering let u is equal to x minus x i divided by h N. So, let we are considering u is equal to this. So, what will be the d u if I differentiate this one then it will be equal to d x divided by h N to the power d because in this case we considered what we considered x is equal to x 1 x 2 x d d dimensional. So, that means I have to differentiate d times. So, that is why it is h N to the power d for this d dimensional vector.

So, that is equal to what is h N to the power d h N to the power d is nothing but the volume h N to the power d is nothing but the volume. So, we are getting this. So, what we have considered in the previous expression p N x d x is equal to 1 by N summation i is equal to 1 to N 1 by v N that integration we are considering integration pi x minus x i divided by h N d x. So, which can be written like this 1 by N summation i is equal to 1 to N. So, it is equal to integration pi u pi u we can consider pi u d u is equal to 1 because we know if I take the integration pi u d u that is equal to 1.

So, that is actually the property of the window function in this expression actually what we have considered d u if you see the previous expression d u is replaced by d x divided by v N. So, that we obtain. So, that means this d x is nothing but so the d x is nothing but v N d u. So, if I consider this one then this v N v N will be cancelled out and I will be getting 1 by N summation i is equal to 1 to N integration phi u d u is equal to 1. So, that means this value we are getting d x is equal to 1.

So, that means so this estimated density is a proper PDF the probability density function. So, P N x is a proper PDF. So, that we have determined now this P N x that is the density P N x is equal to 1 by N summation i is equal to 1 to N 1 by v N pi x minus x i h N. So, we have this expression for the density. So, now let us consider delta N is equal to 1 by v N pi x divided by h N.

So, corresponding to this this expression will be converted into is equal to 1 by N summation i is equal to 1 to N delta N x minus x i. So, I will be getting this one. So, this is the expression the final expression for the density. So, this is a very important expression for the density.

So, let us move to the next slide. what we have obtained, we considered the delta N x is 1 by v N pi x by h N and corresponding to this we have determined the density the density is P N x 1 by N i is equal to 1 to N delta N x minus x i expression for the density. So, in this case, you know that the v N is equal to h N to the power D the volume is v N is equal to h N to the power D. So, you can see the h N clearly affects both amplitude and the width of the delta function. So, h N clearly affects both amplitude and width of the delta function of delta N x that is the window function. So, now let us consider how actually h N affects the estimation of the density function.

So, there are two cases the case number one if h N is large then what will happen amplitude of delta N is small and x the vector x must be far from x i. So, this is the outcome if the h N is very large the amplitude of delta function delta N x already we have determined. So, amplitude of the delta function is small and x must be far away from x i. So, then this estimated density is the superposition of N broad slowly changing functions. So, that is the estimated density is the superposition of N broad slowly changing functions.

So, that means I can show like this pictorially. So, this is the estimated value of the density and what I will be getting I will be getting something like the flat PDFs the superposition of N broad slowly changing functions. So, these are the flat PDFs. So, it actually corresponds to this condition actually corresponds to less resolution it corresponds to less resolution that is actually I can say out of focus estimate of the density. So, if I consider this case in the case number one if the h N is very large amplitude of delta N is small and x must be far from x i. So, corresponding to this you can see this estimated density is the superposition of N broad slowly changing functions.

So, this is the case number one. So, move to the case number two. So, in the case number two if h N is very small peak value of delta N there is a delta function x minus x i is large and occurs at x is equal to x i. So, if I consider h N is very small the peak value of delta N and that is the delta N x minus x i is large and occurs at x is equal to x i. So, that means the meaning is this estimated density is the summation of impulses at every sample points. So, that means I can show this is the density.

So, I will be getting only the impulses. So, that corresponds to this corresponds to much statistical variability. So, that corresponds to much statistical variability. So, that means we are considering the h N is very small and corresponding to this the peak value of the delta function delta x minus x i will be very large and it occurs at the point x is equal to x i and corresponding to this the estimated density is the summation of impulses at every sample points and that corresponds to much statistical variability. So, this is a condition. So, two conditions we are considering in case number one we are considering h N is very large in the second case we are considering h N is very small.

So, what is actually this delta function if you see the delta function what is this delta function. So, delta N x we considered is equal to one by V N pi x by h N this is the definition of the delta function. So, this is actually the direct delta function. So, this delta N x meaning is the delta N x is an impulse function impulse function at x. So, as the h N tends to zero that it approaches zero, then what will happen this delta x minus x i it approaches a dirac delta function direct delta it approaches a dirac delta function centered at the point the point is nothing but x and in this case this the density the estimated density is a superposition delta function centered at the samples centered at the samples.

So, that is the meaning of this case number two. So, that means, you can see here what is this impulse the dirac delta function is the impulse at the point at x this is the direct delta function. So, when h N tends to zero, then this delta x minus x i it approaches a dirac delta function centered at x and corresponding to this the estimated density is a superposition of delta functions centered at the samples and this corresponds to much statistical variability. So, in one case, I have less resolution and another case the much statistical variability the more statistical variability.

So, these are the cases. So, move to the next slide. So, what we have considered the recap is just I am writing it again, we have estimated the density 1 by N i is equal to 1 to N 1 by V N pi x minus x i divided by h N. So, this is the expression for. So, this is the expression for the density and what we have defined the delta function is delta N x is equal to 1 by V N pi x divided by h N.

So, we considered this. So, this expression already I have discussed. Now, corresponding to this expression, corresponding to this expression what we have considered two cases, h N is very small and another one is h N is very large, h N is large. So, these two cases we have considered case number 1 and case number 2. So, the summary is if V N is because V N is equal to h N to the power D, if V N is large is too large what will happen? If V N is too large the estimate will suffer from little resolution and another case if V N is too small because you know the V N is nothing but h N to the power D. So, what will happen? The estimate will suffer from too much statistical variability.

So, these are two extreme cases. So, V N is too large and V N is too small. So, one condition we have to consider now. So, if I consider N tends to infinity the number of samples are very very high, then we can consider the V N may be very very small that we can consider. So, that thing we will discuss later on the conditions for the convergence. So, in my next class I will be discussing about the conditions for the convergence.

But this condition is important. The condition is if I consider the number of samples are very very high, then the V N may be very very small. So, this condition I will be explaining in my next class. Up till now I discussed the concept of the Parzen window and I have shown the expression for the density. So, we can estimate the density with the help of the Parzen window technique.

After this I discussed two cases. In the first case, I discuss when the V N is very very small the volume is very very small. And another case if the volume is very very large. These two cases we have discussed. And after this I will discuss the conditions for the convergence.

So, one thing is very important. The number of samples are very very high. And in this case I can consider the volume may be very very small. So let us stop here today, Thank you.