NPTEL

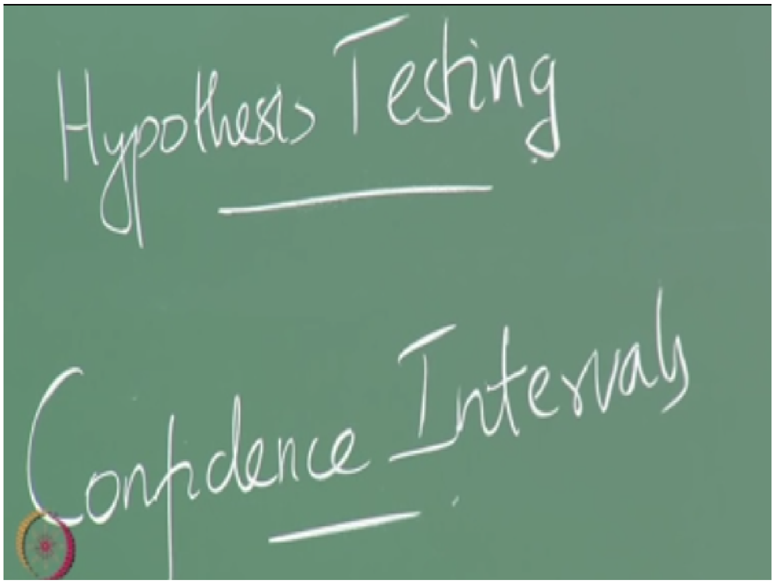
NPTEL ONLINE CERTIFICATION COURSE

Introduction to Machine Learning

Lecture-58
Confidence Intervals

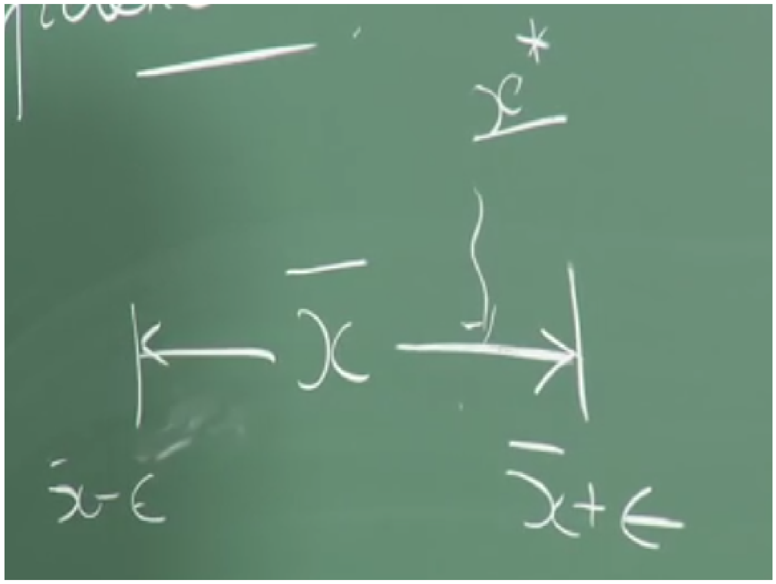
Prof. Balaraman Ravindran
Computer Science and Engineering
Indian Institute of Technology Madras

(Refer Slide Time: 00:17)



Confidence intervals right so we talked about confidence intervals all right so essentially the question I want to ask is the following right. So if you think about it what we are doing is right we are trying to estimate some parameter some performance measure by looking at some statistics that we compute on a sample of size n right, so that is basically what we are trying to do be trying to measure some performance as on a statistic in a sample of size n right.

(Refer Slide Time: 01:18)



So I am doing this repeatedly right suppose I have done this with one sample and I have some number let us say X bar okay let us say that is the average error or average whatever okay I am giving you some performance measure $\bar{x}$ right so in what fraction of samples of size n right I draw this $\bar{x}$ and then I will give you some interval around $\bar{x}$ so right and I will give you some interval $\bar{x}$ - $\bar{x}$ and $\bar{x} + \varepsilon$ right.

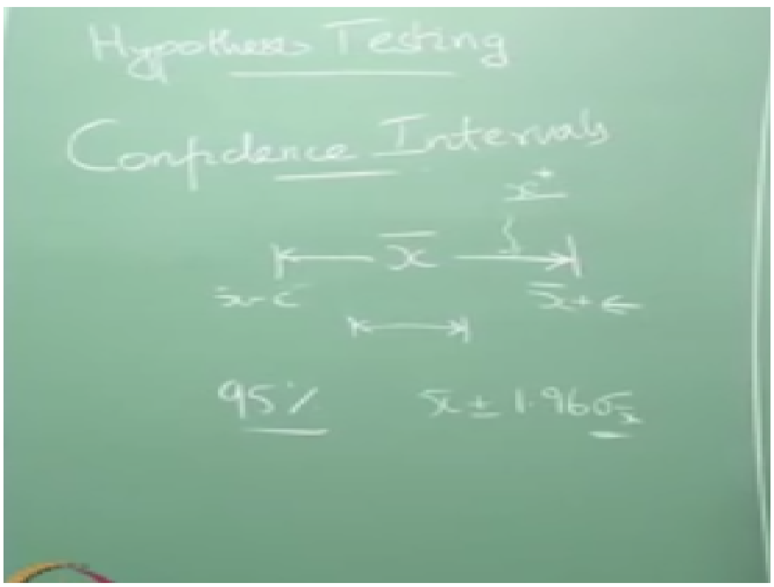
So and there is some true performance measure I do not know called x star right so ideally I would want to give you this plus and minus ε such that X star lies somewhere in this interval right so I give you dot $\bar{x}$ I give you $\bar{x} \pm \varepsilon$ says that with a high probability I want my x^* to lie within that interval okay, so in fact the confidence interval essentially the amount of confidence you have in this interval essentially means the following okay.

In what fraction of the samples of size n that I draw suppose I keep drawing samples of size N and I tell you that I have give you a ninety-five percent confidence interval right so what does this mean exactly 95% of samples of size n, X star will lie within the spar would lie within this star would lie $\pm \varepsilon$ ($\bar{x}$) right so right you understood what I say right in 95% of the samples of size n right x star will lie within $\pm \varepsilon$ ($\bar{x}$)okay.
Is this is the same thing as saying that with ninety-five percent probability the x star is within $\pm \varepsilon$ ($\bar{x}$) no, why? Could I am talking about samples of size n here right so depending on my sample size may sample is very large right then possibly this will approach that probability I am talking about samples of size n okay so whenever I give you a self-confidence interval remember that it

really does not mean even though people often mistake it for the probability of $\bar{x}$ being within ε (x^*) is really not the case.

What it really means is if you repeat this with samples of size n right in ninety-five percent of the samples X^* will lie within ε $(\bar{x})$ okay some kind of an assurance right but not suppose I want to reduce the confidence interval what does it mean, sorry what does it mean to reduce the sampling I mean confidence interval reduce the ε right I want to reduced ε I want to make it smaller right so you say reduce the confidence interval I really mean that okay I want this as opposed to that right. If I want to do that what is the best way to do it increase the sample size n, right.

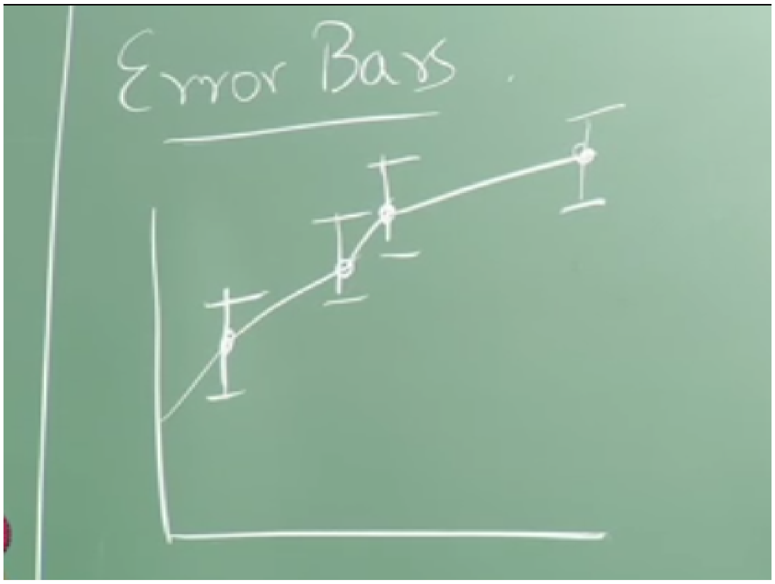
(Refer Slide Time: 05:25)



So if I want a ninety-five percent confidence interval right so what would that be, we look at some magic numbers in the last class increase from this is z, so we cannot use the standardized thing here because we really need to give actual values here right so we cannot use the standard normal Gaussian so it has to be 1.96 x σ $\bar{x}$, so it is 1.96 because it is 2.5 that side 2.5 this side right so it is1.96 right.

So if I want a tighter confidence interval right then essentially I have to reduce the σ right. This is assuming that your end is fairly large right if a tent is very small then you have to use the T distribution you cannot use 1.96 lot to use the corresponding statistics from the appropriate entry in the tea table so appropriate table find appropriate row in the tea table corresponding to the degrees of freedom right so typically for n > 20 right. You can use even something simpler 1.96 or even two times σ is good enough so related to the confidence interval right.

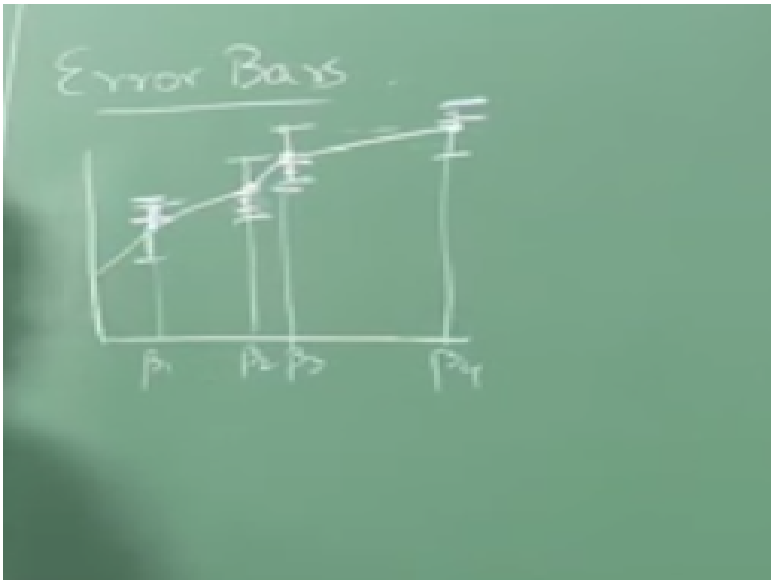
(Refer Slide Time: 07:24)



You also have this notion of eradiate what error bars are what is that somebody said something remember standard errors what is it standard error yeah but what is it really it is see it is the variance of the standard deviation in the sampling distribution right that is what we call the standard error so error bar are essentially things that you plot around your estimates so that it tells you what is a what is the variance that you are likely to see in the estimate that you are getting.

So typically what you do is right so you make some estimates and then you try to make some plot right I am varying some parameter then I say okay and that is how my performance varies right so instead of just plotting these points and trying to draw a curve right what I had like you to do is essentially give me a error bars around that right so each of this point I would have run an experiment I am varying some parameter here right and I am looking at the performance right some parameter I do not care what it is and looking at the performance. And in each of this point I would have run an experiment right.

(Refer Slide Time: 09:12)



Assuming they are all equidistant around the thing right each of these outer run some experiment and for each of those I can give you the standard error right so I plot these error bars now the question that you have to ask is okay from let us say I have these values some values ß 1 ß 2 ß 3 ß 4 right so I have just run it at some intermediate points I have these curves right, so can you tell me if ß 2 ß 3 is there any difference in the performance between ß 2 and ß 3?

Not really because my error bars overlap significantly and I cannot be sure if there is a difference between the performance of ß 2 and ß 3 just on evidence of this curve alone right what about ß 1 and ß 2, No right ß 1 and ß 2 also I cannot say that that is a actually a difference what about ß 1 and ß 3 barely ß 1 and ß 4 surely what about ß 3 and ß 4, not really right so I mean so yeah in the means are different if I just gone by means I would have probably said that ß 4 gives me better performance and ß 3.

But on the evidence of the experiments that you have run so far I cannot conclude that because the error bars significantly overlap right so this is why whenever you are running empirical series you are always supposed to plot these error bars if you just give me an average performance right it is not at all clear so if I am comparing two things then I can run your two t-test and so on so forth this gives you a rough idea of which of the performances are actually different.

So ß1 and ß 4 are certainly different that ß 4 is certainly better than ß 1 so for other things evidence is kind of shaky sorry, good they could so the way for you to verify this is now go run more experiments with ß 1 ß 2 ß 3 try to see if you can get a better estimate because as you know that so with the now the true estimate could be anywhere in this interval right so rather with ninety-five percent of the cases that to estimate could be anywhere in this interval so what I do is I run more experiments so if I run more experiments with ß 1 I might actually see that my mean shifts here right and my confidence interval becomes much narrower.

But now remember this is not the same confidence interval as it before because this is a confidence interval of a larger sample size look you cannot directly compare these two it is a confidence interval of a larger sample size so I might actually rerun the experiment and this might be the values I end up with this if there is a better color that I run the experiments again with a lot more data and we can see that now things are little clearer.

So ß 1 ß 2 there is really no difference right they are the same alternatively ß do could have moved up ß 1 could have moved down and in could anything could happen and this giving an example here where no ß 1 repeated or almost like more likely to be the same and ß 3 is certainly better and ß 4 is certainly better than all the other three that this could happen one potential scenario another potential scenarios this could move down this could move up right.

So it could whole thing could change right essentially what the error bars tell you is what kind of conclusions can you draw from the experiments you have done so far could very well be that it is enough for you to find out which is the best ß, ß for seems to be the best interms of the experiments that you ran even the first time around but if you want to produce a ranking among the ß you will have to rerun the experiment.

The only conclusion you can make from the previous experiment that you had was that ß 4 is probably the best ß best value forbidden if that is all you are interested in finding out you can be happy with that experiment but if you want to produce a relative ordering of the parameters then you will have to be more careful so that is essentially the use of error bars.

IIT Madras Production

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

www.nptel.ac.in

Copyrights Reserved