NPTEL

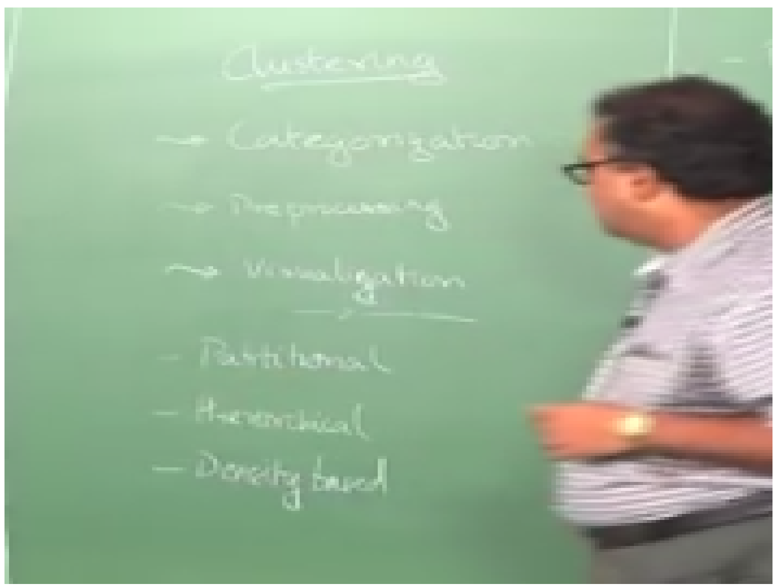
NPTEL ONLINE CERTIFICATION COURSE

Introduction to Machine Learning

Lecture-70
Partitional Clustering

Prof. Balaraman Ravindran
Computer Science and Engineering
Indian Institute of Technology Madras

(Refer Slide Time: 00:17)



Clustering so far you know what clustering is right so we did in the very first class we looked at clustering so all you know what clustering is and essentially is idea is to group together data points that are similar right, so what you are essentially trying to do is find a partition the simplest clustering problem is stated as follows you want to find a partition of the data points such that the similarity between the points are belong to the same cluster is maximized right and the similarity between points that belong to different cluster is different cluster is minimized right.

So that essentially the thing and so if I give you a set of n data points right and I ask you to partition into k clusters, right how many clusters are how many different clustering are possible so when I say a clustering it is a set of k cluster okay so how many clustering's are possible if I give you n data points and k clusters and before condition that none of the cluster should be empty huge number right, so nice question to ask in exams but it is a huge number, huge number of clusters right.

So it is just impossible for you to exhaustively search through all the possible clustering's and then come up with the one that is best right, so inherently clustering is all clustering algorithms we will look at are all some form of an approximation or the other right to, to the actual base solution, right in fact the some satiation so consider clustering a ill post problem right and do not day into solve it you know, okay that is a ill post problem I will not solve it right so thing is it is a real problem.

People do not so all kinds of a place where we look at clustering so there are two main things that we do it clustering right, so the first one is as a machine learning as a task as a data remaining task in itself right, I am interested in producing clusters right so I am interested in producing clusters right so this is some kind of a what we will call some kind of categorization I want to take this data point data is set of data given to me and then want to categorize them into different groups right in it of itself.

I am interested in doing clustering right clustering is also very valuable as a pre processing tool right, so why would I want to do clustering as a pre processing to so I can take a very, very large data set if I have a cheap way of clustering it right I can cluster it reduce it to a few data points right I can take a say 10m no data way date set.

Like 10m for 10m items and then I can say I am going to cluster it into 10, 000 clusters right, pretty large let we take a long time to do it but then if it have only say 10, 000 representatives of this 10m data points so I want that to sample 10, 000 data points from this 10m but I wanted to it in such a way that they are representative of the data is possible so what I do is I use clustering right and so 10, 000 is a large number of clusters right so each of those clusters is going to have few 1000 points right.

Not very large right so then I can just go and pick out one the representative for each one of these as suppose to sampling directly from a 10m road space okay so that gives me some kind of leave here right so pre processing, the rather place where we want to use clustering right and any other things we can think of our clustering is useful, exactly right so for visualization again cluster is something there is very useful right instead of just looking at large table of data something like that if I looked at it pictorially and I show you that okay here are one set of data points at belong to one group here is another set of data points and stuff like that then it makes it lot easier for you to understand the structure of the data that you are looking at right.

So clustering allows you to visualize the data and understand any kind of special structure or structure in the feature space right when you say special structure does not mean that it is actually 2D space right I mean structure in the feature space. So is there any kind of structure in the feature space that you are able to understand that right. So it is a very valuable visualization tool and also very valuable for you to understand something about your data right.

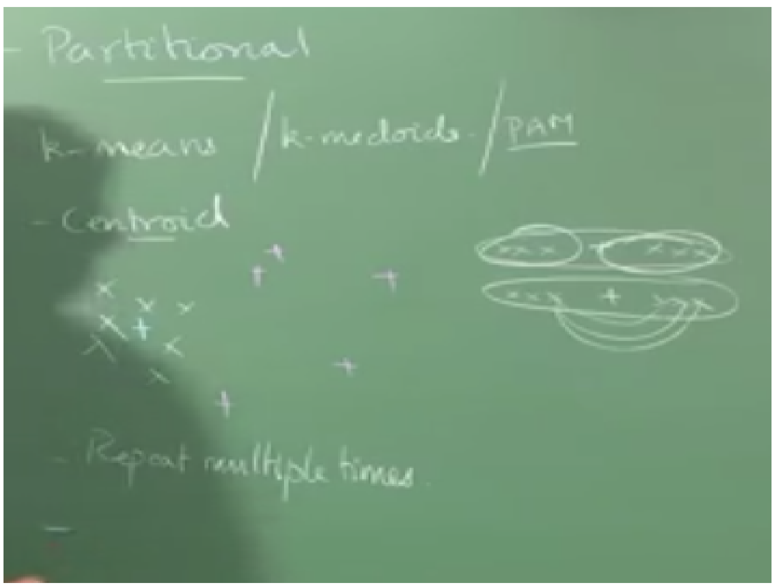
So in fact when I talk about categorization right so in and of itself it can be the problem that you are solving or it could be something that you do as a people are seeing step before you go and actually solve the problem, so I have been interested in finding out how many classes are there in my data so I do not know I am not be given class labels a priory okay I do not given class labels a priory but can I tease out class labels right can I say that the data contains seems to be coming from three different distributions right and then I am going to say each of those is a separate class condition distribution and I can assign a class label to all of those distribution right.

So it is like I am going to look at the data try to understand that and then see okay these are people who are likely to finish a course right these are people who listen to all the lectures but not write the exam I do not know what to be so but we not look at the data at here okay. so this kinds of things which you could do that is what I am saying, clustering is just not a one short okay give me the data here or the clusters in you are done with it in fact quite often it is not even call clustering it is call cluster analysis because you just stop with the clustering alone you actually have to go and figure out what the clustering is telling you.

Classification in some sense is much easier in some so you given something you are basically return the labels and more often than not your done right but clustering it is usually a step on the way to something else okay right. So there are many ways in which you can do clustering so the most popular of this okay are called partitional approaches right so partitional clustering and I tell you what is the others are I get to them on okay.

So first thing is partitional clustering then hierarchical and then density based and they are not really disjoint classifications right but typically methods get index end of this three and then there are many, many other smaller things raise up, so somebody wants to get a paper return the over I propose something that is neither partitional large hierarchical and then they will propose a new algorithm and thing like that, so they are just to sell their papers and not necessarily separate classification right. But these are kind of the three main classifications right and so look at each in turn right.

(Refer Slide Time: 08:55)



Partitional clustering things K means, everyone knows what K means as right, so why do we call them partitional clustering? Clustering is partitioning the data right so why I am calling this as partitional clustering. So I am essentially these are methods which search through the partitions directly right the final partitions that I want right they search through the partitions directly so that is why they are partition clustering methods okay.

So suppose to hierarchical clustering methods right which do not says through the space of partitions on the entire data set right they first try to do it in to two group right they do not search through all the partition suppose I am interested in K things right they do not do all the K they do not do the K clusters right they can start off with two clusters and then split the two clusters in to four and so on and so forth okay.

They can go down in to their, they do search directly in the space of k partitions right while k means exactly does that right this when what did you do in k means I am sorry so what you are doing in k means you start off with k guess for the centers of the clusters right so there are few things that we need to know so the first concept here is the centroid so what is the centroid well this et center of cluster right so essentially you take the average along each coordinates right suppose I have 10 points it belong to the cluster it is in a five dimensional space.

For each dimension I take all the 10 coordinates take the average right so finally I will end up with the single data point so that is the centroid right suppose I have some set of data points like this right so the centroid could be some where there now that I am exactly computing the centroid but it will be some where there so right sure so what is a how do you find perfect solution, yeah so there are many ways of cleverly initializing this right we just talk about vanilla k means right so the many ways in which you can look at a clever starting point right.

So the other things which people do they what they do is they start off with 10% of sample of data and then they repeatedly run clustering on that and figure out which are the good centroids for the 10% of data and then that use has an initialization for clustering the whole data right. So the idea there being rather than repetitions of very small when you are doing the 10% of sample of whole data so this is the one way you do it the other way of doing it is to do this kind of an initialization where you try to move to the further corners.

But then that has it one problems right so essentially it will make it more sensitive to out layers right because you are trying to put your centroid the beginning centroid at the edges of your space right so it will make it more sensitive outline, so those there are issues. Yeah that is k merger another’s so I do not think he meant that I think he meant something else we will come to that right so might there are I mean again am many heuristic so I think the current came in variation champ is I think k – means ++ of some thigh right.

So that gives you a very good initialization and then your drawing a clustering from there and works well right but k – means is an approximate thing so when you look at DM right so you actually see a more well founded derivation for k – means so right now I am just going to introduce it to you as a heuristic right but later on when you look at EM right the canonical EM problem that you solve right first canonical EM problem that you look at will be K – means.

The variant of k means usually we call Gaussian mixture modules but it is essentially like k mix okay so what you do with k – means is that is that you first pick centroids for all your clusters right a randomly right of course I have k clusters then I will just choice k centroids at random right that could even be like that right choice k centroids at random and then I will have my data points so what I do is I assign each data point to the centroid that is closes to it right I say in each data point the centroids that is closes to it okay now I forget the old centroids.

Now I have k groups of data point's right and for each of those groups I try to find the new centroid okay this is the actual centroids the first once I started over of centroids where really not centroids but we keep we still call the centroids anyway so in the just to keep the terminology uniform right now I recomputed the centroids then I go back and assign each data point to the centroid that is closes to it.

Yeah whatever distance measure you are using right so you are in some RP space right remember our data points come from some p dimensional space so in that space whatever is nearest you will you could use ulceration you could use whatever distance measure you want it choose the appropriate distance measure right so k – means or any of these distance based computations that you do right works best okay any of the distanced based computation you do works best if everything is real value right not even integral right.

So everything is real value that works best if you are going to categorical attributes you will have to think of a different distance measure that you will have to defined and I do not know if may I can talk about 1 or 2 things that people use for categorical things but the most popular by far is using some kind of Jucar similarity, right so I have lots of cat values that the thing can take and then I say okay how many of those it actually is similar on right so how many dimension I have may actually agreeing with the other one on so if I have the same value in that categorical dimension then I will say one if I do not and the I will say 0 and the I will try to find how many I agree on right.

And that could be one measure but then defining a centroid there is little tricky right so categorical variable may centroid might come with a value of you know red 0.5 where I mean right 0.5 times red + 0.5 times green right so what would that mean no do not add up the colors please I mean I said brown or something I do not know what red and green would up with but do not know that does make sense right so we have to we, very vary about using k – means when you have categorical attributes.

Yeah so what I mean by probability vector like one of n kind of encoding right 10 encoding right yeah you could do that but then really what would be the mean centroid value for that 10 encoding, yeah that is a interpretation issue so you could still try k – means I mean you do not really have to interpret what the centroid means unless your returning the centroid as a representative point right if I am returning the centroid as a respective point then you get into problems.

So just going to do clustering on it you can go ahead and do crusting on it you do not have to really interpret the value of the centroid right so there are ways of handing this when you have categorical attributes right so k – means is the simplest way then there is something called k medoids right which kind of gets around this whole issue of having to generate an artificial centroid right so instead of centroid it essentially uses a the equivalent of median right so the mean cannot be not actually be a data point but the median is always a data point right.

So likewise medoid I always a data point so it gets around this interpretation issue so computes the centroid and takes the data point closes to the centroid as the representative so the medoid right and the you also have other kinds of things which is called portioning around medoids right so we do portioning around medoids you work with data points as the representative of the clusters right and you do not ever generate an artificial point you always choose a another data point as a centroid I will describe that more right.

So those in such cases you do not have to worry about this you know meaningless attribute values being generated but you still need a distance measure so you will have to come up with some kind of distance measure which categorical attribute so if you use 1 out n or 1 hot encoding or anything else you still use some kind of a Jaccard similarity right so then do that okay.

Right so going back let us finish up the simple k means algorithm right so what you do with k means is so once you have done the assignment to the centroids if you forget the centroids estimate new centroids and keep repeating this until you no longer make any changes right sorry done?

No, no labels is necessary, there are no labels, there are unstable is learning from right. So if you have labels and other things, I mean depending on what your application is you might want to look at labels right, that is what the scameans is concerned, the question is the following. And you start with initial guess for the centroids right, I assign data points to the closest centroid right, recompute the centroids, then reassign the data points the closest centroids, I keep doing this until the centroids no longer change.

The question is have you done? Consider a, yeah, okay. Al right, how many clusters you want me to put them into? Three clusters I am done, you are getting close. So the probability of recycling is really small yeah, yeah whether other things which I can show is slightly more dramatic right, okay. These are my two sets of data points. Obviously, they are in two clusters, are not they?

These are the two initial centroids I start off with okay. So what are the two clusters I will get? Okay, and I compute the centroids for those, where will I end up with, and I reassign the data points to the centroids, I will end up with the same thing right so because of my bad initial choice for centroids, I do end up at a point where the centroids do not change anymore, but it is a really, really bad clusters.

Yeah, randomly chose those two points right, I randomly chose things. Exactly, so my question to you was if the clustering does not change or you are done, the answer is no okay. In this case surprisingly one of the few cases where the answer is not it depends, the answer is no okay. Because if you just do it once right, you are not done, so that is the thing which came in, you will have to repeat it multiple times right.

And please every time you said different random seat for choosing your initial starting point right. If you just the same random seat then you will get the same thing right. Yeah, so we will come to that right. And right, so the people understand why you have to do this multiple times right, because you will get stuck in some kind of local optima right, and you have to start over again with a different random set of case centroids right.

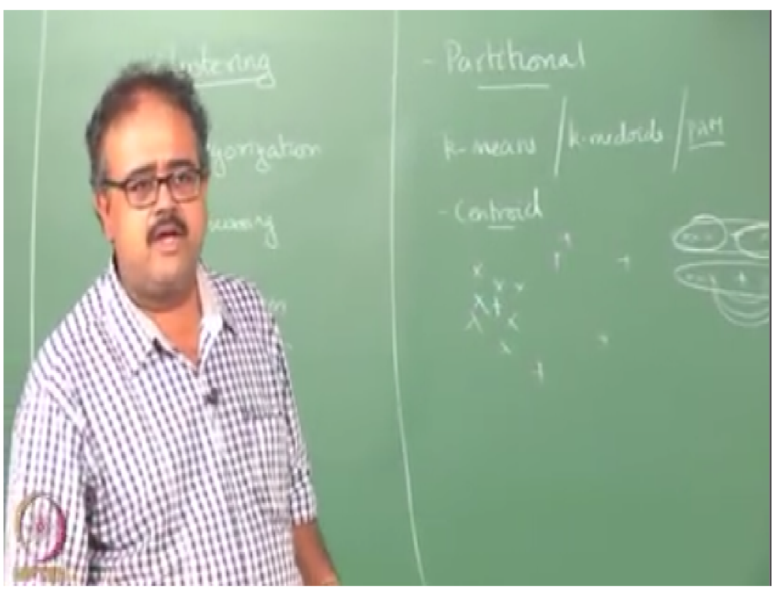
And then keep doing this yeah. Good point, so there are different measures right, I never told you about cluster evaluation measures so far right. So there are different ways in which you can evaluate clusters right. So more of the more popular is this some kind of a dispersion measure right, so I took a diameter, so the way the diameter is different there are two definitions unfortunately for diameter the literature.

So the first one is the average par wise distance between the data points that belong to each cluster right so if you took that right, so I will take the pairs, pairs of data points right, I will measure the distance between the pairs. So but, then I will also have to consider pairs like that right. So I will look at the distance between every possible pair of points that belong to the same cluster right, and I will take the average okay, that will be the diameter of this cluster.

And the average of the diameters for all the clusters will be the quality of the overall clustering okay some kind of a dispersion measure, right so how much spread out my data point this right alternatively if you have if we are completing centroids right we can also compute the average distance of the data point to the centroid right and take the average and cross all the clusters.

And use that as your quality measure side so the either one is referred to as diameter in the literature so some call the and the most popular one is the average pair ways distance okay, is it fine right and I take it back so the average distance to the centriod is called the radius of these average is the centriod is called the radius of the cluster right.

(Refer Slide Time: 26:00)



So the diameter the second definition of diameter is the max of the pair ways distances right either to the average of the pair ways distances right or the max of the pair ways distances both could be call the diameter but the radius is essentially the average of the distances to the centriod, okay so this is one measure right so there is an another measure which I could think of which is essentially is called purity which did you guys look at purity already, will never gave it your in the first then assignments of array.

So purity essentially tells you for each cluster that you have what fraction of the data points belong to a the same belong to one class what is the largest fraction of data points and belong to one class, okay so purity can be used only when you have data sets that have classed label associated with that right so suppose I have 10 data points I am going to one cluster six of them are in class one two or into class two or in class three, so the purity of the cluster is 0.6 right because out of 10 or in class one two out of 10 are in class 2 so 6 out of 10 is largest thing.

So 0.6 is the purity of the cluster so like this so whatever right so which are the level of the hierarchical or evaluating it at we will have some set of clusters, right we just evaluate the purity if you purity is a measure that it choose right so there is matter so purity mixing right, so something related to purity again if you using label data sets we can use something like entropy.

Right and what would be the entropy when the class distributed right look At the class distribution in the cluster so the P 1 fraction in the class 1 and the P 2 class 2 so that we can do the P 1 _1 so since that is entropy so that their classes are more are less evenly distributed then the entropy gives you better measure than the purity so the purity measure is the 0 .5 it is not clear to be weather the other 0.5 belong to the other class or the belong to the many other classes.

So that is the thing so these are the measure that reduces and then we having the silly this is the satiation will be the clustering is the hill port problem because there are many different in the clustering and then trying to list them in the popular one okay and the other one is called ass the ram index which is typically used when you have the reference clustering and we have the reference clustering and I am trying to achieve the reference clustering so what is the ram index is the for the good point I don't have class labors.

So have the reference clustering but I do not have really but I do not have the reference clustering so the new comes in to the I do not have the class label so in the data points to the classes so the new set of the data and the new data point would not come I will learn new set of the data and I should run this algorithm in the in the that set of the data and then produce the clustering right.

So that the point here is I am not really looking to the reproduce exactly that cluster right what I want to do in that is the following I give you data point and run a cluster algorithm so I would like in the cluster on the training data it look like this so give me a new set of the data points to learn the clustering algorithm on it I will get the set of the cluster and then I little bit more comfortable with the t of the cluster and then the whole data point.

And then manage the reproduce the reference cluster a new set of the data points I get like actually get map to the whole data points itself so itself it slightly different problem so I am just evaluating the cluster algorithm and the clustering algorithm not the simple algorithm is produce so then we have the ram index and the one instants could be that we have nicely separated classes but many of them.

So that is the other case I am not talking about it so thermal index is typically used to every points in our data set so the F I and the F J belong to the same cluster in the reference clustering and then belong to the same clustering the cluster that they produce give the score of the one if the F I and the F j are the reference cluster and the difference cluster and the cluster that they produce and then give the score of the one and then how many pairs are there and then choose the pairs of them and then divide the some by then chose.

And the what fraction of the pairs of the data points have you cluster correctly right what is the nice thing about ram indexes suppose these are the more thing in the clustering that the original clustering is given to you right and then the end of the slitting points and then splitting in the two right the original clustering and the larger cluster and the original cluster and the original cluster.

And then this is the original cluster. and the ram index is not suffer a lot and the suffer greatly it is the only there cross and the cross cluster parts that in the phoneless within the cluster parts all fine so it gives you little bit in the terms of the lens in the terms of being in there stricter than the other than the other and the reference clustering okay these are the different measure and we can use whatever you want right because it is not the ram index and then we can use diameter or radius in the case.

Right if we have the label data points and the we can use the label purity or the entropy or the full measure of these okay so we can list about the hundred of these measure and the people are used in the literature so just pick the favorite one yeah good point so how do you know which cluster is what right so I have to figure and then I have to the alignment on the cluster right see I have some K1 cluster I have to use that is my reference cluster and my reference has k1 cluster and I have to use K cluster so which how do I align this k to the k1, right. So ran index text gets rid of alignment problem okay. Oh okay no clusters are- yeah. Yeah there are minimum phase in which you can do this optimization right.

There are literally 100's of paper out there explaining to this kind of optimization so the question is yeah what is overhead and implementing this optimization is that whether in your particular application whether the overhead is justified because you are getting a significant improvement over the usual we have doing things right. So what people do as standard implementation or things which kind of give you good results across a variety of domains?

So if there is something very specific for your domain then you have to welcome to try to this optimization right and that is what make this engineering discipline as supposed to I supposed to theory right, so yeah a many things if you can do so went back right so now I did this I do tis again I keep doing this multiple times and what do I do with all the clustering I produced, take the best yeah that is right.

So is not nothing so you just keep doing the clustering and what are my evaluation measure you have so you take the best according to that evaluation measure, okay. So we threw away the rest so you have repeat it a few times and then take the rest, great. Okay let me re-writing it okay you have to repeat k means multiple times do not please do not come to us after doing k means once and say that Oh it does not seem to work okay.

So I will guarantee you that one time it would not work okay and we do not know but just making that in fact emphatic. So let us go back to the question, how do you fix K? Okay that is one answer what is your answer, domain knowledge that is yeah exactly I mean depending on use of requirement domain knowledge right. So domain knowledge is one way of fixing K right is there are other more systematic way of fixing K.

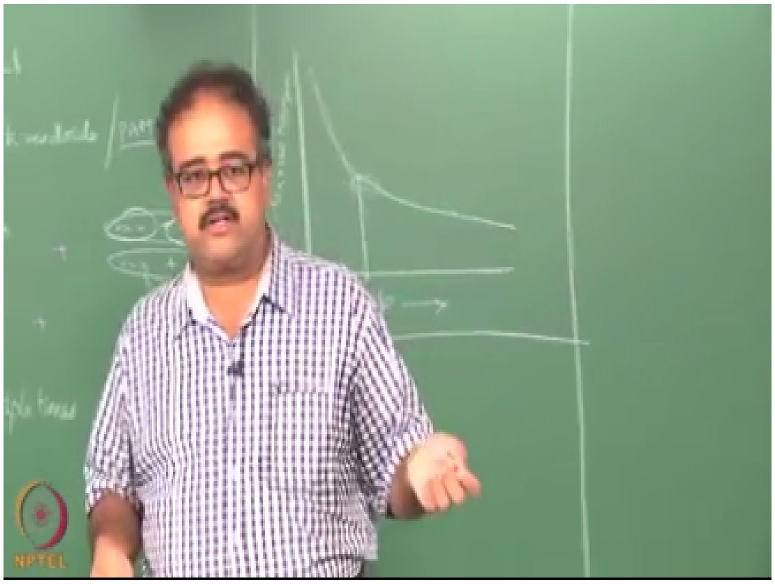
Try all case but then how do I know which k is good so I tell you one thing say suppose I am using diameter as my measure, right. Larger the case and better it is exactly so the right way of doing it would be to say that I am going to have some kind of a complexity performance trade off right but it is incredibly hard to implement in K means right incrementally hard to implement in k means.

So but you can do that I mean people have come up with that so especially if you are going to take a basin approach to cluster right if you take a basin approach to cluster right how will you implement the penalty for the size so the number of k the prior what you will do with the prior who said prior what I will do with the prior, how will you make it finalize larger K exactly, so reduce the prior probability for large K, right.

But the you do the searching the main problem with this is most of this optimization think is for finding cluster work well if I fix K but if K becomes a parameter and optimization it becomes incredibly hard that is why I say it is hard to implement right if K becomes a parameter and optimization it becomes incredibly hard to solve the problem right for the fixing a K right you can think of K mean as a approximation to solve the fixed K assignment problem right.

Even if for a fixed K if I have to do something like K means to solve it if I am going to make K a parameter and try to solve the larger optimization problem and it becomes little tricky so suppose to that.

(Refer Slide Time: 39:18)



So what people do is practical thing is draw a curve between K and let us say draw a curve between K and diameter right, so what would you expect to see? As K becomes larger decrease we will see something like that we will keep going down and keep going down yeah you keep going down to 0 when depends on how many data points you have right so if you at K = n it will become 0 I do not have to do all of them I already told you what the right value of K is here I have shown you an example say criteria what is the thing it bring, so may k initially decreases you can see a rapid decrease in the diameter right, after some point what happens the thick slows down right, so if you think about it that is a significant change of slope somewhere around here, right.

So that essentially quantifies your good curve complexity verses performances trade of, so wherever there is this change of slope right, you pick that point and say that this is the right scatting. Usually it is you do this when you are trying identify a small number of k right, if you your is very large right, then you are probably using clustering for some kind of tree processing rather than as the final means okay, so if you, we said I remember right, I told you that there are different ways in which we use clustering.

So if your clustering is very, very rare the case very, very large right, you are probably doing it using it for some kind of pre-processing right, we which case the exact choice of k does not matter that much right, when you are trying to use it as a algebraic visualization tool even visualization choice of care does not matter that much some of you using it as a actual end in itself right, you want to do clustering, you want categorization in those cases typically you have k smaller, right.

But it can run k up to a few 100 that side, right and then you can find the right value of k like this right, but if determining the right value of k is of some curial importance to you and k is very large then I would not recommend partional methods at all, right so I would recommend other approaches for doing the clustering okay. So this is called the bend now what did you say it is that name for this it is very descript and they must call the Knee method, so it bends right.

So it is called the Knee method for define k okay, so in fact you can use this for in other cases also where you want to do this kind of complexity verses performance trade off and the optimization problem because very hard if it is know in the k into the complexity parameter also if you throw into the optimization, optimization problem because very hard so can use this kind of a empirical method for determining the actual value, okay, good and we move on right.

So what about k mean writes, it is like k means so you find the centroid right, and represented and it uses the k is closes to the centroid right, so the every point of time you have a representative which is in actual data point, right and then when you do the assignment you assign it to the medoids that discloses to you, okay not a very interesting thing.

But the advantage of this is when you move from average to median mean to median what is the advantage, in statistics it does not get effected b y out layers same thing here, so when you move from centroids to medoids okay, it becomes little bit more robust to out layers right. Suppose of all this is the data points right, and I try to cluster it right, so I will end up getting a centroid somewhere there right, then but the medoid might be this a slightly better you know not that much but it is better than having a centroid that is out there, okay.

So one important thing I forgot it in an application of clustering something call out layer remaining, so what is out layer remaining that find out layer right, I mean that is nothing how big deal about it, so why do you want to find out layers, delete term is one of it anything else sorry, fraud detection right, so I want to do any kind of anomaly detection so out layers would be anomalous data points, so who said fraud detection show me yeah okay, yeah.

Yeah so the out layers would be some kind of anomalous data points and therefore you would want to find them I am not interested in deleting them from the data set where I am probably interested in deleting them from the real world right, so that I want to catch these things and put them safe guard against them and things like that. It will be useful for understanding yeah; it is the one of the initial thing I told you right. so instead of randomly sampling around the entire state space right generating 10000 samples from the million sample data base to clustering with k = 10000 and sample from each cluster, that gives you more samplings.

I mentioned that in one of the uses in very beginning okay great. So we have done k- means, we did k mean right okay, so PAM is called the partition around medoids right. In fact it is incredibly expensive algorithm, nobody uses PAM any more. When it was proposed it was very big thing and but then people came up with faster ways of doing PAM, all of these work on very small data sets okay. Really large you want to do 10million data points things like that. PAM is nowhere near competition, any of medoids or not at all competitive.

On a very large data sets and any way I will just talk about it because it was very interesting algorithm right. So in partition around medoids, so what you do is okay. So I have some data points, I start by assuming some, say I am doing some clusters, I will start by assuming some two dada points as my initial medoids right. So let us say that this as 1 medoids right and unfortunately assume that this is the other medoids right. Now I look at the quality of the clustering, let us say I use radius as a measure for the quality.

So I will assign all the data points close to right, end of this is one cluster and I look at the average distance of the data points to the medoids count keep that as my point to medoids, I will keep that as my quality of clustering, right now what I will do is for every medoids right, I will consider swapping it with the non medoids right. So I will say, I will make this a normal data point right that we make that a medoids. Now if I make that a medoids what is the change in the quality of the clustering?

Likewise for this medoids I will consider each non medoids intern and consider swapping with it, or whichever gives me the best improvement in clustering, I will keep that as the my new non medoids and then I will go look at the other medoid. Now I will consider swapping this with each one of the data points intern right and then I will be swap it here. Sorry anywhere, so this how it works, it is very expensive as I told you right. So for every time you do the swapping you do the order n thing.

I just have to check the distance, so checking the distance rather order in computation right, so essentially I end up doing n^2 computation for very swap, that I have to make. People made all kinds of interesting observations and they came up with the ways of cutting down on the number of the computation at you do peroration. So when I make a swap I do not have to go through each and every PAM, so only those change cluster membership. Only those data point change cluster membership I have to really look at it.

Data points belong to the current cluster right the medoid of that change; obviously I have to do re computation for that right. Among all the data points do not belong to the clusters, only those change clusters memberships, I have to evaluate this right. But then I have evaluated the cluster membership anyways okay. So I do not have to do new things but still I have evaluated the cluster membership. So but then you can again organized it little more efficiently so for.

So depending on how you organize this computation people come up with a variety of different things that is PAM I can remember, that is partition is on medoids right. People are interested I can give points to read out on more PAM things like that, like I said it is not very that widely used in the community so we will skip those things. So what is the problem with K- means PAM addresses? Initial random really does not matter anymore because I am any way considering everything; I am any way considering very possible pair gone.

Then well it is medoids it is no long that affected out layers right, what about the choice of K, we still have to choose k that is still there, it is not gone away right. Right what about the issue k-means if you have real value attributes, works well if you have attributes, if you gone rid of that. I still need a distance measure right. So if I am going to have categorical attribute better have a distance measures that takes care of categorical attribute, still the close problem remain with PAM okay great.

IIT Madras Production

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

www.nptel.ac.in