Practitioners Course in Descriptive, Predictive and Prescriptive Analytics

Prof. Deepu Philip
Dr. Amandeep Singh Oberoi
Mr. Sanjeev Newar
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur
National Institute of Technology, Jalandhar

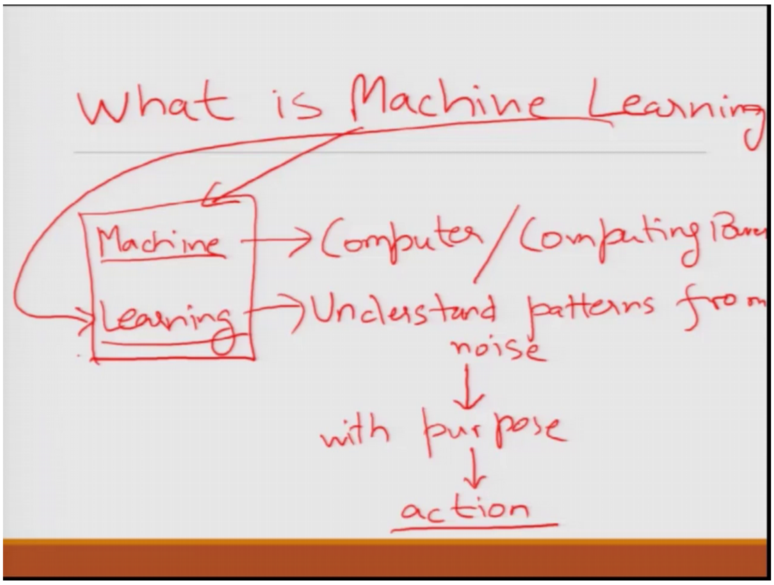
Lecture - 19
Machine Learning

Hello everybody, welcome you all to this lecture on Machine Learning, we will be talking about tricks straps and trips for a practitioner of machine learning algorithms. And I am sure you would be also very excited as I am about this machine learning lecture because machine learning is the cool thing these days, you will see a lot of tutorials, a lot of training material which is available online and elsewhere on machine learning and there is lot to learn from there. In this lecture, what we will do is in this first part at least what we will do is try to create an intuitive framework and an understanding of what machine learning is all about.

And then we will go about some of the algorithms; our focus will be more on the practitioners side. And when I talk of practitioners side do not just mean the users, but an intelligent user a user who uses these machine learning algorithms for smart decision making; rather than as often we see the trend that people adopt some of these algorithms just because they are cool or they are in fashions.

So, we will talk about what are some of this pitfall. So, that you can evolved yourself to be an in lighten machine learning users rather than just being a machine learning user. If you want to get into more maths of it I recommend book by Trevor Hastie and Robertive Sherani; Professors of Stanford university called Elements of Statistical Learning, but that will be very Mathy book a lot of maths in it, but if you really want to get into depths; I believe that amount of maths will be important because then you will be able to understand why certain algorithm works in certain situations, why you should not use certain algorithms even though it sounds so, cool or there is so much of craze about it.

(Refer Slide Time: 02:18)



So, without further delay let us start. So, before we even begin let us try to understand what is machine learning? You will find a lot of definitions floating around and lot of people saying this is machine learning, this is computational method this is statics, this is statistical inference and regression and a lot of things. So, let us evolve or let us have from practitioner's perspective a very pragmatic definition.

So, if you look at machine learning there are two parts of it number 1 is this word machine; now what we mean by machine basically is a computer or I would say computing power in general. And this is the second term the more interesting term which is learning.

Now, what is learning? You can come with a lot of definitions lot of philosophical definition, lot of operational definitions. I would put learning as the ability to understand patterns from noise. So, if you recall I mean I am sure you will not recall, but when you are child as a baby; the entire world you know you were born myopic you did not know how to focus our eyes on certain things and the entire world was a noise gradually the eyes start learning how to looks.

So, basically focus on certain things; ignore rest of the things as the noise and hence understand when you start reading at first you know and even for illustrates; if you if you do not know a particular language for that language basically all you see are some black and white patterns all around you cannot make any sense. The moment you start understanding more meaningful patterns out of those blacks and whites you say that you now know this particular language you know this particular script.

Similarly, there is a lot of data all around us when we are able to make patterns out of the noise or out of the so, much of data that we see or sometimes unless of relevant data more of noise data. I mean even this word relevant where I have that I am using assumes that you have at least some intuitive sense of what is useful in the data?

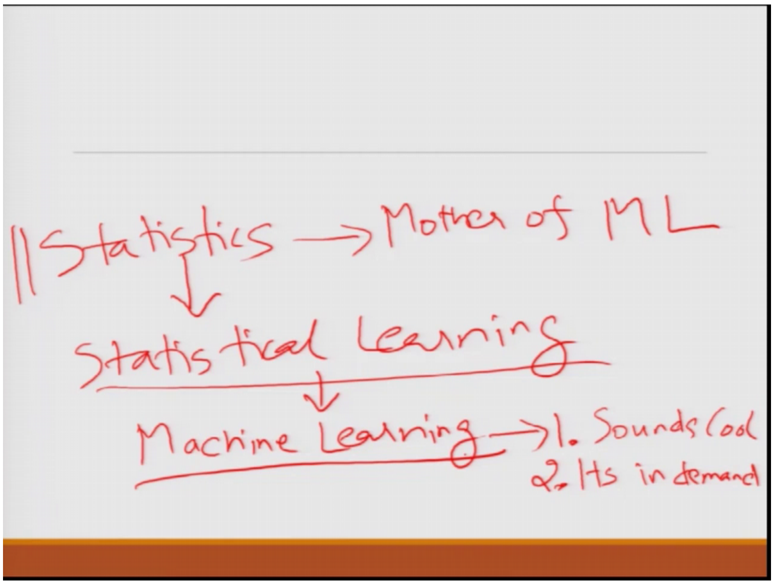
So, basically when you have a huge amount of information data it can be a 1, 0 bit; it can be any kind of information. When you have to figure out and you want to make meaningful understanding meaningful; interpretations out of it that is what is called learning. It is about putting those schematics in place, putting those probably even mnemonics in place, putting those frameworks in place so, that when data appears in particular form you are able to make sense out of it. And learning not only means you just understand the patterns, learning has something more to it ; you understand patterns with the purpose and purpose for a practitioners perspective balls down to action.

So, if you ask me to explain what machine learning is; it is all about using the ability of a computer to make sense of a huge amount or less amount or whatever amount of data that we have. So, as to derive certain patterns out of it which helps us make better decisions or take useful actions.

So, basically this is now even analytics is in general also about that. The slight difference is that when we talk about machine learning what we are saying is that we put those algorithms and put them largely on autopilot mode. So, for the first time when you build the algorithm you may have to do a lot of dirty work, but after that once you have built a particular algorithm; you have connected with the particular source of data you just apply the algorithm and it gives you the result.

So, often the implementation of machine learning is on an autopilot mode.

(Refer Slide Time: 07:39)



So, there are these three terms number 1 is statistics; this was I would say is other of machine learning. From statistics as computers evolved and we started using a lot of statistical techniques through computers, using simulations using a lot of other advanced thing because in the classical era of statistics; so, lot of these calculations would have to be done through hand or through complex tables and stuff the moment computers came it to all revolution likes the thing.

And from statistics we evolved into something called statistical learning which is nothing, but using some of these principles of statistics along with the computer to get all the learning and from statistical learning today we are in the era of machine learning.

So, what is the difference between statistical learning and machine learning? Well I would say the first difference with it sounds cool number 2 it is in demand. So, technically it is statistical learning, but industry loves the term machine learning industry enjoys calling it machine learning, there are lot of huge amount of job demand surrounded lot of craze around it, lot of projects around it a lot of excitement around it. So, we better stick to the word machine learning and of course, it sounds cool as I said.

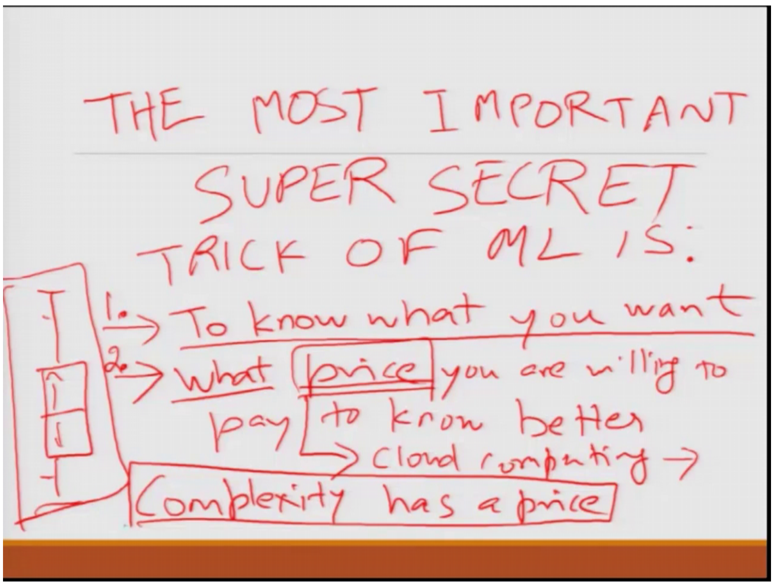
So, before I dwell into machine learning algorithms and stuff in detail because there are a huge amount of algorithms that have come in machine learning and every day as we speak you know every one moment; somebody is improving upon some aspects of algorithms. So, that is a huge amount of work that is happening and you have in machine learning algorithms from very basic simple stuff to the most complex; nowadays we talk about deep learning and all that stuff.

So, there is a huge amount of plethora of algorithms and books after books you read about it you will still feel that you do not know you were a bit about it or you just have scratch the surface. So, in this entire forest of algorithms that we have by the way we also have an algorithm called random forest.

So, in this entire forest of algorithms that we have; we need to have some kind of an understanding I would say a common sense or I would say an analytical brain, an analytical intellect to decide what algorithm will suit our purpose. And often the reason I have emphasising on this is that in industry, you will find that many a times just because some algorithm is very cool for example, these days deep learning is very cool even industries even companies even organizations do not need deep learning ah; they are implementing their spending millions and crores on this the results that do not speak of themselves.

And normally this happens in IT that there is a craze about a particular term that became popular which was a rehash of something which was existing for quite sometime may be slightly modified a bit and then you implement trying to be state of the art and you do not get the results, then you get disappointed and then the cycle of economy falls down and up and we face all the consequences.

(Refer Slide Time: 12:06)



So, when you implement as a practitioner I would say the most important lesson if you ask me; the most important super secret trick of machine learning is and holder breaths. The most important super secret trick of machine learning is to know what you want, you do not implement an algorithm just because we have shared an our code with you and you find it cool.

You need to be very clear why you are using a particular algorithm? Why are you even doing machine learning in the first place? I do not know a lot of analysis you do not even need machine learning to do if the if you do not need that kind of an automation a simple box plot even itself, if you ask me even a box plot is a machine learning algorithm very simple machine learning algorithm.

And if you can just automate certain patterns of a box plot on where the median lies what is the range? Is the median closer to the upper or the lower part; what are the kind of outliers that you are getting and you can design rules around it. And this in practice can form as a better a more robust machine learning solution then some of the advanced techniques that we will be discussing.

So, the first and important I would say the most important secret is to be very clear before you get into any project on machine learning. First of all clearly define your questions that you want answers to; first of all very clearly define what you want to do, what are those perplexing patterns that you want to see out of this? What exactly you what is the purpose? So, and this does not end here; the super secret trick has two parts this was the first part. The second part is what price you are willing to pay to know better ; so, you want to know something.

The second question is, what is the price that you want to pay to know that thing better. See knowledge is not an absolute 1, 0 binary code in machine learning when you see patterns, there is always a pattern and there is an error associated with that pattern. So, what is the kind of accuracy you want? How much of detailing you want? How much of stability you want in your model the service, how much of perfection you want in terms of the accuracy with which the predict something.

Now, these are questions that you have to answer now these are not philosophical question by the way today; because most of these machine learning usages these days happens in what we call cloud computing domain. And when you talk of cloud computing domain to a practitioner apart from all the other bells and whistles that it refers to; for a practitioner it means that any calculation, any activity that you do on a cloud is chargeable.

You are charged in proportion to the amount of computing resources you use, you are charged in proportion to the amount of time that calculation takes, you are charged in proportion of the amount of RAM that calculation takes the amount of data if you are saying I want to analyse big data; the amount of tedious of space you need to store the data. So, and that comes at a price nothing is free.

So, there is a price there is a price in terms of computational complexity, in terms of the amount of data you need for that analysis and depending on that you are price actually increases. So, that is something I would say is simply the price of being on the cloud, but apart from that machine learning algorithms themselves and as we will see; there is a price when you try to make algorithms more and more accurate or you want to get into more and more details that is the price of losing the future predictability of a model. So, we will talk about in detail.

But the point is you need not always consider that is because this is the more sophisticated algorithm I have. And using any algorithm on say R or if you are using python or your losing any other kind of software is about in working one single line or two line command at some point of your entire program. So, it should not be that because of course, I believe that I can create a deep learning out of it and I have this Google’s and tensor flow available to me for free.

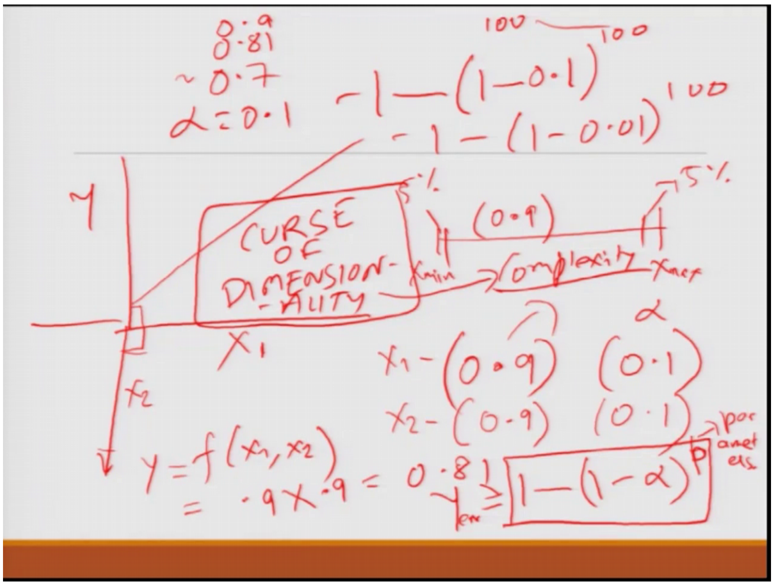
So, why do not I create a four layer deep learning network well you can create, but then to support the deep learning you need those kinds of resource hungry servers or GPUs; you need to hire that kind of a cloud platform to carry on the calculations. You can do a support vector machine it will be computationally complex slightly less complex than deep learning, but nonetheless compute support vector machine will also be complex. Or you can do a very simple regression or a simple binary classification based on some simple rules.

So, depending on the problem; you need to be very clear about this. So, you always keep this in mind that complexity has a price more complex the model you make, more the price you have to pay. To give you an intuitive understanding of this complexity, the price of complexity and this I want to do this because I have seen many people even despite all these disclaimers and all these warnings you know the the the the urge of being state of the art sometimes so, much that we tend to lose this part.

And we have an inherent desire to create more and more complex models because perhaps as human beings we have the tendency to be able to understand more and more complex things. So, we like to if we are working with one complex beast today you want to have something more complex tomorrow. So, that is the human nature and in machine learning actually I would say the struggle your success as a machine learning expert would be on how much you can resist this temptation.

So, let me let me just give you a very rough intuitive example on this which will not translate exactly, but you will get some kind of a field.

(Refer Slide Time: 20:21)



So, for example, you make a model where you are predicting some Y compared to a very well X let us not get into details of what these X and Y are or for example, we talked about simple regression last time. So, consider it to be a simple regression or something like that.

Now, to make a prediction of Y based on X you know that X has a value range from some minimum value to some maximum value. Now suppose you know that every data it is not suppose I mean every data always comes with noise. So, you suppose you assume that mostly the or 5 percent let us to make the calculation simple, let us consider 5 percent of data on the extreme are outliers or the noise data which you do not want to do analysis on. So, these may be some extreme points which you do not want to make analysis or they are the actually noise may be some kind of a mistake. So, these are the outliers you must have discussed studied outliers and Doctor Deepu Philip was talking about the data and its characteristics.

So, in a sense only point ninth of the data is relevant to you; now which means whenever you are modelling, there is bound to be an error and accuracy of up to maximum of 0.9 and 0.1 because the data will come on stream and some amount of data will be junk data. So, roughly 10 percent of the data is junk; so, it is only a 90 percent of the data which is relevant, now which is the 90 percent accuracy which is not a bad thing.

But suppose instead of X and Y; suppose we have this X 1 as well we have considered this as a another dimension which is X 2. Now the same thing happens this is for X 1, this is for X 2 again this amount of data is removed. So, if this be removed what is the overall loss of accuracy when you predict Y?

So, basically when Y is suppose when y requires both X 1 and X 2 it is a function of X 1 and X 2. So, roughly if you consider it as a 3 dimensional space you are cutting off in this in this 3 dimension you are cutting off extremes values of X 1 you are cutting off extreme values of X 2. So, what is left for Y is the relevant data points are the relevant I would say volume the relevant cube of data is does not contain the error points is 0.9 into 0.9 which means accuracy reduces from 0.9 to 0.81.

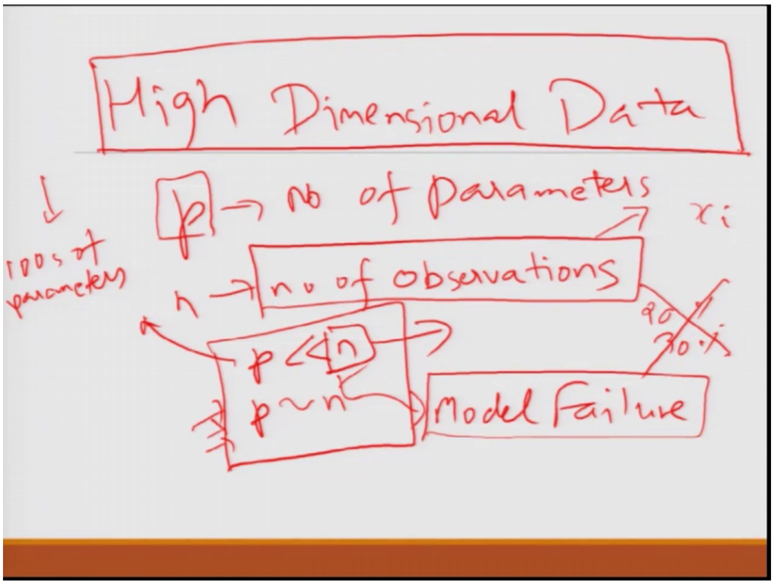
What if there was one more dimension considered to be hyper cube an n dimensional hyper cube. So, eventually you will see that the accuracy becomes a function of 1 minus. So, if this be the alpha 1 minus alpha to the power n, where n is the or in machine learning we normally use p; p stands for parameters. So, if you use p parameters in your model and each of these parameters have an accuracy or have an error of alpha. So, accuracy of 1 minus alpha the overall accuracy of your model cannot be more than 1; it has to be greater than or equal to 1 minus 1 minus alpha by p.

Now, this formula looks cool, but if you do a calculation the for example, when p was 1, it was 0.9; when p was 2; it became 0.81; if p becomes 3; it becomes probably around somewhere around 0.7. Now I give you as an exercise to calculate that for example, you are in a big data per lengths and then you get 100 parameters, very valuable parameter. Just try to see if that if alpha is equal to 0.1; what is the Y error looks like? Basically I would leave it is an assignment to calculate 1 minus 0.9 to the power 100.

Now, you must say that 0.9 is I have more accurate data and in my case you know my error rates are or this will be sorry this will be 0.1. In my case I have reduced error even further try to calculate 1 minus 0.01 to the power 100. Do the calculations and then you will come to know that as the number of dimensions increase, the rate of loss of accuracy is actually literally exponentially.

In other words, technically there is a term for it which is called I do not know if you heard earlier or not it is called curse of dimensionality. This curse of dimensionality is as we increase more and more dimensions as you increase and dimension is proportional dimension is analogous to complexity it is nothing, but complexity. So, as an as you make the model is more complex you put more and more dimensions into it curse of dimensionality comes and you are accuracy actually falls down.

(Refer Slide Time: 27:38)



The other way to look at is and this is again a challenging area a lot of research is happening especially after genomics has come up in a big way is; how to do analytics or machine learning on high dimensional data. Now high dimensional data means if p is the number of parameters, the number of independent variables on which you want to do to prediction; if you would have recall the lectures on regression you would recall.

So, these are the dependent variables or the x i's and n be the number of observations. Now, most of the algorithms are designed and in fact, this is also a kind of I would say the reality or the limitations of statistics that most algorithms will work fine. So, far as the number of parameters is very less than n the moment they start being comparable to n there are issues when the number of parameters are almost comparable to the number of observations that you have the models this leads to model failure.

However in the excitement and for most of the practitioners most of the people, who have who do not know the certain advances there is always a tendency to include as many parameters as possible in some way in the model; even though you do not have adequate data points.

I have seen this problem happening in lot of banks and financial institutions a lot of I would say I would attribute a lot of credit failures, a lot of credit risk in banks happening primarily because they forget this simple dictum. They will have models where they will try to have 100s of parameters 100s of parameters to assess a person's worthiness for a loan. They will have long questionnaires answers to that a lot of data coming from the past account quality record, the past transaction history, a lot of data that comes from other credit scoring agencies sibyl and stuff.

So, they have this huge number of parameters they try to combine all these parameters to build a model of the issues that the number of defaulters in the system are not that high especially if you talk of say small business or large corporates. For credit card of course, there may be a larger number of defaulters you may see, for credit cards for personal loans there may be still because these are small loan, but for large corporate for small businesses even. The number of loans in the banking system are not that huge; so, what happens is p becomes comparable to an or maybe the number of parameter is hardly a fraction of n maybe 20 percent or 30 percent does not work.

So, one of the I would say for one of the most precious take a ways that you can have from this course is always compared to the total number of observations that you have to build the model. We will talk about validation of model and a prediction the testing of model separately. So, not all the data that you have is used to build the model only a fraction is used to build the model test is use for testing the model, validation of the model.

So, the number of observations that used to build the model; if your number of parameters are large compared to it; the models are likely to go for model failure. This may not be the case there are ways to deal around it for when the parameters are more of hard or accurate parameters. So, for example, for things like image recognition and video recognition this may work.

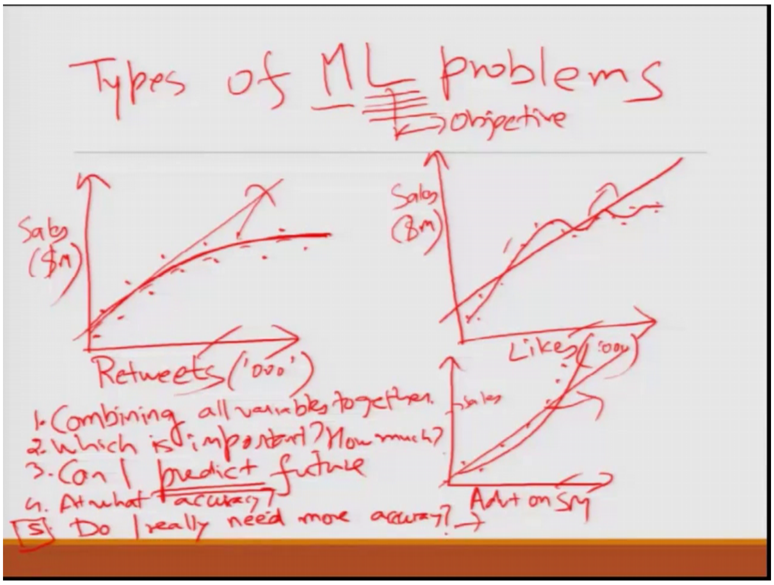
But if you are talking of things you know say management perspective, you are trying to do marketing analytics, you are trying to do financial analytics, you are trying to do risk management credit risk analytics, you are trying to do operational analytics; this actually becomes an issue. In fact, this has been a reason if you know bustle is the guiding authority on risk management for banks.

And it had come with a very complicated algorithm to calculate the operational risk and it wanted the banks to calculate all kinds of operational risks and then combine into a number used very advanced statistical methods. And some of that you have studied some you will study in future lectures to come with an assessment of the operational risk; did not work for one simple reason that banks did not have sufficient number of data points on operational risk a because bank did not have sufficient number of data points on operational risk.

The entire complexity there is a limit to the amount of complexity that you can have. So, in the new edition of bustle again they are proposing to create a simpler approach asked banks to calculate operational risk using a simpler approach rather than going for this amount of complexity.

So, (Refer Time: 33:35) I would say the million dollar take away from this course and something when you get into your jobs or if you are in job you apply machine learning, you will see that will gives you a huge amount of edge compared to many of the peers who do not know this formula is that always try to keep your models simple.

(Refer Slide Time: 34:01)



Let us spend some time discussing what are the types of machine learning problems that we try to solve or try to understand basically the what actually is defined in machine learning. As I said a technically even box plot from where you generate certain rules can be a machine learning algorithm, but in industry in common term there are certain things which are called which are which are included in this cluster of algorithms called machine learning. So, let us understand what all those things are.

The first kind of machine learning algorithms and you will immediately able to figure out what I am trying to do if you have done the previous lectures is suppose you have these number of retweets in 1000s versus I would say sales of a product in dollar million. So, it is a company which does a lot of online promotion; so, it try to find out that whenever there was a particular amount of retweets of my specific products tweets that I sent around particular product this was the kind of sale I received.

So, let us not for example, get into the mechanics of how they exactly built this data let us assume that it was something like that. Similarly it is an assessment of number of say the likes versus sales again in dollar millions and this time something like that; similarly they had or other graph which was on advertisement on social media.

So, here for example, they find this kind of a trend you would have recall that I am doing something similar to regression analysis yes; that is regression is the most basic form of machine learning. In fact, if you pick up any book on machine learning the first thing they teach you is regression. The first thing they try to do is to make your concepts of regression clear and here we will also try to build in tuition around it.

So, what are the questions; from a machine learning perspective let us understand what are the questions? So, as I said machine learning means you want to automate the process and you want to see what kind of learning we can derive this part is important. So, what are some of the objectives and as I said learning means you have an objective. So, what are some of the objectives that you can have with this kind of an analysis or this kind of a data or even before you started doing this kind of an analysis.

So, 1 is combining all variables together; so, here you have made graphs of sales versus retweets sales versus likes sales versus advertisement; what would be a combined thing look like? How do I combine the retweets likes and advertisements on social media together to come with the common model. So, multivariate analysis is the technical word then based upon that from a practitioners perspective which is important.

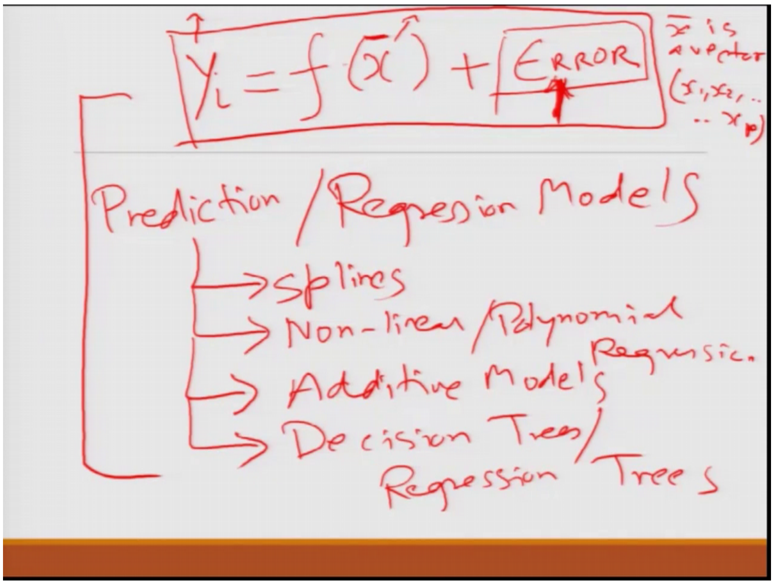
So, suppose I have budgets and I have to allocate only to one of them; should I allocate on this or this or this, it is not a simple of just making a simple may be a linear regression and saying whichever gives me the highest amount of slope or the variable parameter. So, I will choose that you have to also look at what are the kind of errors; you face you have to look at various other parameters and we will discuss some of the pitfalls.

So, which is the most important how much and mind you linear regression is one way one of the ways of doing there can be other ways of solving the same problem. So, which is more important how much and based on this can I predict future? So, tomorrow if I say that I invest this much amount on likes this much amount on trying to get advertisement this much amount of tweets I get can I expect what my sales would look like. So, this is the prediction this becomes important.

Fourth at what accuracy, and fifth which is extension of what we studied in two slides ago, do I really need more accuracy? Now this is the question that will keep coming up and I will keep reminding; do I really need more accuracy? I can look at the figure and I can say that probably if it was very steep and my I might say that likes seem to be the most important; other things do not have much of a bearing. And I can say that just take your decision this is how most of the cos will run the company in a classical case.

But do you want to make something more sophisticated to one say these kind of questions and get more accuracy. So, these kind of models; where you want to predict something try to compare different variables find which are important and then try to predict something these form what we call prediction or regression models.

(Refer Slide Time: 41:25)



So, you talked about linear regression, multiple regression and stuff you have had some understanding of this. But there is more to it this trend for example, a linear regression might fit this kind of a curve, but if you live carefully maybe this kind of a curvature to suit the model better. In this case, probably the linear will go like this; however, the actual data is something like this; in this case it is like this linear may fit like this.

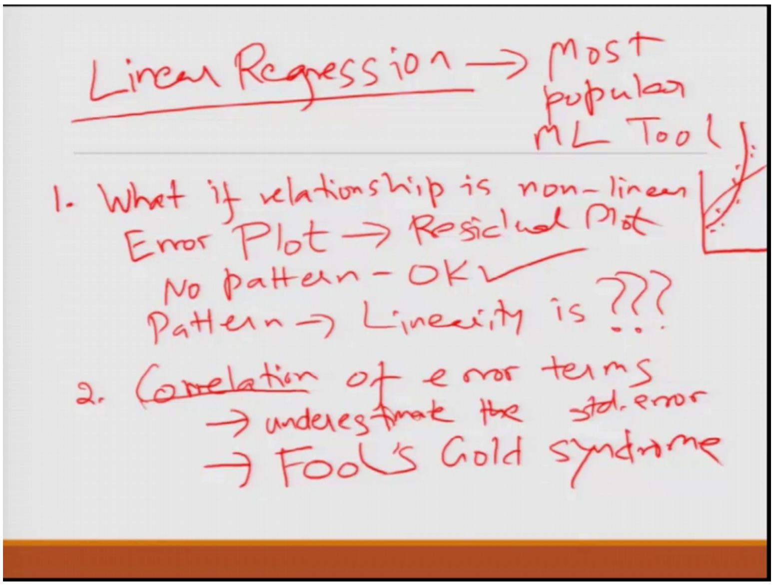
So, linear is an approximation; there is more to it you can make more complex models, some of the more complex models are splines; you can make non-linear or polynomial regression. So, basically instead of just the linear terms you have even x square x cube terms; then there is a whole family of tools called additive models and then to even have decision trees also called regression trees.

So, these are different kinds of models some are more sophisticated, some are less sophisticated as we discussed last time and let me put it slightly in a more matey way; so, that when you start reading some of these more serious texts you can make feel out of it. So, so, this is what basically a prediction is trying to do, this is what I want to predict xi’s are my different let me or rather you know try to call it an x vector.

So, these are basically the different parameters that you observe. So, this is the equation; your goal is to minimize error or is it really we will we will talk about it later, but in general what you want is; you want ideally in a simplistic case you want your f x to be closer to y i and reduce the error.

Now, depending on the sophistication of the algorithms that you use; depending on the methods, the smart ways that you use even within the to tune the parameters within these models themselves you; in a way the intuitively you want to reduce this much to the extent possible and bring f x as close to the y. As we will see this simplistic intuitive approach has its limitations there are other consideration that come up. So, sometimes you do not want error to be beyond a point. So, as I discussed in the beginning of the lecture; you have to first of all decide how much of error you can limit, do you really want more accuracy and we will see as I said more complex model, more price you have to pay.

(Refer Slide Time: 45:36)



So, now we talked about linear regression let me I will not get into details and mechanics of linear regression, but because this is I would say popular ML tool. Because this is the most popular machine learning tool; it is important that we spend some time discussing some of the pitfalls or some of the issues that you can face with linear regression as a practitioner.

The first what we also hinted in couple of slides ago is what if the relationship is non-linear know; actually the data is something like this going like this it is not linear what do you do? Well we had covered this earlier in the course; what you do is you create error plot or residual plots and try to see if the errors are showing any kind of a pattern or not.

If there is no pattern is ok; if there is a pattern; that means, linearity is a question mark; it is a question mark. And if it is a question mark you can still go with linearity because you do not want to pay the price of complexity. But there are other ways in which you will try to reduce this nonlinearity; sometimes you can transform the variable or you can use polynomial regression or more complex form of regression.

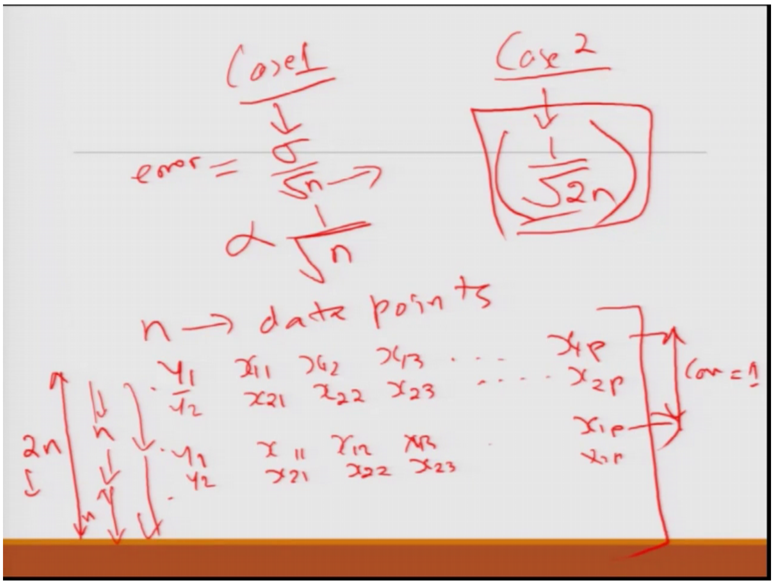
But my advice would be my thump rule would be that, unless you have a compelling reason, unless you see a really significant increase in accuracy. And some other thing that we will discuss in future try to be as simple as possible; remember complexity has a cost the second issue that comes is correlation of error terms.

Now, what happens in this case is that whenever you do a regression, the regression will calculate a standard error for you. Now, when the error terms are themselves correlated you under estimate the standard error. In plain language you do not know what is the amount of error; that you have you are illusion that you have got less error, but when the actual error is high.

And this is a very very common especially in the financial domain; especially in marketing analytics this is a very very common pitfall people simply use a linear regression or use a more complex form of algorithm they do not look into the fact that there is a correlation of error terms. And I would call it fools gold syndrome because you really get low amount of standard error, you feel really excited about it your happy oh you made a wonderful model. But the reason was there was a correlation in the error terms and hence your happiness your euphoria is misled.

Now, just to give you a intuitive sense of what this means and I think it is important for you to have this feel; I am not sure how many will emphasise on this, but as somebody who has practised and who has fallen trap to this and who has burnt his fingers; I cannot emphasise it more. Just to give you a very very intuitive sense and you will realize why this is; so important.

(Refer Slide Time: 50:28)



You assume when if you recall some of those earlier statistical lectures in the course; you would have recalled that when we try to estimate the variance for the standard deviation of the population, we estimate by the sample and the calculation is that the error is sigma by root n.

So, basically I do not want to get into statistics, but basically your error is proportional to 1 by root n. More the number of data points you have the in the sample, your error reduces now this is used to estimate your population statistics. So, and that is why it is important that you keep the sample as large as possible.

Now, suppose I was a very cunning analytics consultant and I knew that you do you are not much aware of this fools gold syndrome. And I wanted to build a model which is really very accurate a model which is wonderful you say wow I have not seen such an accurate model and I want to fool you. So, what can I do? Let me give a very extreme example and then you will realize; all I do is suppose you had n data points. And I talk of data point data point is some observations that you have got on which you want to build a model. So, there is a y i there is an x 1 there is an x 2; there is an x 3; there is an x p. So, this is a typical data point and I have n of these. So, I have got these and data points on I want to build the model.

So, what I do in an extreme case? I simply replicate these n points again. So, the same data point is repeated twice; so, instead of n now I have n more data points, which are again. So, if this was x 1 this was y 2; then again I repeat y 1; let me make it more clear; so, I had a y 1. So, these are my individual data points what I do is I just copy. So, if you are for example, working in excel I just control see I select all the rows in which the data exists I control C; I control the copy beneath it. So, I have exactly the same data sets beneath it.

So, similarly y 2, x 2 1; x 2 2; x 2 3 this was X 1 3, X 2 p and so, on. So, in total now I have two n data points; so, case 1 my error was proportional to 1 by root n just by using this trick. Now my error is proportional to 1 by root 2 n which means simply by control C, control V using the same data set just repeating in it. Once again I was able to show you that now my error has reduced by one by root 2 times. So, 1.41 something bits; so, may be around 70 percent or something. So, and think of it; if I if I do it 3 times, 4 times, 5 times what happens?

Now, what was the issue here the issue was because these are just replica, it means the correlation between the error of these two terms. So, what it means is that the correlation was exactly one here for these data points; you get a reduction if there was a lesser correlation you would have still got at least some amount of error reduction purely because the errors were correlated; your model is still the same your model effectively is the model built on a smaller data set its an illusion of having a more accurate model.

So, today we talked about two traps; in the next session we will talk about some more traps and regression which you need to take care off. We will also talk about some other kinds of machine learning algorithms like classification and like unsupervised learning like reinforcement learning and the philosophy behind it. So, thank you very much hope you like the lecture and let us meet in the next session again.

Thank you very much bye.