Data Analysis and Decision Making - I
Prof. Raghu Nandan Sengupta
Department of Industrial &Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 24

Welcome back my dear friends; a very good morning good afternoon good evening to all of you and this is the Data Analytics and Decision Making course number 1 under the NPTEL MOOC series. And we are on the 24th lecture as you can see and this is a 12 week course for 30 hours that is total number of lectures will be 60 each week we have 5 lectures each being of half an hour and I am Raghu Nandan Sengupta from the IME department IIT Kanpur.

So, in the last class which was on the 23rd lecture we were discussing about linear regression, multiple linear regression the concepts and the ideas what we mean by the assumptions; assumptions being what is this error of the for the multivariate linear regression, then linear regression what is the expected value of the error. What is the variance of the error how the xs are dependent if there are more than 1 x; what is the relationship between the xs and the errors what is the relationship between the how do you find out I would not use the word relationship how do you find out the variance of y given the variance of x and the variance of the error. How you find out the expected value of y given the mean value of the expected value of x and expected value of the error.

And we also in the last part out of third lecture we discussed that what we mean by rank why rank is important what we what is basically the concept of errors being independent, the concept of I just mentioned the words homoscedasticity and then heteroscedasticities they would be coming up later.

And we also discussed that given the n number of values for ys and n number of values of xs considering there are k number of xs, we can predict the alpha hat the beta 1 hat, beta 2 hat till beta k hat, alpha may be there may not be there depending on the problem formulation has been done. And using those the estimated values or the hats and they are based on the fact that you minimize the sum of the squares of the errors and once these hats are found out you basically plug them in the nth plus 1 value utilizing the x 1 to xk; that means, x 1 comma n plus 1. That is the first value the nth plus 1 value for the first

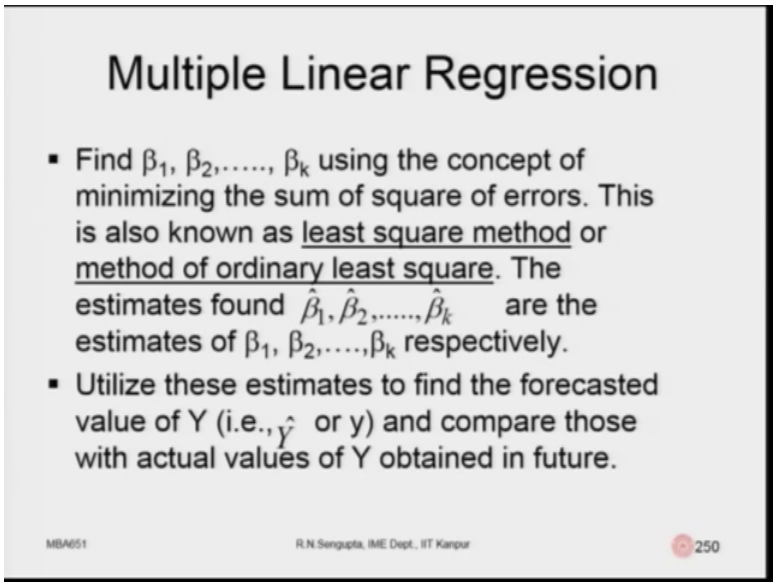
variable till the nth plus 1 value for the kth variable utilizing that you find out the estimated value of y n plus 1, and then utilize that value basically to find out what is the error of the nth plus 1 term.

That means, the difference between y n plus 1 this is the actual value of y minus y hat of n plus 1. And technically if the errors are assumed with the mean value of 0, then if you have different errors from considering from n plus 1 n plus 2 n plus 3 so on and so forth. You can add them up and; obviously, the expected value being 0 the sum of those error should be 0 and the variance can be found out and it should match with the value based on which we are proceeding.

And now, the another important thing which and I drew the graph also in a very simplistic sense how the errors are there in which directions errors are there, they can be positive error, they can be negative errors and all those things. And for the multiple linear regression also we found out that given the mean values of x I am just repeating it the mean given the mean values of xs and the mean value of the error we can find out the mean value of y, even the standard deviation or the variance of xs and the standard deviation of the variance of error we can find out the standard deviation of the variance of y.

And obviously, when you consider the covariance matrix of all the xs and epsilon and alpha, you should remember that as alpha is a constant hence its covariance with respect to other random variables will be 0 0.1; point number 2 the relationship between the xs that is x i and xj as per the assumption is 0 there is in they are independent. And the covariances between 2 different xs i and j and not being equal would be 0 and similarly we also considered the covariance between the errors and the xs are also 0 hence those terms would vanish. So, only you would be left with the variances of all the xs and the variance of the errors because the variance of alpha is not there.

(Refer Slide Time: 05:01)

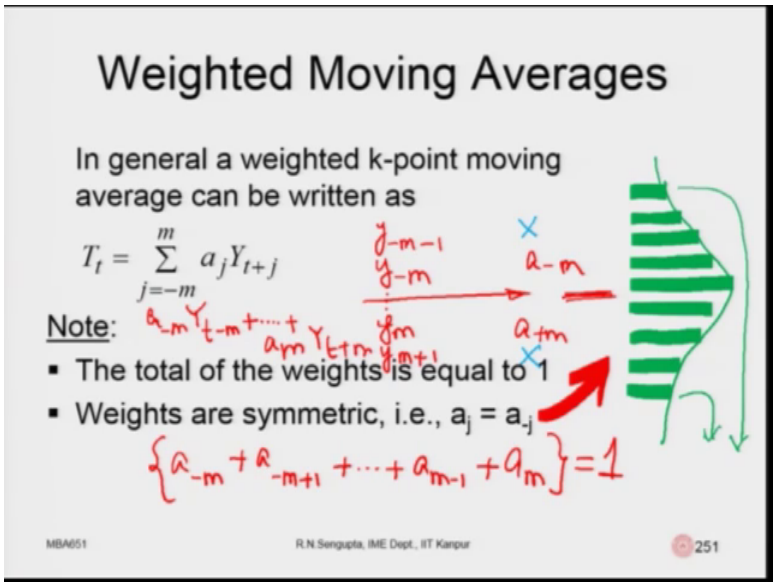


So, I will just repeat it. So, find you should find out beta 1 beta 2 beta k using the concept of minimizing the sum of the square of the errors. This is also known as the least square method or method of ordinary least square the estimated we found out is basically beta 1 hat beta 2 hat till beta k hat they are the estimates of of beta 1 to beta k and they would have the properties.

So, if you remember during the part, where we are discussing about different type of point estimation I did mention time and again the concept of unbiasedness as a consistency. So, basically you find out the concept of unbiasness and consistency for this beta 1 hat to beta k hat. And provided because you assume the underlying distribution for xs epsilon and ys are all normal.

So, we will utilize these estimate to find out the forecasted value of y, which is basically y hat or small y whichever the notational concept you are using and compare those with the actual values of y obtained and then basically say something about the error.

(Refer Slide Time: 06:07)



Now we will consider something to do with I will come to the problems later on. So, in general we will consider another area, which is basically weighted moving average. So, weighted moving average is basically the concept where you where you give different weights to the past values and based on the weights of the past values you try to basically predict the future values with least amount of error with respect to the actual value.

So, this is not related to definitely not multiple linear regression, because in multiple linear regression we consider that they are the y is dependent on the past data of independent variables considering x 1 to xk. But here you will basically try to find out the estimated or the actual and compare with the actual values of y respect to the past data of y only.

So, in general a weighted k point moving average can be written. So, if you want to find out the k number of times of the moving average here, you want to go back and find it out. So, it will be you will take equal weights or equal distances from the positive of the negative sides, how you do that I am going to come to that later. So, you will take the j values from minus m to plus m; that means, you center your actual predicted value based on the weighted concept at the middle and why we do at the middle? That will become very clear as I basically give the concept with respect to the histogram.

So, you will give weights. So, the weights which will be given for the t minus m value would be a minus m, similarly the value which will give to t minus m plus 1 value would

be a m minus minus m plus 1 and go on till say for example, the last value which would be the value of would be Y t m plus 1 all this is are t m plus 1 t minus m or t minus m plus 1 all these are coming in the suffixes; and the corresponding weight would be basically be given by a m.

So, what we are doing is I will draw it so; the values are I will draw it vertically. So, it will be easy for you understand. So, the values are given. So, consider you have y t plus minus m. So, I am going in the downward direction. So, its m minus m space is less minus m minus 1 y minus m dot dot dot dot till y m y m plus 1.

So, I want to find out the middle value somewhere here. So, what I will do? I will give for this; obviously, I am not taking any weights because this is out of the purview because the summation is being done for minus m to plus m. So, I will give a weight of a minus m go, this is also not there. So, this would not come I am putting a cross here, cross does not it is basically no not of variable or symbol is just a cross is not considered. I think you would utilize another color to give the cross let me use the blue one; this is a cross this is also cross. So, the value would be a plus m. So, you are basically summing it up, the sum when you do would be a minus m Y t minus m plus dot dot dot dot and that is like I write the last value would be a m Y t plus m.

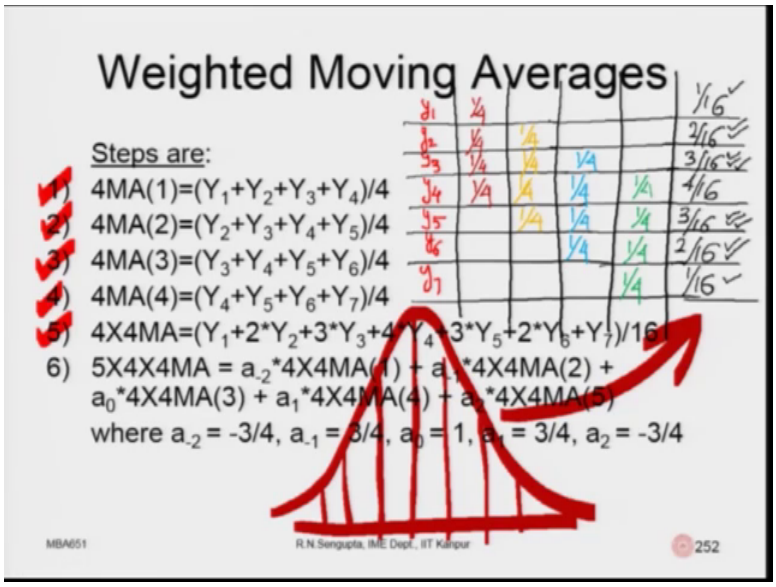
So, if you consider the weights technically, further you go away from the actual value like you are going more to the left or more to the right; that means, if I am standing here I am going more to the back or to the front, the weights would basically diminish accordingly. So, if I give consider the weights as some sort of histogram and I draw the histogram considering a different color and consider I will use the highlighter with this colour. So, I will use this so, the weight for the first one; that means, minus m would of this side the m minus this one sorry then it would be this one I will mark them out there then it will increase to this one then again. So, it means if you are looking at this diagram, and I am using the midpoints of the histogram I have just inverted it.

So, if I draw it will be a normal distribution, in the sense the weight which you have here and the weight you will have here should be of the equal weightages; hence the what you are trying to do actually do is that trying to give the weights in such a way that will basically form a normal distribution with the central value being the predicted or the estimated value using the moving average concept. Note that the total weights another

thing, the total weights technically should all be 1. So, if I am basically adding up and use the different color so, a minus m plus a minus m plus 1 plus dot dot dot. So, this is in the bracket am minus 1 plus a m the this sum value should be 1.

So, because it is a convex combination of the weights we are putting such that you are able to achieve your objective considering normal districation is true and you find out the weights in such a way, that they give you the best prediction weight should be symmetric so; obviously, symmetricity what I have implied would be coming here. So, this part would be taken care as I have drawn using the histogram. So, this is basically a simple concept of moving average (Refer Time: 13:50) which we will take.

(Refer Slide Time: 13:51)



Now I will just give the rules, rules are very simple exactly what I mentioned I will just highlight it. So, basically I will take these three. So, I will take this one, the first one second one and third one. So, the first 2 would basically to the third. So, let me start the concept using a different color. Now consider you have values as Y 1, Y 2, Y 3, Y 4, Y 5, Y 6 and you should have a Y 7 also because it is a 4 year moving average 4 times.

So, I should consider the fourth and the fifth also utilizing this. So, I will utilize the 1 2 3 4 5 in order to explain you this. So, I should basically redraw. So, this is y 1, y 2, 3, 4, 5, 6, 7. So, these are the are the difference. So, and basically highlight it wait. So, this is the first. So, I am just drawing a division in order to make things easy so; obviously, they would be more values, but I will take as given here. Now what does 4MA MA 1 mean? 4

means 4 months taken for 4 time periods taken MA means moving average 1 means the for the first instant.

So, what I do is that I take the first one with one fourth probability because I am taking 4. So, if it is three it will be one third, it is 6 it will be one sixth second one with one-third probability, third one with one-third probability fourth with one-fourth sorry one fourth probability. So, add addition of all the weights are one point one point number 2, I am giving equal weightages now, but you will see later on if you remember it was symmetric as I have drawn that normal distribution. For 4 MA 2, 4 month moving average for the second set of calculations, again I use a different color let me use the orange.

So, now I am using one fourth here, one fourth here one-fourth here and one-fourth here; that means, I am starting at Y 2 and ending at Y 5 naught at Y 1 and ending at Y 4. So, I am just shifting by one notch for 4 MA 3 and basically start with one fourth 4 y 3 one-fourth for y 4 one-fourth for y 5 one-fourth for y 6. So, this is 4 MA 3 4 MA 4 again it will move one notch it would be again one fourth weights, and the weights would be given in this way. It will be one-fourth for y 4 one-fourth for y 5 one-fourth for y 6 one-fourth for y 7.

So, if I so; obviously, if I find out total probability, total probability for this would be let me use the coloring scheme as black. So, it is easy for me. So, the total weights here would be if I find out the whole weights so; obviously, the weights in between for y 4 with the highest. So, y 4 would basically be I am busy leaving the ratio in those ratios for y 1 it will be 1 by 16 and for the last one it is also 1 by 16, as point 1.2 for the y 2 and y 6 they would be 2 by 16 this is 2 by 16 for third and fifth it will be 3 by 16, 3 by 16, and the fourth one it will be 4 by 16.

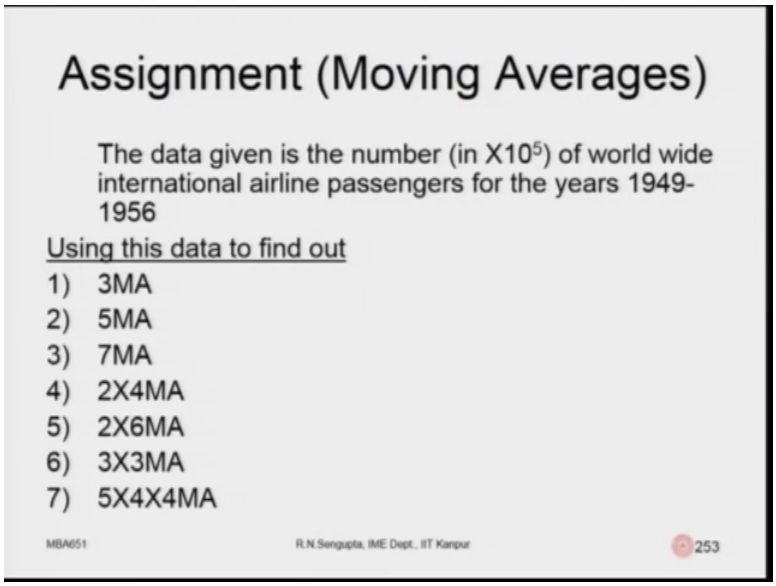
So, if you see the weights this one and this one are equal, this one which is y 2 and y 6 are equal y 3 and y 5 are equal and similarly the middle one is the highest. So, if you look at this histogram and I draw it. So, it is and we draw the histogram this is the base line I am just inverting it. So, one-sixth weight comes here two sixth would be little bit higher, three-sixth would be little bit higher three times. So, let me draw it the three times four-sixth would be much higher this is a proportion not may not be right diagram, but I am trying to basically highlight, then the third one would be the last third from the

last would be again height of three-sixth second from the last is 2 sixth and other one is one-sixth.

So, if I draw the histogram it will a normal distribution exactly as the values have been predicted now if I do it say for example, 4 3 so; obviously, the calculation will change. So, in the case first case it will be one-third, one-third, one-third by y 1 y 2 y 3 and again one-third, one-third, one-third for y 2 y 3 y 4 and one third one third one third for y 3 y 4 y 5 find out the width and proceed accordingly just one thing the weight should be exactly equal to one. So, if you have 2 plus 1, 3 3 plus 3 6, 6 plus 4 10; 10 plus 3 13; 13 plus 2 is 15; 15 plus 1 is 16; 16 by 16 is 1. So, which you basically collaborates the fact that the assumption what we have started with is true.

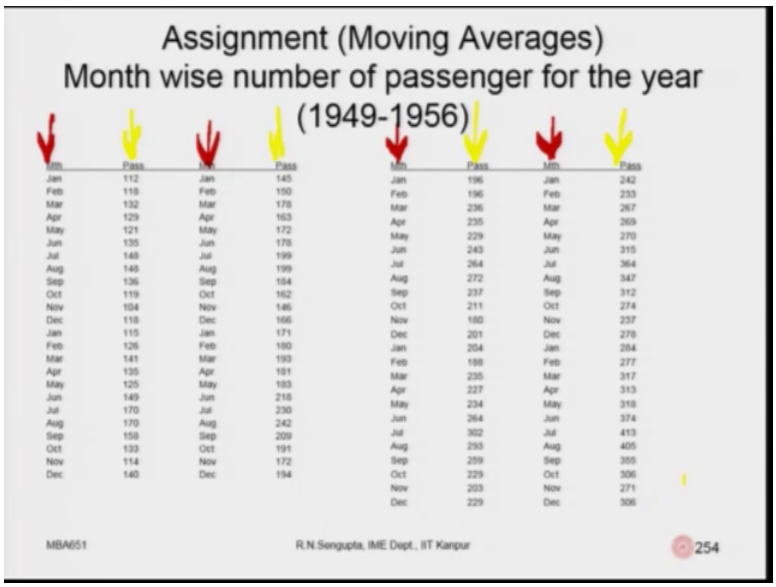
Now remember that this moving average concept whether you are taking 4 month three month 5 month 2 month 7 month whatever it can be changed accordingly depending on what is the level accuracy we want. So, in the case of 5 cross 4 cross one in the case of basically first case, what was 4 plus 4 was 4 month you took it for 4 times and then you did the average. In 5 cross 4 cross 2 you will basically doing it the next stage in order to basically bring more normality in the case as the weights are more spread out and the sum of the weights are is equal to one; obviously, that will be true. And you can do the calculations accordingly put negative weights positive weights such that the overall sum is one and it basically mimics, which would be symmetric mimics the normal distribution.

(Refer Slide Time: 22:33)



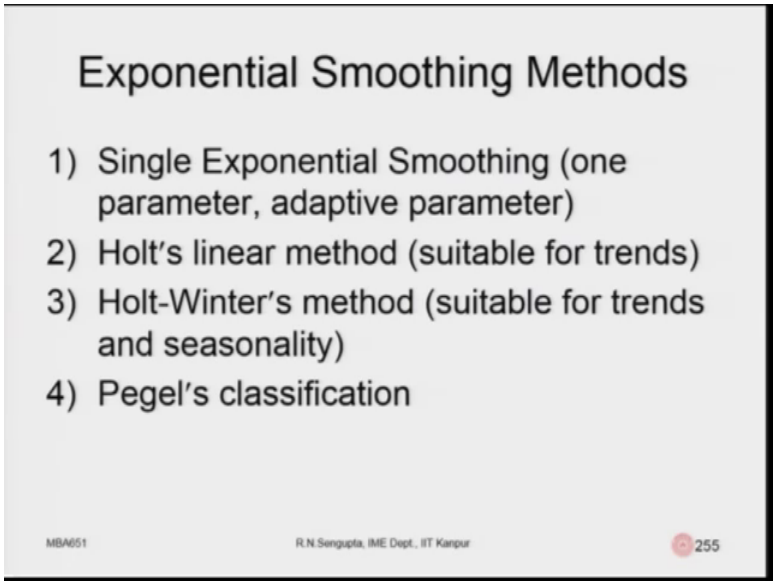
Now, we have the data given in the number in 10 to the power 5 of worldwide international airline passenger for the year 1949 to 1956. So, you have to use this data to find out the 3 month moving average, 5 month moving average, 7 month moving average 0.1, then you find out 4 month moving average for 2 times, 6 month moving average for 2 times; that means, you are trying to find out the average; 3 month moving average 3 times and 4 month moving average into 4 month moving average into 5 depending on the accuracy which you want.

(Refer Slide Time: 23:11)



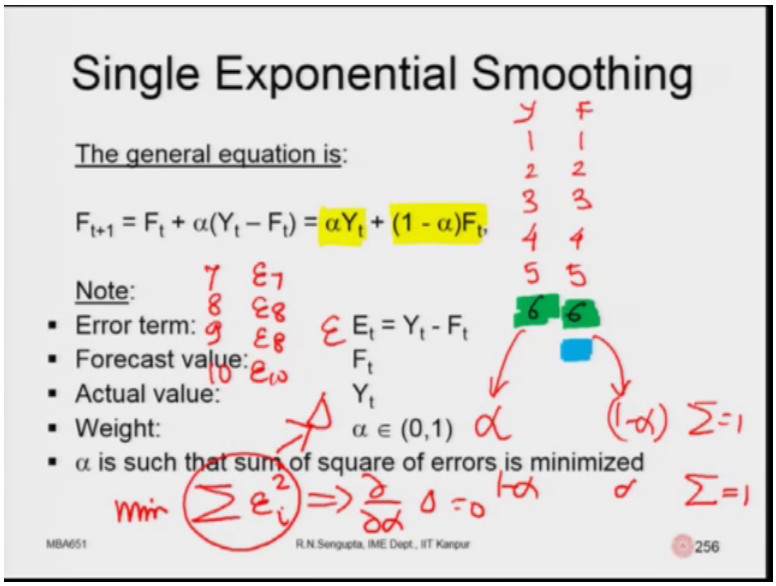
So, this is the data which you have from 1949 to 1956, the 1st columns which you have here as well as the 3rd column as well as the 5th column as well as the 7th column. So, 1 3 5 7 are the months and the corresponding values are the passengers 2nd, 4th, 6th and 8th. So, basically they start from 112 in January 1949 and n to 306 in December 1956 and you can do it.

(Refer Slide Time: 23:59)



So, based on that you can find it out so, if its 3 month 4 months 6 months you can do it accordingly. Now exponential smoothening methods I will come to one simple solution later on. So, exponential smoothening one parameter and adaptive method, this is one method we will consider then the Holt's linear methods which is suitable for trends and there is Holt's winter method, which is suitable for trends and seasonality and the Pegel's classifications which also consider in some details later on.

(Refer Slide Time: 24:33)



Now, in the single exponential smoothening method what we have is basically you give weights to both the forecasted value as well as the predict as the actual value from the past. So, consider that you have F t which is the forecasted value and Y t is the actual value. So, what we will do is that, if I am standing today in say for example, January 2017, then I will use the predicted value as well as the actual value for December 2016 in order to predict the best and; obviously, I will give weights accordingly.

So, if it is F t plus 1 that means, 40 plus 1th T period, the forecasting is being done, it will basically give weight or will be utilize depending on the forecast for the tth period one period back and the error of one period back; that means, you are giving weights accordingly to the actual value which is Y t and the predicted value which is F t.

So, in that case what you have? You are giving a weight of alpha to the actual value for time period t, and 1 minus alpha because the sum of the weights is equal to 1 to the value of the predicted value one period pass. So, what you have is actually take this. So, you have y, you have F. So, you have 1 1 I am just writing the time period 2 2, 3 3 4 4 5 5 6 6. So, if you are basically trying to predict for predict for the 7th one 7 time period. So obviously, the actual has not arrived yet for Y 7. So, you are trying to basically be utilize 1 2 3 4 5 6 7 data points for the 6 time periods using the Y and F to predict f for the 7th period. So, when you are trying to do that you will try to utilize the actual and the

predicted for the 6th time period which is. So, you give a weight it can be anything between alpha 1 minus alpha or 1 minus alpha alpha it does not matter.

So, you give a weight of alpha here, give a weight of one minus alpha here so, as the sum is equal to 1 or it could be basically 1 minus alpha here alpha here. So, the sum is equal to 1. Now the question would arise that can I go more back; that means, rather than 6th period can I go into 7th period can I go into no sorry can I go into 5th period can I go into 5th 4th period answer is yes. So, more back you go you are trying to basically utilize the information from the past in order of predict for the future.

So, here the error terms is Y t minus F t which is basically epsilon as we write the forecasted value is Y t, the actual value is forecast value F t actual value is Y t. Weights are given between 0 1 inclusive and alpha is such that the sum of the square is minimized. So, what we are actually telling to do is that for the 7th period. So, I will just write the period number 7th, 8th, 9th, 10th.

So, you basically find out of epsilon 7, epsilon 8, epsilon 9, epsilon 10 go on such that you square the errors sum them up it should basically minimum in this case when you are trying to be basically minimum and you will ask the question what is their not known its alpha. So, you will basically in order to do that you minimize with respect to the so, consider this is delta you minimize that put a 0 and solve the problems accordingly.

.So, with this I will end the 24th lecture and continue with the discussion in the 25th and so on and so forth have a nice day.

Thank you very much for your attention.