NPTEL

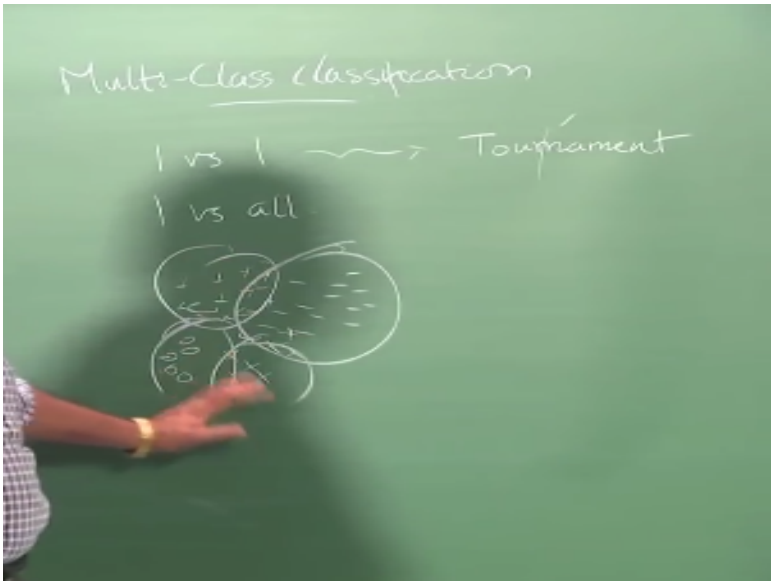
NPTEL ONLINE CERTIFICATION COURSE

Introduction to Machine Learning

Lecture-86
Multi-class Classification

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

(Refer Slide Time: 00:15)



I want to talk about multi-class classification and so there are some classifiers that we looked at which are naturally multi class classifiers right, which are they you know networks yeah with a little bit of work here they are multi class something which is more immediately multi-class this chantry is immediately multi-class right no need to worry any fiddling around with anything else Navy base at all the Bayesian classification that we looked at read all those are immediately multi class classifiers and then there are things which we looked at which are inherently to class classifiers right.

SVM's is one of those every popular of those and any other two class classifiers that you know logistic regression is inherently a two class classifier but we have multi class variants of it I did the two class, classifier in detail in class right but there are multi class variants of autistic remission but again the two class one is the one that is best understood right and any kind of discriminate function based classification that we looked at right or inherently to class classifiers right.

You mean you can think of ways of converting them into multi-class classification but inherently to class classifiers right so suppose I give you a very powerful mechanism for constructing a binary classifier, can you solve the multi-class classification problem using that let us make it even more confident I give you an SVM right I will give you this packaged code for an SVM right I am telling you this is the best possible SVM implementation but it does only binary classification can you use that and convert it into a multi-class classification think how do you do that what is the advantage of one versus one potentially be balanced right hopefully I mean depends right so you still have an unbalanced class classification problem I might have 30 classes in which each class has 10 data points and one class has 10,000 right.

So then one versus one will be a problem then right but one versus all will always be a problem even if you have equal number of data points in every class one versus all will be a imbalance problem right but the disadvantage of one versus one, how many classifiers you need in one versus one n choose 2 n choose 2 ray so that is a large number of classifiers I give you a hand 100 class classification problem right so how many classifies will you need within 1 versus 1large member right.

So you are in the time and we actually want to the classified we need not run them all like at the time interval timer feel to the creation which would probably be a one-time process yeah we would need all okay and when you run when you actually want to classify how many do you need how many would be actually if I whenever you guys while you get rid of but one person any person do you how come because you for example if you have class A B C and D huh, I will run the classifier for a versus b a versus c a versus b then you run it for a versus B you through one of them out then suppose it was yeah b versus c and one come over here.

You could do that right but then for that you have to be little bit careful because the guarantees that you would have or slightly weaker right so this is really not called 1vs1 okay so that is one

version of running 1 vs 1which is a caller tournament is essentially what you are suggesting is run at ornament right. So you train lot of 1vs1 classifiers and then you run a tournament so you need to train that many classifies but we are deploying it you will have fewer numbers to use right.

But the problem with that is suppose you are a versus B classifier was weak right then you would throw out a incorrectly and suppose we are running a versus B and then b one but if you have done here a versus c and a versus d also a might win against all of those right and then be might lose to c and d then it becomes an issue right so a would have gotten two votes well be would have got only one vote and then what do you do right so classifies are good then tourneys or be great if classifiers are weak and you have problem in tournament that you might eliminate things a little early right.

And then another problem in the tournament is you can identify only the most likely class but if I wanted to give ranking of class labels okay that cannot be done with tournaments okay but if you have hundreds of class labels right, so you have some you have to give up on something so essentially give up on the correctness and they essentially try to run a tournament if you have a lot of class labels try running a tournament on this right.

So the psychic learn implemented tournament automatically you know in this one versus one yeah that is that is fine what about SVMs? It supports multi class SVMs but there is nothing called multi classes here we have to do one of these I will take it back that is a multi classes fear we will take it back but there is not work psychic learners yeah so but this might be something you might want to employ okay now going back.

So I told you that it is possible that we have severe class imbalance in a 100class classification problem even right, so you have one class that has like a million data points and each of the other classes have like a thousand data points right so what would you do in that case so we spoke about some ways of fixing the problem right the class imbalance problem so weighing some one class more than the other under sampling over sampling did we talk about this at some point I vaguely remember discussing class imbalance in the class, no we discuss class imbalance in the class right.

So there are different ways of fixing that you could try that alternatively you could try some kind of a hierarchical classification right, so what you do in hierarchical classification is you essentially try to split the classes into two groups okay. And then say that okay first level I will see whether it goes into group one or whether it goes to group to the next level or in within group one I will try to assign it to a specific class right or I can split it in to groups three and four and then within that group three I will assign it to a specific class right, you could do some kind of error K classification right.

So what is the challenge here? So sorry choosing the hierarchies yes using the groups right so unless the groups come to you somehow from the domain itself right sometimes you could have like people classify web pages right and then you can go and look into some open directory project or something like that and there are nicely classified web pages for you right. So you have a hierarchy of web pages there you can then look at the classification down the hierarchy.

So you will start off with saying okay entertainment versus news and then within news you could have say politics sports and then within entertainment you will have movies and I do not know and sports comes into the news or entertainment. So wherever right so you could I will have this kind of hierarchies and then you can use this hierarchy to give you your hierarchical classification right in the absence of that right how would you want to do this?

If you want to induce the hierarchy I give you a flat set of 100 labels if you want to induce a hierarchy on this hundred labels how would you do this? Hum, based on the number of data points, clustering it is clustering so what do you do with clustering how do you do clustering in this case just cluster of the data points blindly that person you are essentially solving we do not know which way to solve it right so how will you cluster it here people are throwing up all kinds of terms now right.

So the point is you have, so the intuition is the following I have this class conditional densities I know okay this is class one what are the data upon this is class to what are the data points right, so I would like to group them in such a way that the class condition densities belonging to one group are very different from the class conditional densities belonging to the other group right does that make sense right.

Suppose my data is like this right so I have all my class one here class two here class four here and class five data points are here right four okay, so which what is what is a grouping that suggest to you itself suggests itself to you one and two should be one thing and three and four should be the other right. So if you think about it so this is the class condition if we are assuming these are drawn from Gaussians and I will have a mean here and some variance over this and the mean will be somewhere here and some covariance oh these look closer right then the means of these distributions.

So that is a basic idea one way of achieving this is to say that okay I will do clustering right and then I look at the class labels that fall together right which class labels fall together more often this is very nicely done right. So what if my classes are actually like this the classes look like this now it is harder to harder to separate them out right, so what you can possibly end up doing something like this and I find some clusters some groups of data points like that.

Now predominantly in this okay I have class one predominantly in this I have class two predominantly in this I have class three in predominantly this whatever prominently in this class three class four now I start looking at which is which of the clusters are similar and then I can do some kind of predict which values the training data is given to you right. So you have the training data you are just being clustering on the training data.

So the training data will tell you what the class labels are right, this is there is this really a formal way of doing it I am just giving you tips practical tips forgetting addressing some very large problems, sorry so I have done these clusters now right I have clusters and I can I can figure out which clusters are similar to which cluster which clusters are close that I have some description like suppose I am using some kind of a Gaussian model for describing my clusters I will of some description of the clusters right.

So I can I can now figure out which cluster is close to which cluster I will talk about hierarchical clustering depending on today or Friday, Friday stands for next class okay. So today or Friday I will talk about the hierarchical clustering then you can see that okay there are four clusters and then these two clusters get merged first and then these two clusters get merged so at that point I can say hey okay now I am going to say all the classes which are more prevalent in these two clusters should go together the classes that are more prevalent in these two clusters should go together.

The data points it belong to those clusters you go and build a classifier right first classifier you build on all the data points that separates these two clusters from these two clusters that spacing so I do not want to do a distinction at the very beginning I do not want to do a distinction between this and that this and this and so on so forth right. Then how does this help us in class imbalance?

Yeah what if you originally you had class imbalance what if originally 1class had a million points and all the other classes had a thousand points each, yeah if that is supported if the data supports that million and this is extreme but say ten thousand to one class about the video get real data like that. I did not get it, we were making clusters right on the glass labels so it does not matter what the size of the cluster is so all points belonging to same level would fall into in the same cluster.

How it goes is class imbalance, no it will not cause imbalance I am asking how will it relieve you from chi class imbalance we are getting is less than has only see I am not I am not using the clustering itself to do the classification right I am only doing the clustering turn that I can group the class labels and then I go back and try to solve the classification problem after that yeah. So I suppose all of you are going to try out different things.

IIT Madras Production

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

www.nptel.ac.in