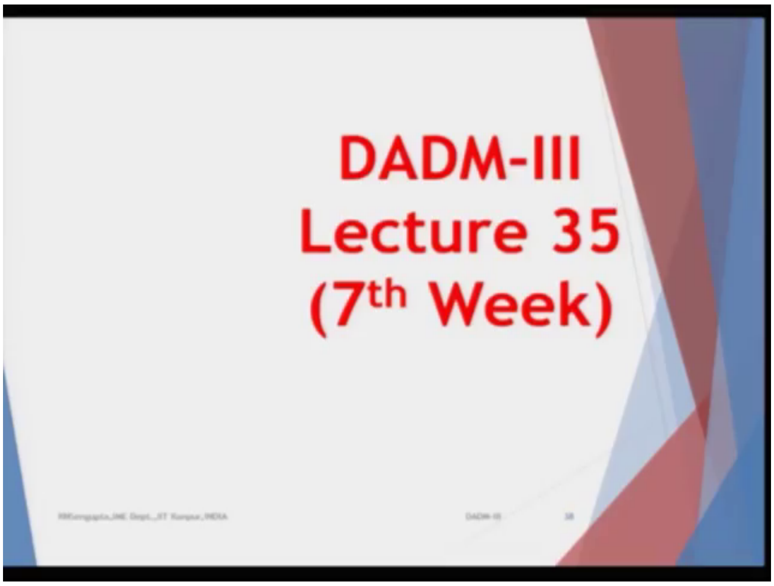
Data Analysis and Decision Making – III
Professor Raghu Nandan Sengupta
Department of Industrial and management engineering
IIT Kanpur
Lecture – 35

Welcome back my dear students. Very good morning! Good afternoon! Good evening! To all of you, wherever you are in this part of the globe. And as you know this is a DADM 3, which is data analysis and decision making 3 course and the NPTEL MOOC series and this total course duration is basically for 12 weeks, the total contact hours is 30 thirty hours which when split into the number of lectures is 16 number. And in each week we have basically 5 lectures and after the end of this 5 lectures you have 1 assignment.

(Refer Slide Time: 0:47)



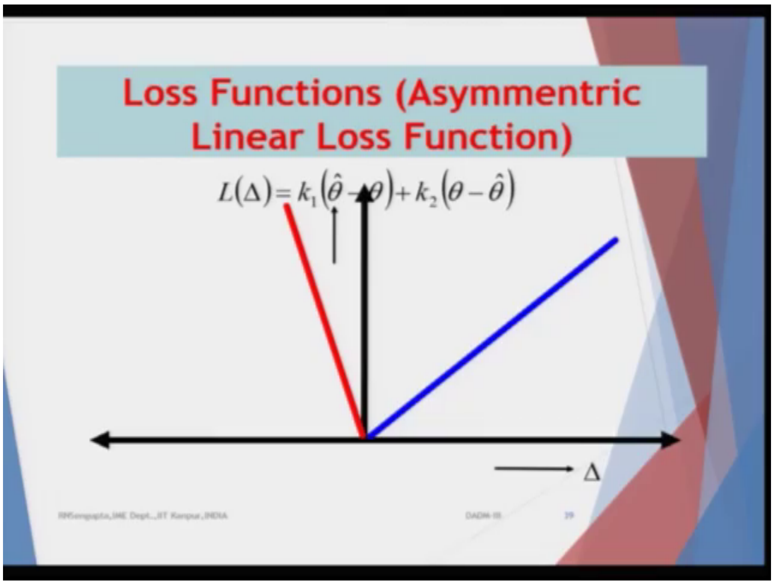
So as you can see from, from the slide which is the seventh week. So after this you will be basically taking the seventh assignment and after the end of the twelvth, 12 different assignments you will be taking the final examination. And my good name is Raghunandan Sengupta, from IME department at IIT Kanpur.

So if you remember we are discussing about the concept of loss functions. Loss function in the sense that provided that you want to estimate an estimate the parameter using the sample statistic which is theta hat, so theta hat minus theta if it is considering as a quadratic loss function it is something to do with the utility functions also I am not going to go into that detail. So if it is a quadratic loss function then trying to find out the minimum of that loss function also is again trying to basically find out the minimum variance which basically makes very good sense in the statistical concept.

But other type of loss function which are utilised in statistics and other studies is basically the function where the both for the estimation on the right hand side on the right hand side which is over estimation, underestimation is equally paralyzed but it is straight line in the sense if you are taking them modulus of the theta hat minus theta both on right hand side and the left hand side. So the only problem which I mentioned is that the at the point of zero the which would be considered the point, we are trying to find out the dy, dx of the function would be difficult.

Now, if I considered that the non linear loss function considering the fact that if it is a module loss function I mentioned that just when we wrapped up the thirty fourth class, I said that I will go through the diagram accordingly and then come into the practical applications or some of the actual the realistic loss function which we generally utilise.

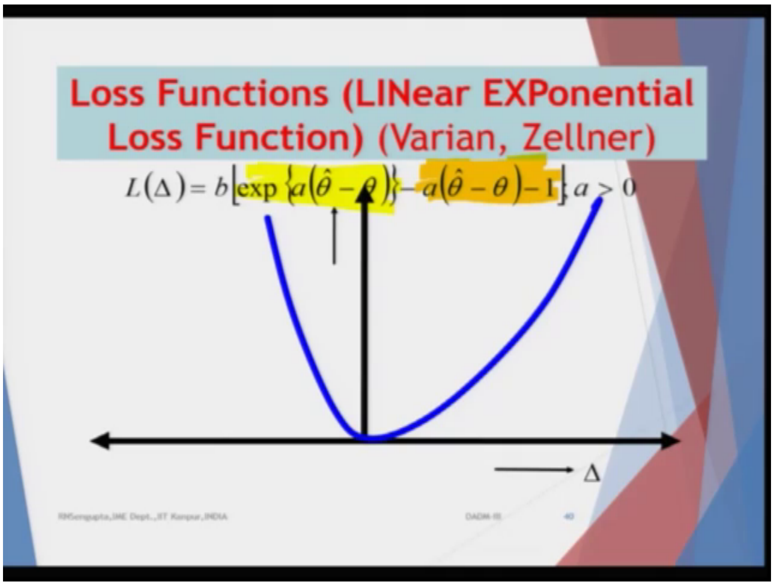
(Refer Slide Time: 2:49)



So in the first instant we consider that the value of k 1 is greater than k, hence the overall loss which is there if k 1 is greater than obviously the overall loss which is or the gain let us put it in the way if you are looking at right hand side but generally there would be loss but because there is difference between the estimated value and the actual value so the overall loss function which is there on the right hand side when theta had is minus theta or delta is positive it will be more penalized on the right hand side and less penalized on the left hand side.

The moment you basically consider the values of k1 and k2 are swapped then the as it is shown in this slide underestimation over a loss would be much more penalized then, then the overestimation. Hence the slope of the red line is much sharper than the blue line. Now again the obviously the problem remains trying to basically find out the dy, dx of this function at the point of zero is impossible, Since, hence even though it looks interesting you would not be utilising that for our practical purposes.

(Refer Slide Time: 3:44)



Now in the 90s, 70s, 80s, 60s a work was basically done by the Hal Varian who is basically in area of economics. In order to basically the predict the prices of houses in or trying to basically study the overall effect that weather overestimation, underestimation can be broken down in such a way that they would basically make some practical sense. Now this was basically, that was basically initially from the economics point of view later on Arnold Zellner took this up as statistical study, where he proposed the statistical properties of a loss function which is known as the linear exponential loss function or the LINEX loss function.

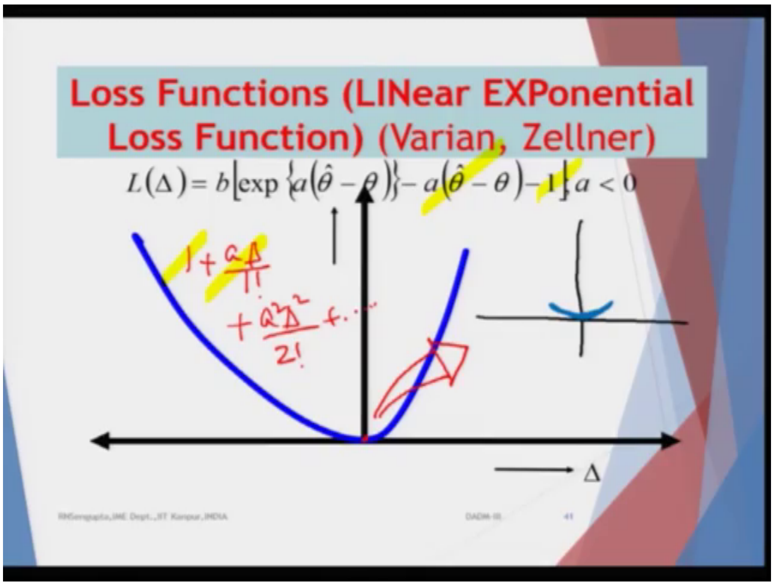
So Linux loss function basically looks like, there is a diagram so I will come to the diagram why this diagram looks like this? But firstly let us basically concentrate on the equation. So basically there are 2 parts the first part is the exponential function which I will now try to highlight, ok using different colour so this will be the exponential part and if I am considering the, the linear part it will be this. So now see the beauty of loss function and obviously b and a are some parameter value so b if it is multiplied a scales up or scales down the overall loss function which you have.

Now consider the theta hat minus theta is positive and a is also positive so in that case what will happen that if you consider any value on to the right that is delta is positive that is theta hat minus theta is positive then the exponential part will start increasing exponentially as you go and increase the values of delta. But it will happen in such a way that there is obviously a negative part, negative part is what? Is basically minus a in the bracket theta hat minus theta but the rate of increase of the exponential part would be so high that you will basically dominant the linear part, Hence the right hand side of the equation which is in the first quadrant would be basically increasing exponentially.

While if I consider the left hand side of the equation, left hand side of the diagram which is there on to the second quadrant in that case obviously a is positive which is fine but the value of theta hat minus theta is non negative, so if it is negative so obviously you have an exponential function with some negative values. So as it starts increasing the left hand side increasing as if we go more and more on to the left, then in that case it will be 1 by that exponential function which would basically decrease exponentially.

Hence if you mean that the linear part would slowly start dominating the exponential part so hence the left hand side even though the diagram may not be drawn very perfectly the left hand side would basically almost be a linear function while as I mention earlier the right hand side would be an exponential function because in the right hand side exponential part will dominate the linear part while in the left hand side the linear part would basically dominate the exponential function. Now consider the fact that I basically reverse the a value so I take it negative. So in the case when you have taken it negative when I go into the right hand side which is the value of delta greater than zero or theta hat minus theta is greater than zero.

(Refer Slide Time: 7:15)



So in that case the exponential part which is already there it would now become 1 by a to the power some value because that negative value would bring it to the denominator, while the positive part would be the linear part would be with a minus a it will become positive. So any values which is linear would basically start dominating the exponential part. So the right hand side of the equation would basically look exactly what you have seen in the previous diagram, so that will be a linear in nature while if I go to the left hand side, the left hand side would basically would be such that it will be with the minus sign if I am considering and also theta hat minus theta is also minus.

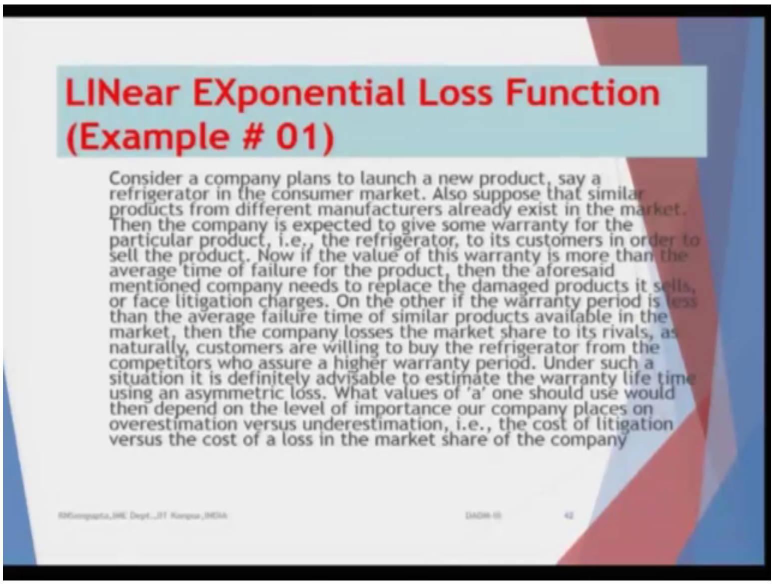
So in that case exponential would be a positive value hence the dominance of the exponential part would be so large than the linear parts, Hence in the second quadrant you will basically have a exponential dominated curve while in the right hand side in the first quadrant it will basically linear dominated curve. So the linear exponential loss function would basically be more penalized when the value of theta or when the value of a is positive when you are concentrating on the right hand side which is the first quadrant and it will be penalized more on to the left in the second quadrant if it basically a value is negative.

Now the interesting part of this equation is that if I basically zoom in and in and around the value of zero, so if I basically zoom in here, so if I zoom in here then the graph would look like this so I use the microscope zoom in, so it will be almost a quadratic function. Very near to the value zero so I am not considering any higher value and this will be true from the equation perspective also why? Let me draw, write it down. So consider the exponential loss function, so you will basically have e to the power some a delta so let us expand it.

The first term would be 1 then the next term would basically a delta by 1 factorial, the next function would be a square, delta square by 2 factorial, I am expanding the a part and then the cubic part to the power part I will continue so let me put it as dot, dot, dot. Now if I consider 1 and 1 here they will cancel out because that other one is basically negative so basically this value and this value cancels. The value of a delta by 1 or a delta basically cancel so this plus and minus sign as it is there and this exponential I have already expanded.

So if I consider the third part as quadratic and higher parts to be neglected considering that they are delta cube which will basically tends toward zero as we increase the value of a as well as the value of delta. So the only equation form which will be true for the linear exponential form in around the value of origin would basically be the quadratic equation. Or if you consider the utility function which you have already done and based on a quadratic utility function which basically gives the information that the returns of the decisions are normal in nature. So this is, is a very typical property of the linear exponential loss function as proposed by Variance and Zellner.

(Refer Slide Time: 11:07)



So let us consider 2 examples here. So this examples will make the concept of loss functions clear and why it is positive? Why it is negative? Consider that you have a product or company to launch a product and I will come start with the very general example then, then come to the specific. So the company wants to launch a product and for launching of product is basically wants to find out that what should be the actual price of that product. And it gets some information from its competitors from the market and then it launches the product. So considering the price of the product is say for example 100 and I basically trying to sell it in the market at the value of 90.

So in that case what will happen that an also I consider that depending on price I gave a high or low warranty life. So in this case what will happen is that I am not specifying weather to give a higher warranty of life and lower warranty of life. So the moment basically I go for a lower price people will attempted to buy my product but the fact would be a higher or lower warranty life would mean that the if a higher warranty is there actually when it is not required it will mean the product will start failing much before the warranty life which has been given by me.

So in that case I will basically be incurring a huge amount of loss. In other case if I have given the warranty life which is less actually had it should be and still the people buying then in that case my warranty life would not be validated and I can increase my sells accordingly. Because by the word of mouth my product will be considered as good. Now in this case weather the warranty life should be higher or weather the warranty life should be lower or weather the price is higher on an average or lower on an average would basically depend on other data points or other information which is available with me.

So let me read the problem considering the company plans to launch a new product see a refrigerator in the consumer market also suppose the similar products from different manufacturers already exists in the market then the company is expected to give some warranty for particular product. Warranty means guarantee that is the refrigerator to its customers so they want to give some warranties such that to the costumer. They would be interested to buy the product or the company would be able to sell the product provided the warranty life or a guarantee life actually makes some positive sense to the costumer.

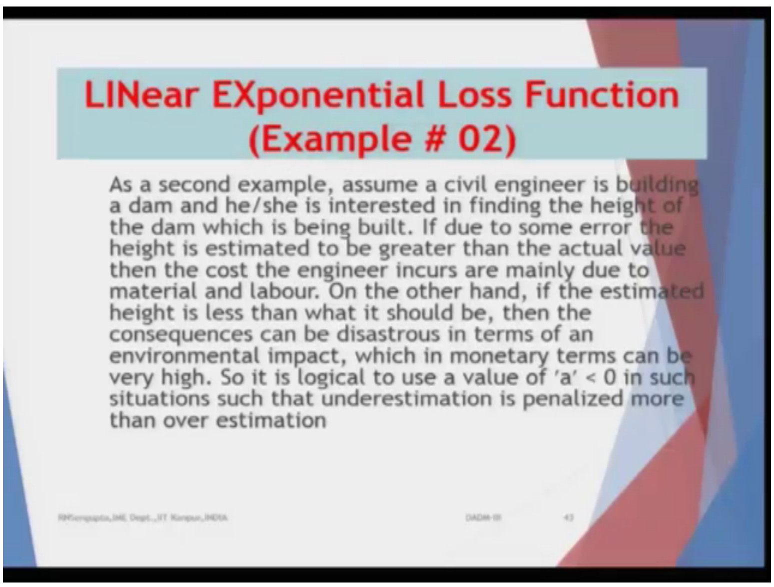
Now if all the value of this warranty is more than the average time of failure of the product as I just mentioned, then the mentioned company needs to replace the damaged product so that means they would definitely not able to cover the warranty time. Replace the damaged product it sells or face litigation charges. In the case that if I do not replace the product obviously the costumers will file the litigation charges. So in that case I have to basically find out the balance in the fact that if I gave the warranty life is lower weather I should replace the product or basically face the litigation charges. Whichever is high will take the decision accordingly such that we do not face the huge amount of loss.

On the other hand if the warranty period is, is less than the average failure time of a similar product. So basically the warranty, warranty now over a value will switch. Than the average failure time of the similar product available in the market then the company losses the market share to its rivals. Because the warranty life is higher than obviously the people will attracted if it is lower sorry your product is lower people will be attracted to the rival product. Under such a situation it is definitely advisable to estimate the warranty life time period using an asymmetric loss function.

What values of a should be so if it is weather positive or negative should use so when you are trying to utilise a you should use the concept. That depending on the level importance you want to face for overestimation, underestimation you will basically put the value of a as positive or negative. So overestimation means that you want to penalize much higher on the right hand side. Underestimation means you want to basically penalize the underestimation part much higher.

So what values of a one should use would than depends on the level importance our company or your company places on the overestimation problem versus the underestimation problem, that is the cost of litigation versus the cost of loss in the market share. So in case if I am losing the market share and also basically losing the litigation then it is disaster from. But in the case when market share is increasing but even that I am able to offset the cost then obviously it will mean that I do make some sense when I am trying to give a warranty life which is higher or lower than the ordinary average value which I want that to be.

(Refer Slide Time: 15:55)



Consider second example and consider a civil engineer is trying to build a dam. Now assume that the dam height is 120 feet but there are 2 instances. Consider in the first instance the engineer builds the height of the dam to say for example 122 feet with respect to 120. And in the second case the engineer builds the height of 118 feet with respect to 120 feet. So in the both the cases the difference would be 122 minus 120 which is to whole square that is 4 and in other case it will be 118 minus 120 whole square of that is 4. So if I am basically trying to plot it on a quadratic utility function or quadratic loss so both the values of positive and negative would be equally penalized.

If I go on to the right hand side or left hand side of the, the estimated difference of the estimated value. Estimated value and the real value which we have try to find out or the parameter value which I try to find out that is delta and plot it along the x axis. Now what happens in the case if it is built to a height of 120 feet the initial cost is much higher. So it is man hour cost, material cost so and so for. But when see for example when the flood comes the propencity of flood to breach the dam which is of height 122 feet with respect to 120 feet is much lower.

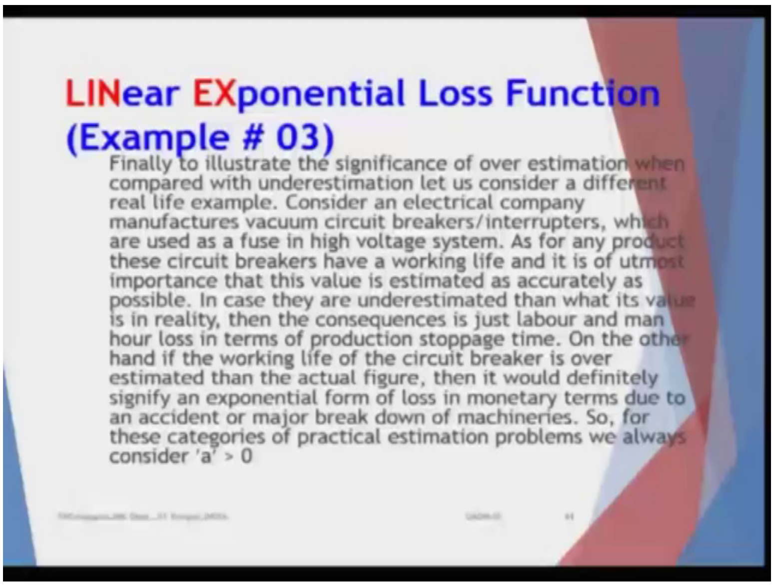
So hence we do not face any catastrophic loss. Or in the other hand if the height is 118 feet and if a flood comes then the propensity of flood to breach the dam is much higher. So initially we may not have spent that much of money to built in comparison to 122 feet we have not use that money because we did not use materials, we did not use labour cost so we have been able to save it save that all total of money. But in case as I said if the flood comes the propensity of breaching the dam height which is 118 feet is much higher than in that case the overall catastrophic loss is much, much higher. So in the first instant when you have overestimated it and in the case when you underestimated it the overall losses in the underestimation case would be much higher. So you will be attempted to utilise a new or practically it is useful to utilise a value of a which is zero such that you would penalized underestimation much higher than the overestimation case.

Now consider the case you have an electrical huge equipment and there are this circuit breakers, circuit breakers are basically type of switch which are there trip switches. Now consider the warranty life of that machine based on to the fact that you will basically stop that machine and change this circuit breakers is say for example 6 months, so if you do not do that there may be a chance that there may be your catastrophic accident. Now in case say for example you gave a estimated life for a warranty life in one case as 8 months and another case has 4 months and in both the cases it is 8 minus 6 which is two, two whole square is 4, in the other case it is 4 minus 6 is minus 2, minus 2 whole square is 4.

So we are happy that both overestimation, underestimation are equally penalized you will use the quadratic loss function solve our problem and proceed accordingly. But does that solve our actual problem the practical sense. Let us see it in case if I have overestimated it and I say that is overestimated in the value of plus 2 then we will attempted to basically to stop this machine. In a sense overestimated that if it is 6 and we have overestimated to 8 and in that case we will basically be, we will be basically be tempted to start the machine in a sense that for negative values or positive values will be stopping the machine at a higher level based on 6 or on the lower level based on 6, 6 months.

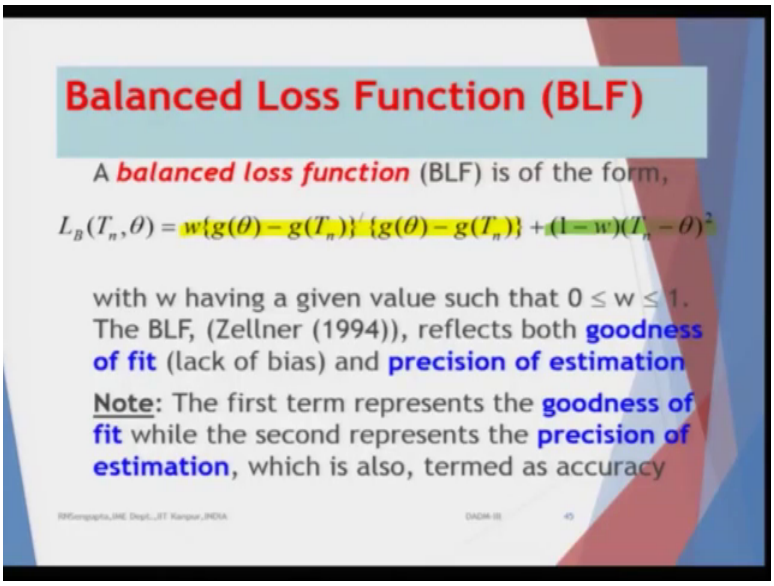
So stopping the machine at a level when the actual time period of change of circuit breaker has not come. It will only mean that there is stoppages of worker, worker man hour there is a stoppage of production. But we will be changing the circuit breaker much more often. And in other case when basically we underestimated and then we and we have not been an we basically be a giving a guarantee life in such a way that will be basically not changing the machines or not changing the circuit breaker are the optimum point of time obviously the production will increase but the flipside is that the chances of catastrophical loss is much higher.

(Refer Slide Time: 20:36)



And in this case we will be tempted to find out the value of the overestimation is more important for us so we will basically take a value of a as 0. Which means that if overestimate they will basically stop the machine later on so the initial cost of production is very positive to us in the sense that we are making more and more profit. But in case if a catastrophic loss happens because as we will stop the machine later on hence the overall loss or overestimation will be much higher but in other case when we do with the underestimation we will be tempted and we will be forced to stop the machine much before the 6 months change it. So obviously they would be some losses on the production fund but there would not be any much catastrophic loss for human life or overall accident which may can happen. Because you are trying to stop the machine much before hand.

(Refer Slide Time: 21:21)



There is another loss function which is used very, very interestingly in the area of statistics or in area of regression or in forecasting is basically known as the balanced loss function. So balanced loss function is also utilised as been predicted or given by the work of Zellner in the 1994. So balanced loss function works in this way see for example you are doing a regression model and in regression model our idea is basically to find out errors. Errors is basically the difference between the actual value which is y and the predicted value which is y hat.

And we find out the difference of this errors and as per the concept of the quadratic loss function which you consider in initial part when we are discussing we basically try to square this overall loss add up all the losses for all the estimates which are there. Estimated and the forecasted value and the actual value. We add up this square of this losses and trying to basically minimise this losses. So in the case when you trying to minimise the losses you will basically try to differentiate this losses with respect to the all the parameters which are unknown to us put it to 0 find out this parameter values using that estimated value which is higher value and then our life is done.

But Zellner propose a very interesting loss function which I just mentioned which is a balanced loss function where you take there is errors both on 2 accounts. Point 1 the error is there for the estimation problem as such that is the parameter values which you find out that is the values which I just mention alpha hat, alpha 1 hat, alpha 2 hat, alpha 3 hat which are the concept of the regression coefficient which we have. Now that would be 1 error, another error would be once you use those estimated values to find out the value of forecasted value of y so obviously they would be of error when we try to find out the difference between y and y hat.

Now what Zellner proposes basically you try to basically penalise both the concept of estimation error when you are trying to find out theta with theta hat and also there the forecasting error when you are trying to find out y by y hat by utilising the concept of quadratic loss function and it can be shown that the properties would be such that overestimation or underestimation can also be considered in unequal proportions in order to come off with some practical good examples.

So the balanced loss function is given as such where w is some, ok another thing which I forgot to mention is that when you are trying to basically find out the overall errors, so there are 2 parts 1 error is basically the estimation error where as I mention if you try to find out theta using theta hat and the other error would basically be the difference between the y value and its forecasted value which is y minus y hat. So when, when you have to considering some w value which is a lambda value it is the overall weight we have going to give to the forcasting error and as well as the estimation error.

So if I am multipling w or delta with the concept or the forcasting error so here which is which I will basically mark is yellow and the other one when I am trying to basically utilise the estimation error so here it will be 1 minus delta, 1 minus w. So in case w is 1 then the overall effect of the loss would all be subsumed on another forecasting error in case the w is 0 the overall loss would basically be subsumed on the estimation error.

So with w having a given value such that it is 0 and 1 Zellner found out this, this concept of loss where it can be split into 2 parts. 1 is the goodness of fit which is a lack biasness and 1 is the precision of estimation. So the precision estimation would be the concept of the estimation which you have doing which is on the right hand side which is green in colour. And the goodness of fit would be on the left hand side where you want to find out the difference between the actual value and the forecasted value.

The first term represents the goodness of fit as I just mention while the second represents the precision of estimation which is also termed as a accuracy of the prediction or the forecasting which you want to do. Because based on the, on the estimation of the forecasting nice it is we will try to basically utilise that to do the further forecasting and trying to basically utilise the values of y hat in order to predict the values of y actually which is needed.

Now before I start the DA concept I will basically give you a very brief introduction, even though it would not be possible for me to consider everything in this thirty five lecture, but I will just give you a brief introduction and I will start from this slides again. Later on when we start on the, the next lecture which you would be on the starting of the eight week. Now when we are considering the concept of the estimation problem as I mention and when we are considering the DEA, DEA basically means the concept of data envelopment analysis where the main focus of the problem is try to find out the efficiency.

Now we are not changing the track I am basically trying to basically analyse the problem from the point of view of operation research. Now given any factory which will consider as the decision making unit. We will consider there is some bundle of inputs and that some bundle of outputs are produced. So let us consider that you have a hospital in the case when you have hospital the number of inputs or type of inputs can be the type of, of surgery you can perform, the number of type of doctors you have how good or bad, there are the number of beds, number of nurses, proximity to, to say for example localities where people are there and output would be how successful the operations are.

What is the total number of operations? Or output the particular hospital is able to perform? What is the recuperative of percentage of people who are admitted? Or what are the critical cases of, of health checkups which the hospital does? So some of this variable or some of this factor would be objective and some of the factor would be subjective. So trying to basically combine the subjective and objective part would be the main focus when we doing the DA studies.

Consider next example, see for example you have a school or a college or a institute and you want to basically rank this institute depending on different criteria. So what would be the inputs? Inputs can be the type of students who are going to come, the type of faculty members, type of infrastructure, the library, the curricular activities so an so also it can be the staff members, proximity to the city where all the people who are from different status of society can basically study there.

What would be the output? Output could be see for example the how many students have really secured very high marks in either class 10 or class 12 or it can be say for example how many students have been admitted in different type of professional degrees or it can be see for example how many students have basically excel in sports. So all this input or output bundles would be considered in such a way that you will be able to rank the (28:43) depending on some concepts of efficiency. I am going to come to the concept of efficiency very fast.

Now another example who can be say for example we are considering a government institute and the government institute or, or the government and we are considering that the government has different type of schemes to perform see it can be mid day scheme, it can be the performance of the government and trying to basically maintain the crime in the city or in the state, it can be whatever say number of public schools are there, number of public utilities whichever are there, what are the road conditions and all this things can be there.

So the input can be say for example number of amount of money which is being put, number of type of engineers, the efficiency of see for example engineering department then output can be successful mid day meal, successful number of roads which are in good conditions, the electricity supply connections and so on and so forth can be consider. So in all this problem we will see that the overall input and output can be classified into some subjective and objective part such that combining them would give you some scores and we will trying to solve this problem using the simple optimisation problem where we will see linear program will be heavily utilise. With this I will close this lecture and have a nice day and thank you very much.