Indian Institute of Technology Madras

NPTEL
NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING

Pattern Recognition

Module 06

Lecture 05

Visualization and Aggregation

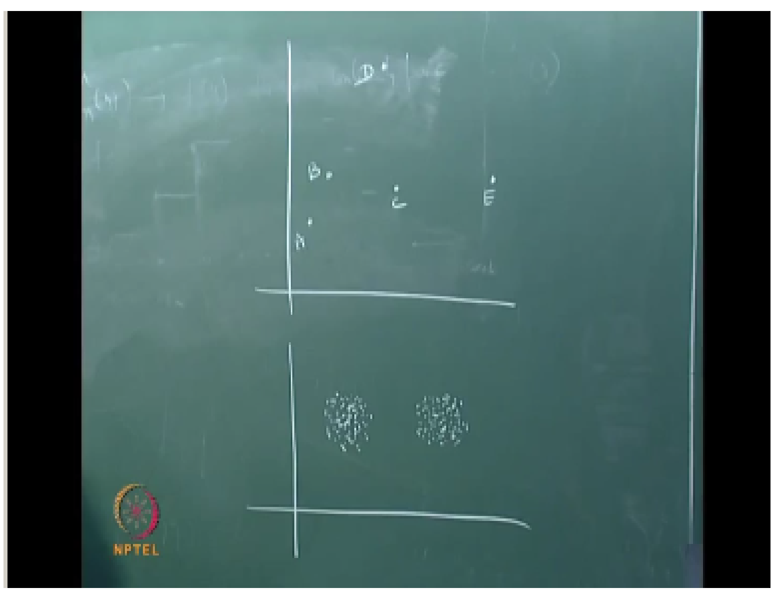
Prof. C.A.Murthy

Machine Intelligence Unit,

Indian Statiscal Insistute. Kolkata

I shall be talking about visualization and aggregation visualization means for supposed to look at the data and then somehow draw some conclusions aggregation means you are supposed to sum up the whole data set and then somehow you say a few words about the data set as examples.

(Refer Slide Time: 00:41)



Let us look at this diagram there are five points plotted ABCDE okay and from the look of this five points the distance between a B is probably the least among all the distances then probably the distance between D B and C comes probably then C and E okay then probably be and d so just by looking at this data set of five points you know that a is close to be than to any other thing and afterwards I mean between four be between a and C a is the closest then C.

So these are the sort of conclusions that we are drawing about the data set just by looking at the data now let us look at this, this points are plotted now from the look of this diagram we are aggregating it and saying that there are two clusters here this is a set of point consisting of one cluster this is a set of point consisting of another cluster just by looking at the data set just by looking at the data set we are making statements about the data here.

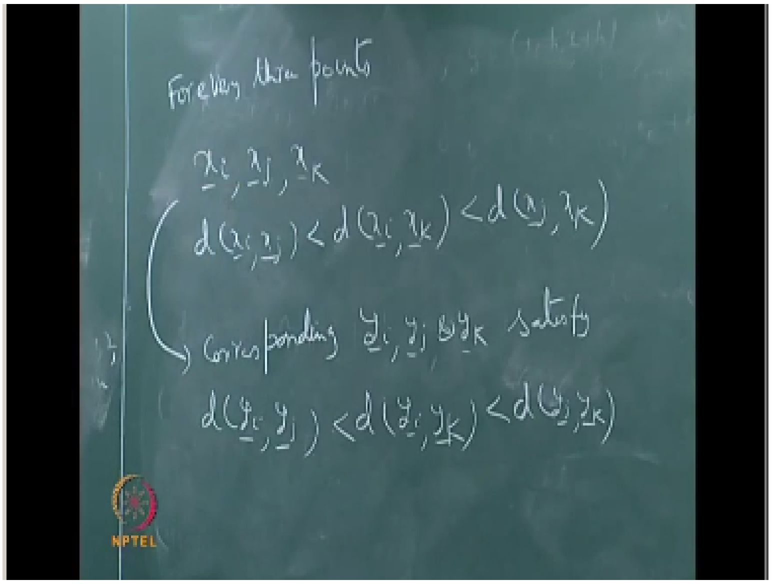
Also just by looking at the data set we are making statements what is the meaning of looking we are basically able to plot the points and we are making judgments about the data on the base of the plots unfortunately you can do this plotting of the points only when the data is in our to the moment the data is in our three are any other higher dimension you are not able to plot so you are not able to visualize and you are not able to aggregate.

So the question actually boils down to how do you plot points which are in higher dimension to a lower dimension preferably dimension to preferably dimension to this problem has several connotations this problem has several connotations people have been working on this problem for a long, long time for along long time you can make this dimensionality reduction I am using our dimensionality reduction intentionally.

Because we are original data is in very high dimension you would like to bring it to your lower dimension so that you can visualize or you can plot you can make this dimensionality reduction in many ways we have already read about PC a principal component analysis where you make some sort of dimensionality reduction we have already discussed several feature selection criterion functions several feature selection algorithms where we reduce the dimension.

But then Here I am going to talk about it in a slightly different way again I will come to this diagram in this diagram distance between a and b is less than distance in B and C suppose you want to decrease the dimensionality from a higher dimension to lower dimension then one of the properties that you may want this dimensionality reduction to have is probably the following.

(Refer Slide Time: 04:48)



Suppose your data x1 x2 xn it is in some capital M dimensional space and you would like to reduce it to say x1 is changed to y 1 x2 e change it to y 2x any change it to YN to some small M dimension where small n means less than capital M X 1 is changed to y1 x2 you change it to y 2xany change to yn and from Capital m the dimension is reduced to small M small M is less than capital and you would like to the reduction to take place in such a way that suppose for every three points every three points X I XJ xk for every three points xixj xk such that distance.

Between X I and X day is less than distance between say X I and XK say this is less than distance between XJ and XK there are three distances xiij is the least XD X K is the highest and the middle one is this for every three such for every three points satisfying this the corresponding why I YJ and YK why I YJ and YK they satisfy why I and YJ is less than distance between why I and YK and that is less than distance between y j +YK.
So you won the Lima logical deduction to take place in such a way that somehow this less than greater than relationships among the distances they are preserved okay this is the condition that you would like to impose now the problem here is this is the condition that we would like to impose this is fine but the moment you impose this condition note that this is to be satisfied for every three points not just for only those three points for every three points xi xj.

And XK so how many such through built-ins you can have NC three NC three like three buttons you can have and forever three points there will be some such relation let me also include equal to so that you would not have this problem of equality and inequality let me also include equality so if you take three points you have three distances and they will satisfy one on this one then the corresponding the transformation once it is transformed to why they should also satisfy the same inequality.

And this should happen for every three point now what is the shown in topology is that such a thing is not possible such a thing is not possible for every data set in fact you have too many too many examples where these things are not they cannot happen for example I will ask you to do it if you have time to take points in unit not unit key you can even take it in unit cube unit cube how many points are there are eight points.

And you have the corresponding distances right HC to distances right unit cube you have eight points just look at the nodes only do not look at the edges look at the nodes there are eight nodes so you have eight points you try to get eight points in our two so that the equality and inequalities are satisfied you will see that you cannot get it you will on your own you will be able to say that you would not be able to get it you can try to do it okay.

And if you increase the number of points then it will be that much more difficult so in fact it has been shown there is also a theorem in topology that for in fact for very rare cases such a thing is possible such a transformation is possible in general such a transformation is not possible for infinitely many points most of the times it is not possible and if you take discs etc that is just simply impossible it is simply impossible.

Now there is a phrase that you generally use topology preserving I do not know very how come across this phrase are not topology preserving maps okay these topology preserving maps from higher dimension to lower dimension they are generally not possible you cannot have them so the problem is that since you want something to happen and the results are somehow the theoretical results are not there in fact.

They are they are just not that there the results are stating that they are not possible this is not included in our course of lectures there is something called coronal's Coonan self-organizing map it is something called Conan self-organizing map where he tries to bring data from higher dimensions to two or three dimensions usually people generally take two dimensions where Conan tries to approximate to the best to the extent possible from higher dimension to lower dimensions okay.

So that it is in some way but then note that that is only an approximation and no such thing is possible always this is one thing that I wanted to say this is partially it comes into visualization and there are some other results that are there those results are concerning what are known CHERNOFFS faces if you do if you go to any searching use any search engine and then just write cheroot's faces you will get many faces basically they are like this.

Okay what CHERNOFFS faces CHERNOFFS use a very well known name in statistics okay there are some CHERNOFFS bounds on errors what you are not thought that we people we generally we human beings if there is little bit of change in the phase we immediately notice it so he thought that probably faces are a best are a way of representing data using faces you can represent a data.

Basically what is it that we want to have small changes are also we must be able to find out when you want to represent a data figuratively are by some plotting or whatever it is basically what is that we wanted to observe we want to observe small changes also we would like to actually get them from the data set now one of the places where we always notice changes small changes that is in phases itself okay.

So how do he thought that why do not we represent the data set by a phase so he represented data as a phase note that if you write a face like this, this is two-dimensional okay this is two-dimensional now depending on he basically changed there are ears also okay please forgive me for very bad drawing there are s also so length of I mean earlobes the measurement o fear lobes may be it may correspond to some of the features.

Okay the distance between eyes it may correspond to some of the figures the radius it may correspond to one of the features so like the, the features that he has he represented them basically as a phase but then he used the data set that he presented for 32 33 features it can take care of but then if you are looking at 100 or 200 features then you really I mean at least he is not able to do it I do not know whether we are able to we can do it or not that is one.

But then every observation he is writing as a phase but then what we are doing there is between observations also we are having some relationship that we are able to find here he is writing each observation as a phase but then between observations what is the difference that only when you see all the faces then you understand it but if you have to see to 300 phases then you have a problem in aggregating right if you have to see to 300 faces.

Then from one phase to another phase you go from the first phase 2 when you come to50th phase probably the information in the first phase we want we may not remember whereas if you see all the 50points are these two hundred points in this way we do know that this forms one cluster this forms one cluster such a conclusion if you write all those two three hundred faces whether you can draw or not that is not quite clear.

But then this is one of the best-known papers on visualization of datasets CHERNOFFS face it is available on internet even now this was some time in 1960 s this paper was written this is one of the best-known papers afterwards you have probably the best-known paper is Conan's paper on self-organizing maps for dimensionality reduction in between many other persons wrote many articles where people have represented data by three.

And they used in fact people used various ways graph is a way of representing data but then they are not exactly successful so this is still how best you can represent a data set of higher dimensions in two dimension so that we understand many facets of the virginal data set well by used by looking at the two dimensional data this particular problem still does not have always satisfactory solutions.

I will be giving a brief talk of probably around ten minutes on n symbol classifiers technically the meaning of n symbol classifiers we have been discussing about which classifier is better which classifier is not better on data set show to resolve this problem this we have been discussing about this earlier now instead of saying which classifier is better or which classifier is not better can we say can we take a combination of these classifiers to get a new classifier.

So that this new class fire which is taken as a combination of the classifiers maybe this works better than all the individual classifiers so this particular thing is known as n symbol classifier as you can see as you can understand the word and symbol means you have a collection of classifiers from there you are going to get a classifier that is why it is called n symbol classifiers.

Now there are many ways in which this can be done one is which is known as bagging what is done here is you have a training data set and you have a classifier now what you do is that you take simple random do simple random sampling on this data set with replacement what is the meaning of that say you pick one observation randomly note the observation and you know the corresponding class label also put it back.

And then you draw the next observation so the next observation could be the first observation also so if you have smaller number of points in the training set you will draw smaller number of points randomly with replacement so it may so happen that you may get the original data set itself are many times what happens is that you get n number of points all right.

But some points will be repeated so what you are going to get will be a multi set okay now on every one of these sets multi sets you apply or you learn the parameters of the classifier are you do the training whatever it is okay so for every one of these multi sets you have a classifier so if you draw 10 such multi sets you have done such classifiers okay.

Now when unknown observation comes you apply all these 10 classifiers apply all these 10classifiers maybe six of them will say that it will go to class 14 of them will say that it goes to class too then the majority wherever it is going you say that that is your rule majority voting this procedure is known as bagging I will repeat it basically what you do is we take do simple random sampling with replacement do simple random sampling with replacement.

And apply the class I am learn the parameters of the classifier that means you train the classifier on this training set basically when you do simple random sampling with replacement you get multi you may get multi sets not always but many times you get multi sets so this even if you have two or more classifiers you can do the same thing that is if you have three classifiers you have got these many multi sets for each multi set.

For each classifier you learn the parameters of the classifier are you train the classifier so if you have 10 such multi sets and three classifiers then you have totally 30different classifiers then when you get an observation in the test set apply these 30 classifiers on them then the majority voting to take it that is bagging okay are and then you have like that you have several other techniques you have boosting.

There is something called ad a boost which is a very, very nice method for in simple classification and there are many other methods basically in all these methods basically let me not say all the methods basically in most of these methods after you get the classifiers how you get the classifiers that there may be many ways after you get the class face when you are applying it on a test data set usually majority voting is taken usually major Keewatin is taken.

And there are some papers way instead of majority voting people have done some other ways other means of doing this finding which is the best classifier are finding which particular combination is to be taken one way is majority voting there are some other ways where you take something like a linear combination.

Or some combination of the classifiers some combination of classifiers also people have taken in some situation but on this the many papers are being published now and I expect to see many more papers this n symbol classification thank you.

End of Module 06 – Lecture 05

Online Video Editing / Post Production
M. Karthikeyan
M. V. Ramachandran
P. Baskar

Camera
G. Ramesh
K. Athaullah
K. R. Mahendrababu
K. Vidhya
S. Pradeepa
D. Sabapathi
Soju Francis
S. Subash
Selvam
Sridharan

Studio Assistants
Linuselvan
Krishnakumar
A. Saravanan

Additional Post – Production
Kannan Krishnamurty & Team

Animations
Dvijavanthi

NPTEL Web & Faculty Assistance Team
Allen Jacob Dinesh

Ashok Kumar
Banu. P
Deepa Venkatraman
Dinesh Babu. K.M
Karthick. B
Karthikeyan. A
Lavanya. K
Manikandan. A
Manikandasivam. G
Nandakumar. L
Prasanna Kumar. G
Pradeep Valan. G
Rekha. C
Salomi. J
Santosh Kumar Singh. P
Saravanakumar. P
Saravanakumar. R
Satishkumar. G
Senthilmurugan. K
Shobana. S
Sivakumar. S
Soundhar Raja Pandian. R
Suman Dominic. J
Udayakumar. C
Vijaya. K.R
Vijayalakshmi
Vinolin Antony Joans

Administrative Assistant
K.S. Janakiraman

Principal Project Officer
Usha Nagarajan

Video Producers
K.R. Ravindranath
Kannan Krishnamurty

IIT Madras Production

Funded By
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in