NPTEL

NPTEL ONLINE CERTIFICATION COURSE

Introduction t Machine Learning

Lecture 52

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

Minimum description length &
Exploratory Analysis

So in minimum description length principle the idea bandit is very simple right among equally we already looked at it in some form or the other okay, so lasso is one way of thinking of it as minimum description right among all equally good classifiers I would like to pick the one that requires the least amount of bits to describe it right. So the description length should be as small as possible given that it has some acceptable levels of performance right.

So if you think about it so what does this tell us well if the classifier is very complex then I am going to need a lot of bits to describe it right if a neural network with a lot of weights the support vector machine with a lot of support vectors right or a decision tree with a lot of branches, so the more information you have the more detail the classifier is the more number of bits I will need to describe it right.

But then the better that it gets in performance right ideally why would you want to make it more complex only because it is making fewer errors right then you have to come up with some way of trading of this the description of the classifier versus the error it makes right ,you have to have to specify what the classifier is you mean you have to decide on how you want to specified suppose it is number of support vectors right so I will have to tell you what are the individual support vectors you remember support vectors or α times the excise right xi yi right.

So I need to tell you what the xi are I need to tell you what the α are for me to specify a SVM completely to you all right. So how many bits do I need for specifying those α and the xi, so the xi yi I can take it as a product and I do not have to describe the way I separately but I need something to describe that and maybe to describe the α also and or maybe I can describe α xi to you if you can use that somehow to produce the inner product but if I have it there the kernel version of it then I cannot do that right.

So I account pre multiply the α into that right if it is a linear thing I can do this I can give you α xi so there are things to think about it so how do you encode this right, so you want to write a program to implement SVM in end of the day right what is the point in me doing a learning algorithm and then not letting you use it right. So for me to communicate to you how we implement it I need to give you the description right.

And the second part is the errors are there right, so I make some mistakes right and I have to tell you on a training set let us say there is a fixed training set and on that training set I also want to tell you what are the errors I made right the smaller the amount of errors I make the lesser the number of bits n heat to tell you how many errors I made, makes sense. So for me to make small errors I met need a complex classifier that will need more bits for me to describe the classifier right.

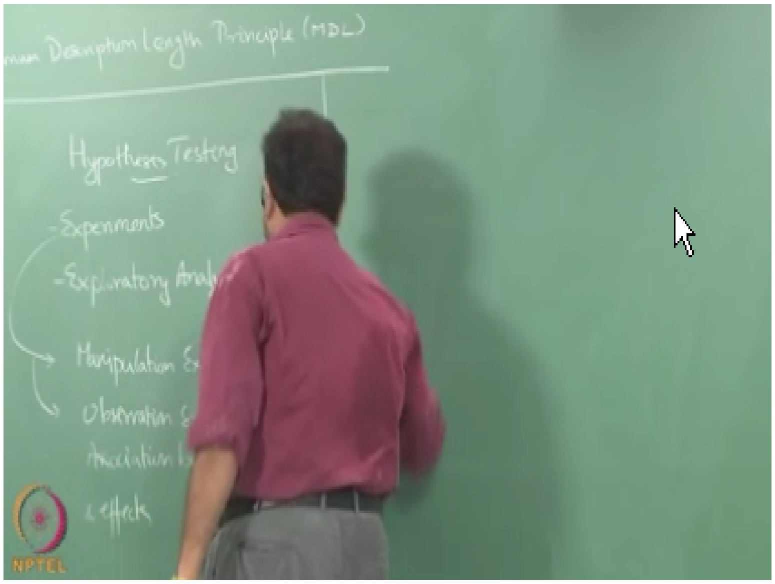
So if you want to reduce the number of bits I want to describe the error I might increase the number of bits I want for describing the classifier and vice versa, I am sorry just to set up a trade-off on equal footing right I am just talking about information on both sides now right. So that is I can make our classified arbitrarily complex right and we keep giving me better and better error right or I can make a classified very simple than it can give me a lot of error so how do I do the trade off ?

Between discrete the size of the classifier and the amount of error I make, so to put these things on an equal footing so that we can compare let people talk about the amount of information required in both cases you need some amount of information here and some amount of information here. So the more information you need here the lesser you need here, so that is the trade-off right.

So that is the idea behind minimum description length right and this huge theory behind it right it is actually a proper Bayesian approach and we talked about Bayesian learning at some point we also talked about ml and map estimates and so on so forth you can show that MDL is actually a proper Bayesian approach and people have derived a lot of complexity measures performance measures based on MDL right.

So I never talked about it earlier that but I think you guys are all ready to now read up on your own about MDL right so the brief introduction I gave should be sufficient right.

(Refer Slide Time: 04:54)



So if you think just one minutes top and think about how people actually use machine learning right it is very heavily empirical right when women doing a lot of math and other things or pseudo math in the class so far right but really at the end of the day when you start using it right it becomes heavily empirical it is actually a very applied subject believe it or not I mean of course you guys are all finding it out now with all the programming assignments but it is actually a very, very applied thing.

So whenever we have these kinds of empirical work right so you have to do our experiments right you really have to do experiments there is nothing like you know an analytical solution to your machine learning problem right. So when they give you a data set right you really have to experiment with the data to figure out what is it that you are going to do so all the theory and everything that we study now is all fine but when you actually get down to doing something you have to run experiments you have to do all kinds of things.

So you have to do experiments you also have to do some kind of some kind of accelerated analysis. So in fact we have not really talked about exploratory analysis at all in this in this course I have to do a lot of different things, so I actually do that whenever I teach my flavor of data mining I do that so what do you do with external analysis right so there are many things that you have to do first you have to figure out.

So how distributed variables are you know so we have to figure out so what is the range of the variables I give you data right I do not do you do not really know what the data is all about I just give you an the simple form is I give you an excel file at the complex form is I give you like few terabytes of data on a disk right but then let us say I give you a file and then you have to figure out what are the different variables are there right and what kind of values do they take right and what is the variance of these values right or their outliers on these is that some values that I can ignore right.

So the whole bunch of things that you have to core and do some kind of exploration right or that way variables that are important to my prediction maybe know about then we talked about some variable feature selection and things like that but all of this you have to do essentially you have to understand the data before you even think of what is the machine learning algorithm I am going to use.

So if I give you some data you do not just straight away plug it into a decision tree algorithm or straightaway plug it into an SVM right so you have to go around try to understand what the data is all about right. So that is part of it is through exploratory analysis and say a little bit more later but as far as experiments are concerned typically they fall under typically they fall under two kinds of experiments.

So there is the manipulation experiments and observations experiments to who do you think this mean so in observation experiment I basically try to figure out correlations I try to figure out associations between variables right I would make a lot of observations and then I try to say okay if this variable whenever this variable was at this level then the output was at that level right so maybe I can observe the wiener, so it is essentially trying to find associations between right associations between factors and effects.

So what do I do in manipulation experiments in manipulation experiments is essentially these are things where I have some control over some of the variables that constitute the experiment right, so typically I set up these kinds of manipulation experiments whenever I want to test a theory a causal hypothesis right whenever we want to say that a causes B right. So I cannot stop yelled you know from happening right.

So I mean those kinds of that does not make any sense but I can say something like okay, so by learning algorithm A is better than my learning algorithm B right whenever the load on my system is high right. So I can actually make a hypothesis like this that A is better than B whenever the load on the system is high right, no I have made my models I have model A I have model B I want to make a make a statement that saying that yeah is better than B forget about under heavy load.

I want to make a statement that okay you have a learning algorithm A he has learning algorithm B I want to make a statement that learning algorithm A is better than learning algorithm B. So when will I make such a statement when can he make such a statement haha right, so that is the whole thing that we are going to worry about here I am going to test what I mean what I call.

IIT Madras production

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

www.nptel.ac.in