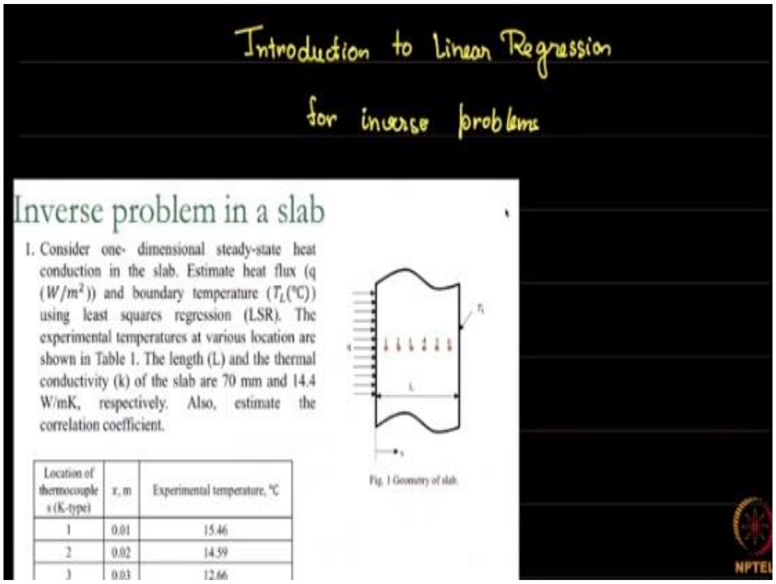
Inverse Methods in Heat Transfer
Prof. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology, Madras

Module No # 02
Lecture No # 06
Introduction to Linear Regression for Inverse Problems

(Refer Slide Time: 00:19)



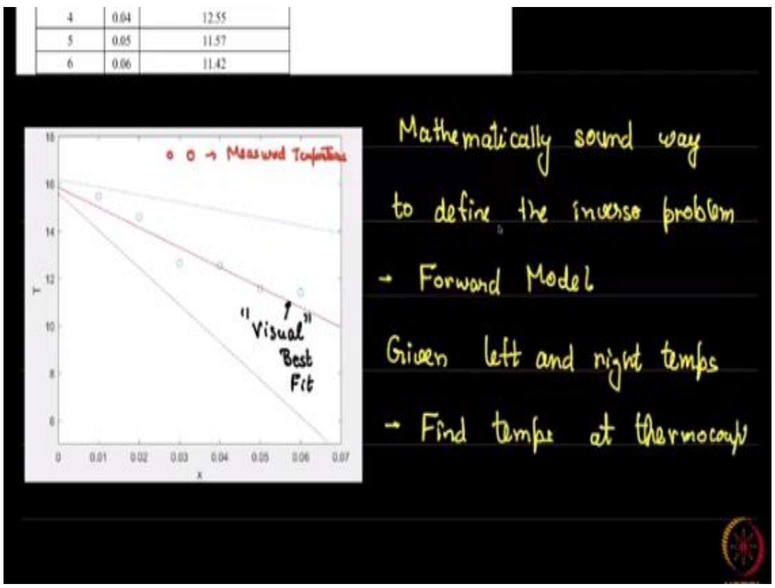
Welcome back. In this video, we are going to look at an introduction to linear regression. If you have seen linear regression before, this will just be a repetition for you. But we are going to do this in the context of inverse problems, as you know the theme of this week is linear models for inverse methods in heat transfer. So, I am just going to repeat a brief context of this problem that I did in the previous video also, just quickly, and then we will come back to how linear regression is relevant to this.

And then we will look at some details of linear regression here. So, recall this problem that i had shown before about the inverse problem in within a slab. So, for example, you could consider this slab here and let us say it has a few thermocouples, here we have taken 6 uniformly placed thermocouples within the slab. And let us say we know the length of the slab; we also know the thermal conductivity and we have some measured temperatures here at these positions at these specific positions.

And we want to find out what the temperatures on the left and right are? So, this is as I had labeled last time too so we are going to call this T_0 and T_L. Suppose we do not know what these are this as I said in the previous week is an example of an inverse problem. So, this is what we wish to figure out and once of course you can figure out T_0 and T_L you can also figure out the heat transfer. So, this 2 let us say we do not know given the thermal conductivity, so that you know from a simple basic forward heat transfer.

Now what makes this problem relevant or slightly difficult as we had discussed earlier is the fact that the solution is not unique. So, as we had discussed earlier this non-uniqueness is a feature of inverse problems. Why is the solution non-unique this should be fairly clear when we look at the actual plot.

(Refer Slide Time: 02:32)



So, if we look at this plot the circles here are actual measured temperatures. So, these circles here are measured temperatures. So, these points here, refer to measure temperatures and these x, of course, are our locations. So, at these locations we have some specific measure temperatures, but we want to figure out what the temperature at the left and what the temperature at the right would be and depending on what fit we choose?

Suppose we choose this fit as linear then you get a different measure temperature and if you choose this you get a different measure temperature. All of us can see that probably somewhere like this is what we would call the visual best fit. So, this one looks like visual best fit. This is what looks the best but is that all can we have a more mathematically nicer way of putting this. So, the question really is what is a mathematically sound way to define the inverse problem?

Now I am going to just mildly set up the inverse problem in this video and in the next video, we will actually discuss a formal way to set all inverse problems. But for this video the purpose is to just motivate this inverse problem quickly look at linear regression and see how relevant it is and then move on. So, the purpose here is to just set up the linear regression problem and give you a quick solution which you would be familiar with from school days.

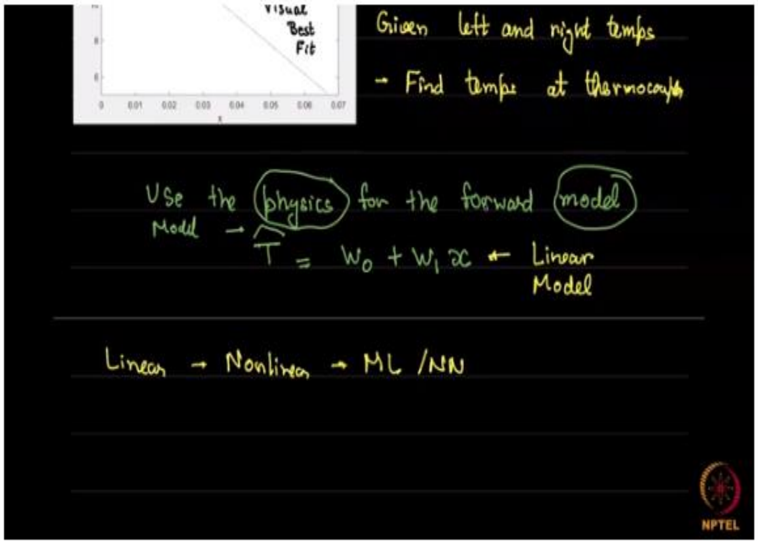
Now once we do that, the point here is that the forward model for this problem. So, the forward model is going to look like this given left and right temperatures, find temperatures at the thermocouples. So, as I discussed in the previous video too, suppose I give you T 0 and T L can you find out T_1, T_2, T_3, T_4, T_5, T_6, which we can and we already saw that the forward model looks like T equal to some function of x,

$$T = a + bx,$$

which is linear, the explanation was then the previous week.

So now we want to go and do the reverse model, which is given T_6 can you find out T_0 and T_L and as we just saw the solution is non-unique. Which one of these fits should I choose? If I choose this, I get a different T_0 and T_L, in this I get different T_0 and T_L.

(Refer Slide Time: 05:55)



But notice one fact we are using the physics. where do we use the physics? we use the physics for the forward model which we will use in the rest of the video. So, this is a model and as follows the physics says temperature, I am going to put a $\hat{T}$, here the hat denotes a model. So, temperature is something some parameter plus some other parameter times x, i.e.,

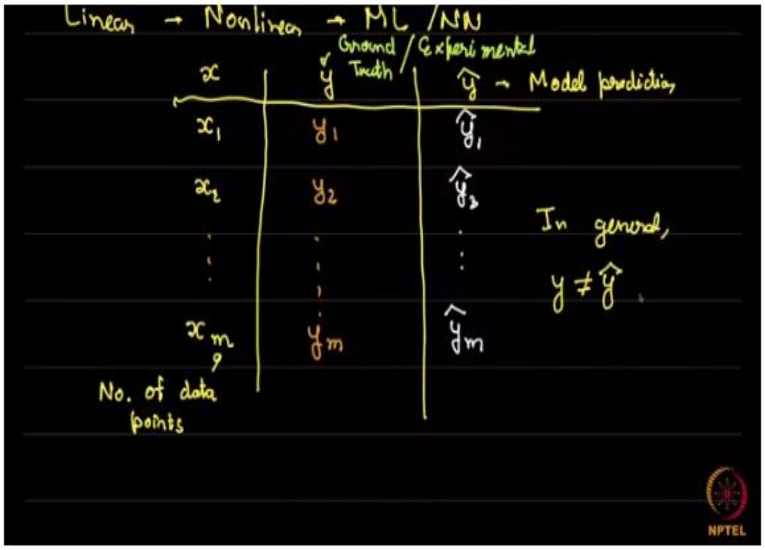
$$\hat{T} = w_0 + w_1 x$$

This information comes from the physics of the problem. So, notice this since this problem is linear or this equation is linear, we call this a linear model.

And we will deal with the fact that linear models themselves can be extremely powerful and we will deal with those both this week as well as the next week and in the subsequent weeks we will see how to move on to non-linear models. Now some of you might think, what happens in case I have no direct information from the physics. suppose I have a turbulence problem we know that no such formula is forthcoming what do we do in such cases?

So, we have several options one of those options is machine learning which we will be covering towards the end of the course. So, the progression is as follows as far as the course is concerned linear, then we use linear and non-linear and as you will see we will use these ideas that we have developed in non-linear to build machine learning and neural network models. So, this is the hierarchy of models as you can see even these are reasonably simple models. we will look at more complex models within the final week of this course, I will just give you a brief introduction but that will lead you to state of the art.

(Refer Slide Time: 07:37)



So let us now come to this linear model. If You observe all we have done is as follows. We are made a table. So, x here is the location so I have multiple locations x_1, x_2, I went till x_6, but let us say in general we go till x_m; so, where m is the number of data points. Similarly, I have y, let us call this the ground truth. Now this is not a language that is used as a standard language within the inverse methods community, but this is a standard language which is used within the machine learning computing.

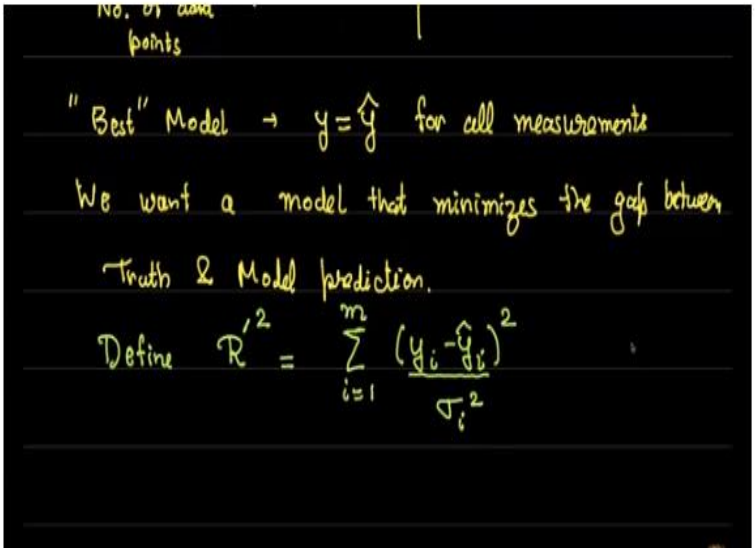
I am going to stick with it because I expect that as you go further and further in your career and you use these methods that you will encounter machine learning more and more, as you can see it is getting more and more common within science and engineering to see this. So, I am going to use this word ground truth, if you wish you can say this is the experimental measurement. So the measured temperature which I am going to call y here, y would be any variable that we measure but in this specific case, this is temperature that we are measuring.

So, we have let us say y_1, y_2 so on and so forth up until y_m so the m^{th} location has a temperature y_m. The third thing that we all carry is $\hat{y}$. This is what comes from my model. Now suppose my model is $w_0 + w_1 x$, for each value of w_0 and w_1 that I guess I will get a different value, for let us say if I take it 0.04, even though the truth is somewhere around 12 point something my model depending on what values I choose for these parameters it might give something like 9 or it might give something as high as 15.

So, this value here $\hat{y}$ is the model value or the model prediction. We denote the fact that it is a prediction by this hat on top. So similarly at each point once you know the model parameters, you will get some predictions here so these predictions will look like $\widehat{y_1}$, $\widehat{y_2}$, so on and so forth up until $\widehat{y_m}$, alright. So, you can see all that reflected in the picture here. So, for this model, this is $\hat{y}$, for this model this is $\hat{y}$ and this of course is y.

So, in this case this is $\widehat{y_4}$ for some model and this is y_4 for hat for another model in general y and $\hat{y}$ will not match. So that is the thing that we can notice. I will write that down in general y is not equal to $\hat{y}$. But another thing is true in general we want a great model.

(Refer Slide Time: 11:23)



So, the best model would be such that $y = \hat{y}$, for all measurements. Now we will see later on in this course that this need not actually be the best, but we have an intuitive sense. That in case you are able to get your prediction to match the truth. So, remember y is the truth and $\hat{y}$ is the prediction, in case the 2 match for every single measurement that you make it looks like you have actually got a great model.

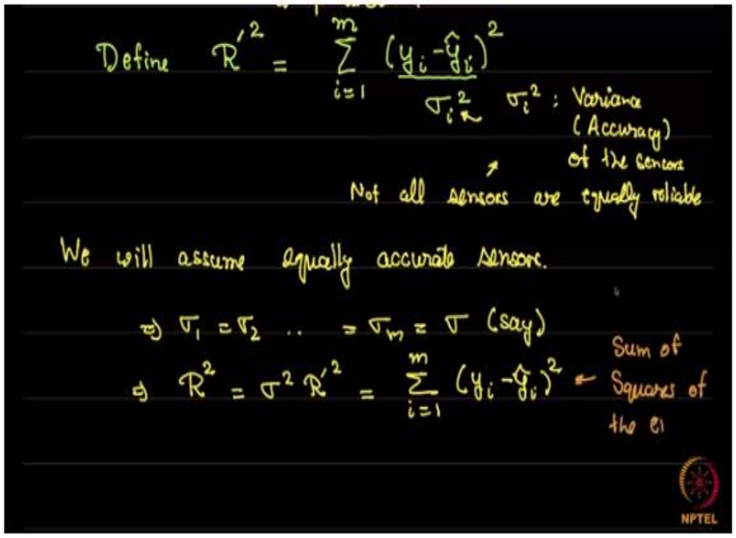
So, what we want is as follows, we want a model that minimizes the gap between truth and model prediction. So, we want some measurement of what the gap between the truth and the model prediction is and the way we go ahead and do that is to define a new quantity. So, we define this quantity called R prime square (R'^2) I will tell you why it is called R prime square (R'^2) shortly.

So, this is a summation of the gap between y_i and $\hat{y}_i$ square, sum of the square divided by something I am going to call σ_i square.

$$R'^2 = \sum_{i}^{m} \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2}$$

This goes from i = 1 to the number of measurements m. what does this mean? so if we go back here to this picture let us take the model below. So, this is let us say model one it looks like a bad model, but let us take this model. So, if we take this model 1 and we look at this gap the prediction is so much, the temperature so much here there is a gap between the 2 you find the gap square it and then divide it by some σ_1^2, find this gap square it divided by σ_2^2 find this gap so on and so.

(Refer Slide Time: 13:56)



Now what is σ_1 and σ_2 so what these are is they are measurements is the variance. I will mention this more when we come to the statistics part of this course. So right now, let us just say accuracy of the sensors. so going back to our example, if we come here, we made these 6 measurements at these 6 points. Now it does not mean that each one of these sensors, some of these might be old, some of these might be new, some of these might be more modern, sometimes some of these might be more expensive so on and so forth.

So, each one of these would have a different accuracy so in general this one could have an accuracy it is like σ_1, σ_2 or have an error within it as $\sigma_1, \sigma_2, \ldots \sigma_6$. So, what we are talking about is σ_1 being a measurement of what the error or that is what the variance within each sensor is. we will make this a little bit more mathematically formal when we go to the statistics portions.

But this account for the fact that not all sensors are equally reliable. Now why does this matter? This matters because see I am saying that how a measure of how good or bad our model is dependent on the gap. Now let us just say that one was a really good sensor a huge gap here should actually be weighted hugely so you want to give a lot of emphasis to that. Whereas a small gap in a good measurement device might mean that even that gap could be amplified by a lot or a big gap in a measurement device, which is less accurate does not have to be that person is or this sensor has always given a lot of error, so maybe the fact that the model and the truth are very different need not be given so much emphasis.

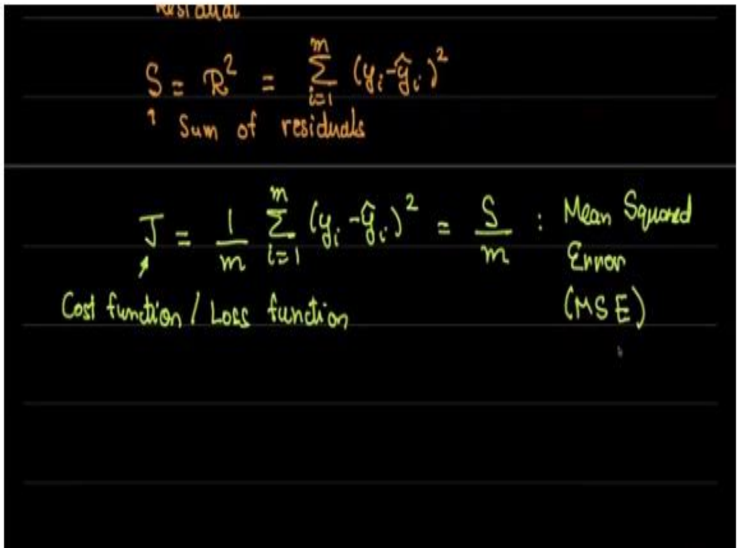
So, what the σ does is gives different emphasis to different sensors. But to start with, we will come to what is known as weighted linear regression which gives this different type of emphasis later. But for to start with we will assume equally accurate sensors.

So that means we are going to assume $\sigma_1 = \sigma_2 \dots = \sigma_{m =} \sigma$. So now we can define one new quantity R^2, I will again tell you what this means this is equal to σ^2 times R'^2. And that is just going to look like

$$R^2 = \sigma^2 R'^2 = \sum_i^m (y_i - \hat{y}_i)^2$$

So, this is simply the sum of squares of the error. This R^2, this R is called residual, R stands for residual why have we put square just to emphasize that this is actually a positive quantity.

(Refer Slide Time: 17:50)



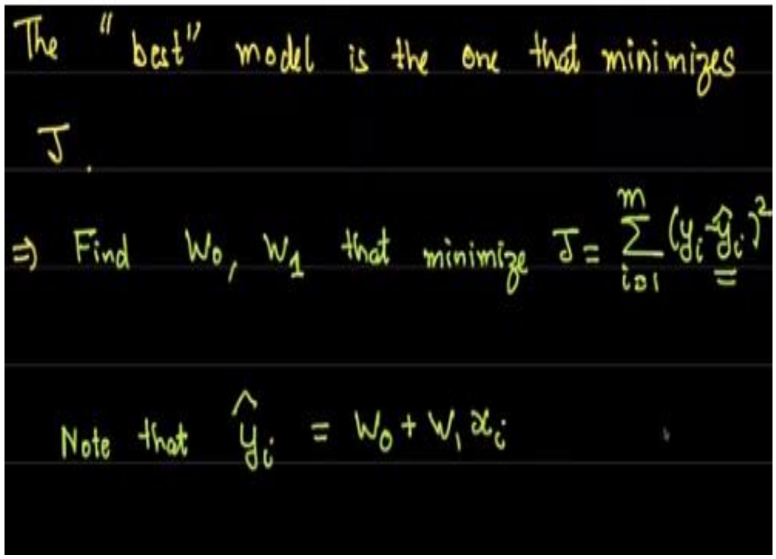
Now there are 2 names, that I am going to use one or the other depending on the notation. So, within the inverse Community this R square is typically called S. S stands for sum, as you can guess, sum of residuals and this is equal to $\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$. Now another name which is more common within the machine learning community is J. So, J is called the cost function, I am not sure why this letter J came I maybe starts with German or something or that sort or the loss function.

All that means is what is the cost of the fact that your model and your reality are different? So, J is mean. So, the mean measurement is

$$J = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 = \frac{S}{m}$$

So, this is basically called the mean squared error. All of you would have seen this at school or even at college also sometimes called which as MSE starting standing for mean square error. So, J is the cost function or the loss function what do we want?

(Refer Slide Time: 19:37)



So, our model, so the best model to solve the inverse problem is the one that minimizes J or S, S is simply one factor the number of measurements multiplying J. So, the best model is the one that minimizes J. Now what are the parameters that we have within J. so this means we need to find w_0 and w_1 that minimize J. And remember J is simply the sum of the square of the error,

$$J = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

Now if you look at it only in this form it might look odd, I mean you want to find out w_0 and w_1 which minimize this but where are w_0 and w_1, where exactly is w_0, w_1 in this expression.

So, let us make that explicit. So, note that w_0 and w_1 is actually hidden within here. $\hat{y}_i$ our model is $w_0 + w_1 x_i$. what does that mean? Let us go back to our example once again which is why we introduce this first we are saying that this here is y this is x this is for example x_1, x_2, x_3 ... y_1, y_2, y_3 … y remember is the truth or the experimental temperature corresponding to this. we should have a $\hat{y}$ which we will see in the next video.

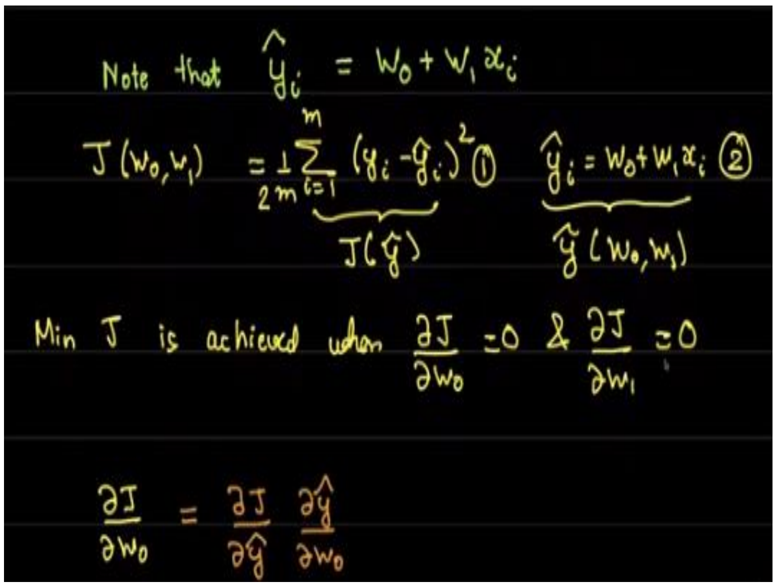
We will see the actual values here. But suppose I guess that, $w_0 = 0, w_1 = 1$. Then the model would have said $w_0 = 0, w_1 = 1$, then,

$$\hat{y} = w_0 + w_1 x;$$
$$= 0 + 1 \times 0.01;$$
$$= 0.01;$$

So, this will be 0.01 so on and so forth. So, if we take this model, this will become 0.02, 0.03 etc. As you can see the gap between y and $\hat{y}$ is huge, this is not a great model. But that is just to tell you how $\hat{y}$ is calculated and $\hat{y}$ itself depends fully on w_0 and w_1.

So, $\hat{y}$ is a function of w_0 and w_1. so please remember this as we go forward. So, $\hat{y}$ is a function of w_0 and w_1. So, we want to minimize J. I will put the mathematical language for this in the next one, but for now let us use a sort of casual language.

(Refer Slide Time: 22:52)



So J depends on w_0 and w_1 and it is given as, I should have put,

$$J(w_0, w_1) = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$$

So, this is equation 1. Whereas,

$$\hat{y}_i = w_0 + w_i x_i$$

This is equation 2. Now when we have 2 variables like this w_0 and w_1 and we want to minimize J, so Min of J is achieved. When,

$$\frac{\partial J}{\partial w_0} = 0; \frac{\partial J}{\partial w_1} = 0;$$

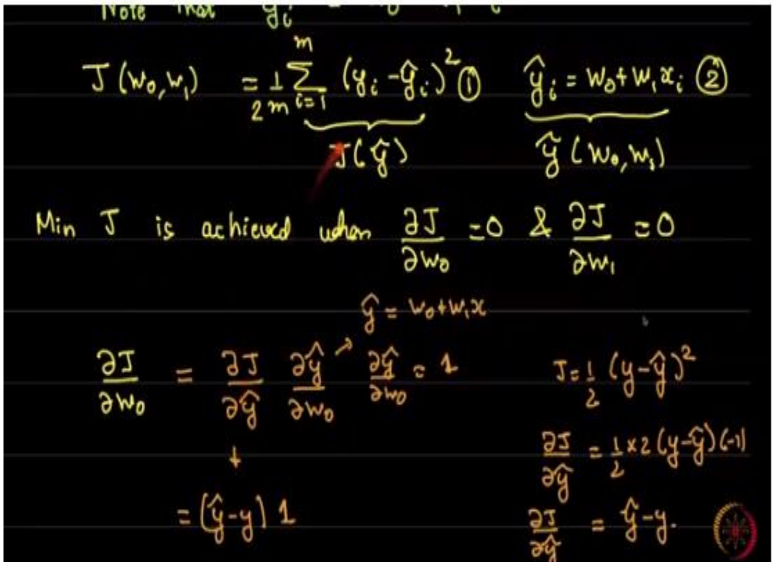
If you do not recall this from your multivariable class then I will be doing this sometime later during the course maybe around the seventh or eighth week I will be again kind of deriving or giving you some physical intuition.

But at least from one variable calculus you know if there was only one variable then it should be $\frac{dJ}{dw} = 0$, if there are 2 variables it turns out that both the partial derivatives have to be 0 like. I said the geometric intuition for this I will talk about later once again towards the you know the third half of the course or the third part of the course. So, now let us see if we can differentiate this equation and get these terms, what are the terms that correspond to $\frac{\partial J}{\partial w_0}$ and $\frac{\partial J}{\partial w_1}$ or what are the equations that correspond to this.

So, $\frac{\partial J}{\partial w_0}$, you notice this is J written simply as a function of $\hat{y}$. This is $\hat{y}$ written as a function of w_0 and w_1. So, we can differentiate this in 2 steps and that turns out to be the easiest way of doing this so $\frac{\partial J}{\partial w_0}$ you can write as,

$$\frac{\partial J}{\partial w_0} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_0}$$

(Refer Slide Time: 25:09)



We want to set this to 0. but let us first start with this equation. Now a convenient thing that we can do just to do this $\frac{\partial J}{\partial \hat{y}}$ is to put a factor 2 up front here, it is not very important but typically we do that we put a half. So that when we take a derivative you know this 2 and 2 cancel out. So, it is a little bit easy. so instead of taking mean square error, we take half of mean square error that is just a convenience that we use.

So if we do that what this comes to. So, if you do $\frac{\partial J}{\partial \hat{y}}$, so I am going to take a simple example and then extend it. So let us say instead of having this $\sum_{i=1}^{m}$, I have a simple equation. so instead of this I am just going to do,

$$J = \frac{1}{2}(y - \hat{y})^2$$

The $\sum$ I am going to erase out. And simply going to keep as if it was just $(y - \hat{y})^2$ and I will tell you how to extend it to the general case shortly.

So now if I do that then $\frac{\partial J}{\partial \hat{y}}$ simply becomes this square comes down here,

$$\frac{\partial J}{\partial \hat{y}} = \frac{1}{2} \times 2(y - \hat{y})(-1)$$

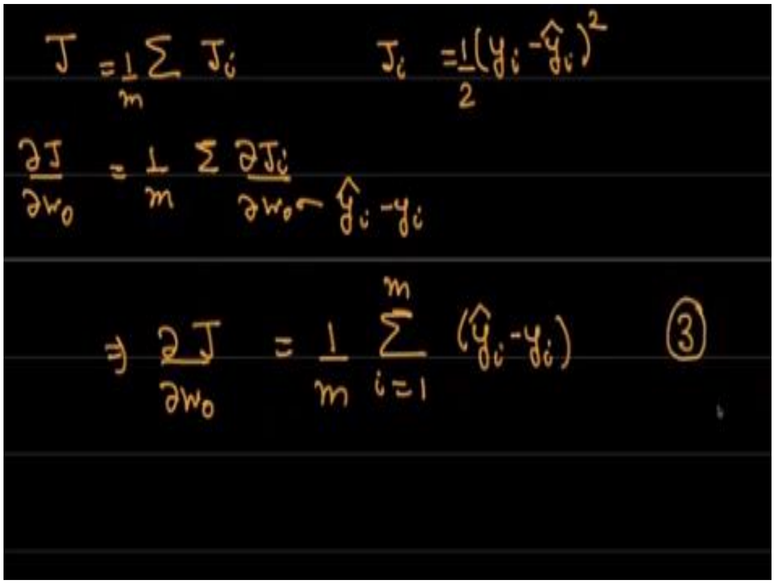
So, I hope this is not too complicated for you, but you should be able to do it fairly straightforwardly if you wish to do this course for credit. So, this is a simple derivative $\frac{\partial J}{\partial \hat{y}}$ is,

$$\frac{\partial J}{\partial \hat{y}} = \hat{y} - y$$

So, I will write that here, $\frac{\partial J}{\partial \hat{y}} = \hat{y} - y$.

Now what about $\frac{\partial \hat{y}}{\partial w_0}$, remember $\hat{y} = w_0 + w_1 x$. I have dropped the i's as you can see, but it is just a convenience so $\frac{\partial \hat{y}}{\partial w_0}$ is simply equal to 1. So, we now have this $\frac{\partial J}{\partial w_0} = \hat{y} - y$. Now suppose I bring this $\sum$ back, which I dropped here. So, this would be for just one term so if I am just taking the first term this would be $\widehat{y_1} - y_1$. Another way of looking at this is like this.

(Refer Slide Time: 27:43)



So, suppose,

$$J = \frac{1}{m} \Sigma J_i,$$

where $J_i = \frac{1}{2}(y_i - \widehat{y}_i)^2$, and I am going to keep a half. All I have done if you look compare the expression here and here is I have just added a subscript, the summation is moving. Then I know that,

$$\frac{\partial J}{\partial w_0} = \frac{1}{m} \Sigma \frac{\partial J_i}{\partial w_0}$$

And we now know based on our expression so far that,

$$\frac{\partial J_i}{\partial w_0} = \widehat{y}_i - y_i$$

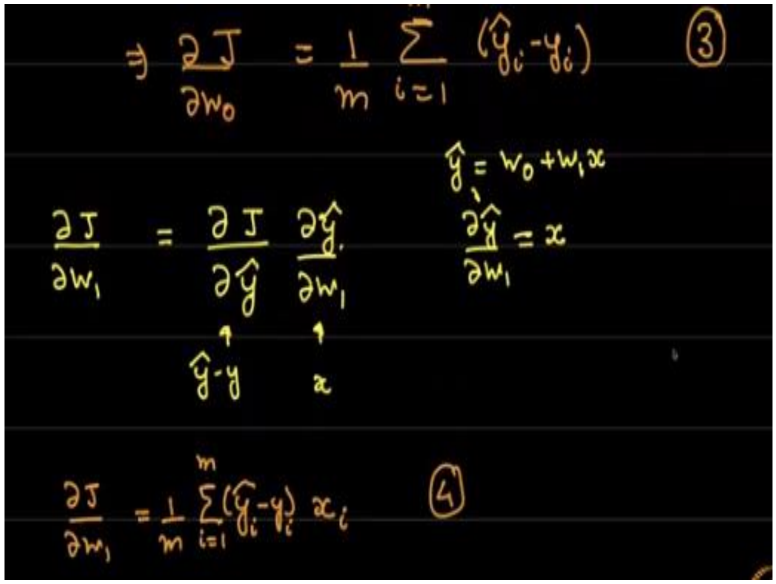
Because all I have done is just replace this J with a J_i and $\hat{y}$ with a $\widehat{y}_i$ right.

Once we do this, we basically get,

$$\frac{\partial J}{\partial w_0} = \frac{1}{m} \sum_{i=1}^{m} (\widehat{y}_i - y_i)$$

So, this we will keep as equation 3.

(Refer Slide Time: 29:22)



Now suppose we do the same thing with $\frac{\partial J}{\partial w_1}$, so I will do the same trick again $\frac{\partial J}{\partial w_1}$, I am going to call it $\frac{\partial J_i}{\partial w_1}$ or let us drop the i for now, just to keep the whole thing a little bit manageable.

$$\frac{\partial J}{\partial w_1} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_1}$$

and $\frac{\partial J}{\partial \hat{y}}$ we just calculated as you saw here, $\frac{\partial J}{\partial \hat{y}}$ is simply $\hat{y} - y$. So, we are going to call this $\hat{y} - y$ and $\hat{y}$ was our model which was $w_0 + w_1 x$. So, $\frac{\partial \hat{y}}{\partial w_1}$ is equal to this derivative the derivative of this with respect to w_1 this is simply x. $\frac{\partial \hat{y}}{\partial w_1} = x$.

So, this is going to be x. So, we can now write this down as,

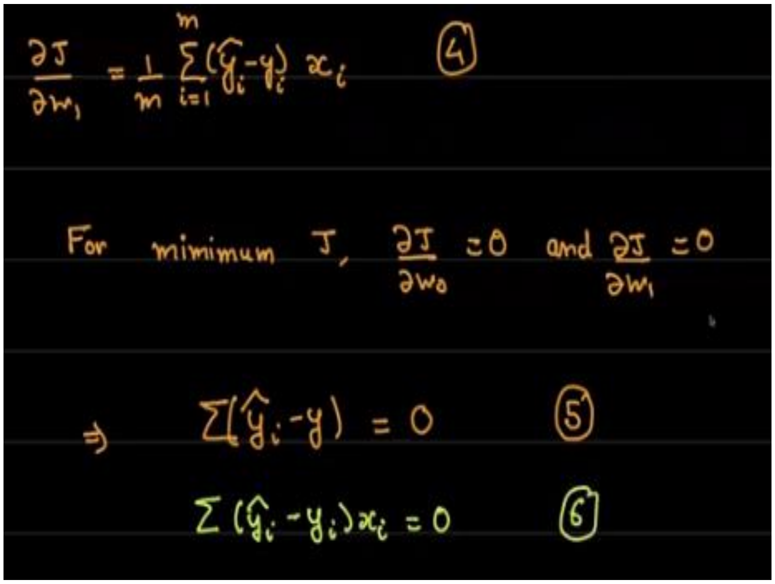
$$\frac{\partial J}{\partial w_1} = (\hat{y} - y)x$$

except, when we actually do this, just like the last time you will get a summation. And get a summation. So,

$$\frac{\partial J}{\partial w_1} = \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)x_i$$

So, this calculation here was just for one sensor this is the full calculation so let us call this equation 4. So, you can notice equation 3 here is for $\frac{\partial J}{\partial w_0}$, equation 4 here is for $\frac{\partial J}{\partial w_1}$.

(Refer Slide Time: 31:13)



For minimum J, we need,

$$\frac{\partial J}{\partial w_0} = 0; \frac{\partial J}{\partial w_1} = 0;$$

So let us put these 2 equations in play. So, this first condition will say $\sum$, I am going to eliminate the divided by m,

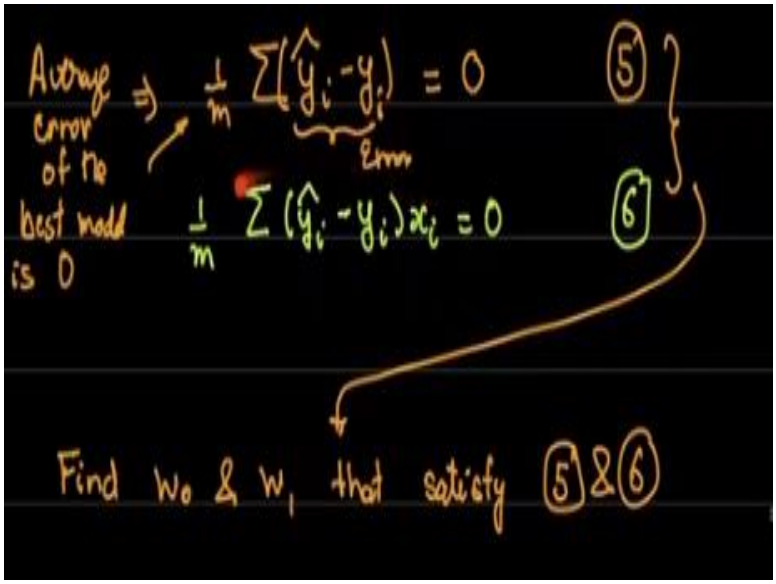
$$\sum(\hat{y}_i - y_i) = 0$$

Let us call this equation 5. And the next equation is going to be,

$$\sum(\hat{y}_i - y_i)x_i = 0$$

This is equation 6. So, these are the 2 equations. Now by the time you come to these 2 equations we might forget. What it is that we are looking for?

(Refer Slide Time: 32:29)



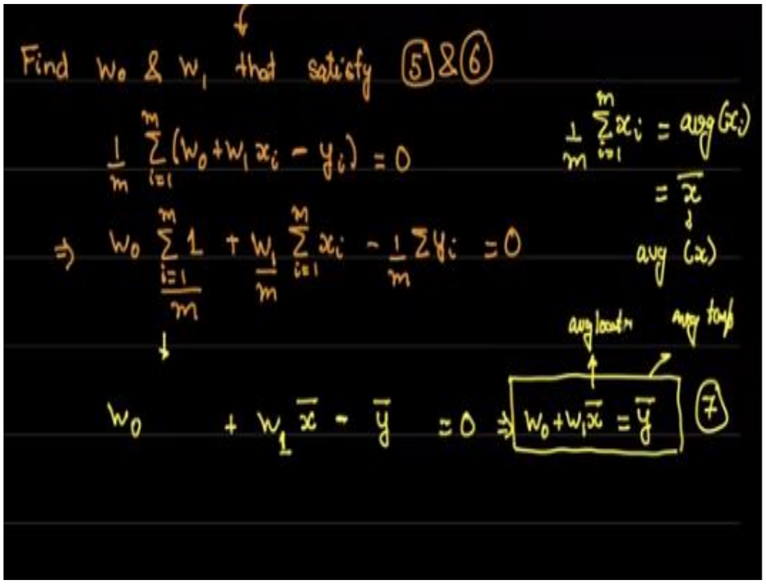
So, we should remember what it is that we are looking for in fact let us retain the m so that might make things a little bit more convenient for us in a little bit of time. So, these are the 2 equations and we want to solve these 2 equations for what were our unknowns. our unknowns are w_0 and w_1, that satisfy equations 5 and 6. That is really what we are looking for all right.

So, at this point you might think what is the meaning of these 2 equations? So, what is the meaning of equations 5 and 6 and how do we proceed in solving for w_0 and w_1? So, let us look at that in a short way. what does this mean? so what this first thing means? This simply means that the average error of the best model is 0. This is what this equation means. you notice this is the error and that is why I retain the m, what this says is the average error of this model is 0.

This one says that the weighted average by x is also 0. This one is a harder meaning, but at least let us look at the physical meaning of the first term here if you notice is look is looking like this. So, suppose this middle red line is our model, you will see that somewhere the error is positive and at the other places the error is negative, whereas the bad models are all on one side. So at least we have a physical intuition that a good model should go on both sides of the truth and that is the intuition, that this equation here is re-emphasizing.

This on the other hand is a little bit harder to interpret and we will see that in the next video so let us see how to solve these 2 equations. Remember we are looking for w_0 and w_1.

(Refer Slide Time: 35:04)



So, equation 1 is,

$$\frac{1}{m}\sum(w_0 + w_1x_i - y_i) = 0$$

Remember y_i is the truth, x_i or the actual temperature, this is the location of the sensor and these 2 are the parameters we are trying to solve for. it is useful to just remember what we are solving for throughout. So, this is 0 this is equation 5 let us just continue with simplifying this $\sum_{i=1}^{m}$, w_0 is a constant with m or constant with i, the only thing that depends on i here is x_i as well as y_i.

So, now if we look at the first term the first term simply is $\sum w_0$, which is a constant which you can take out. So, this is w_0 times $\sum_{i=1}^{m}$, I am just doing it term by term this just becomes 1, the whole thing divided by m. I am going to take each term then w_1 terms times $\sum_{i=1}^{m} x_i$, the whole thing divided by m and the final term is minus $\sum_{i=1}^{m} y_i$. Let us just assume that 1 to m is 0.

$$\frac{w_0}{m}\sum_{i=1}^{m}1 + \frac{w_1}{m}\sum_{i=1}^{m}x_i - \frac{1}{m}\sum_{i=1}^{m}y_i = 0$$

Now let us look at these terms one by one, again, in the first term is simply $\sum_{i=1}^{m} w_0$, is simply m. so, m by m, so that is 1. This is $w_0 + w_1$, let us use some terminology. This is the average value so if you see, $\sum_{i=1}^{m} x_i$, this is the average value of x_i. So, the average location across the domain. So, we are going to denote this average by a bar, so $\bar{x}$ denotes average value of x over the domain.

So w_1 times this term becomes $\bar{x}$ minus this term, now you can see is the average value of y, so we will denote this by $\bar{y}$.

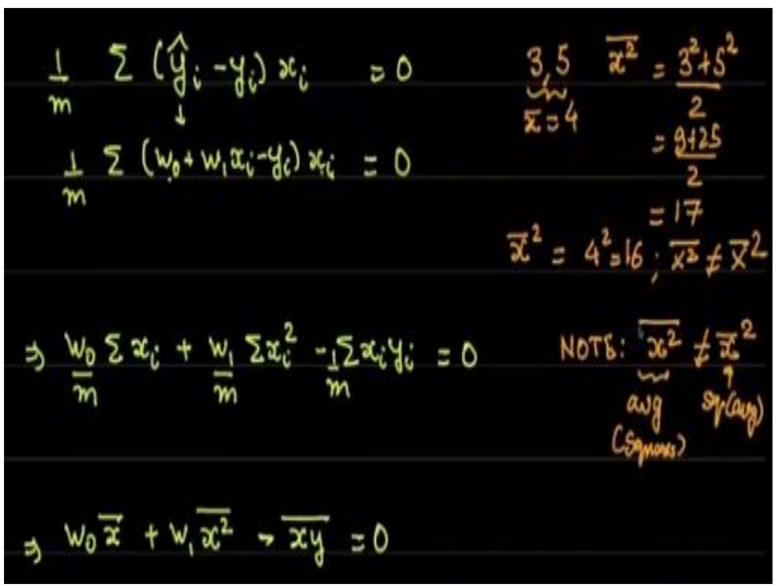
So now you have an equation I am going to write this compactly,

$$w_0 + w_1\bar{x} - \bar{y} = 0$$

$$w_0 + w_1\bar{x} = \bar{y}$$

So, notice this this is one equation notice how nice it looks also. Because it tells you at the average location, I should accurately predict the average output or in this case the average temperature so this equation actually has a physical meaning so this let us call this equation 7. Now let us simplify the next equation which was equation 6 so equation 6 is here this equation so I am going to write that down.

(Refer Slide Time: 38:12)



$$\frac{1}{m}\sum(\hat{y}_i - y_i)x_i = 0$$

So, you can see it is exactly the same as the previous equation but like I said it has a weighted weighting here x_i, so we will open this up,

$$\hat{y}_i = \frac{1}{m}\sum(w_0 + w_1x_i - y_i)x_i$$

I am going to drop the $\frac{1}{m}$, just for keeping our notation a little bit compact. So same thing the first term is $w_0 \sum 1$ the whole divided by m it is, I made a mistake here $\sum x_i$, so w_0 times x_i, so w_0 can be taken out. So $\frac{w_0}{m} \sum x_i$.

The next term is $\frac{w_1}{m}\sum x_i^2$ and the final term is $\frac{1}{m}\sum x_i y_i$ this whole thing equal to 0.

$$\frac{w_0}{m}\sum x_i + \frac{w_1}{m}\sum x_i^2 - \frac{1}{m}\sum x_i y_i = 0$$

So now let us do the same interpretation that we did the last time and write this down. So, the first term becomes w_0 times this combination of terms, so that is simply $\bar{x}$, you can say $\sum x_i$ divided by the average value of x, + w_1 times now this is the average value of x square.

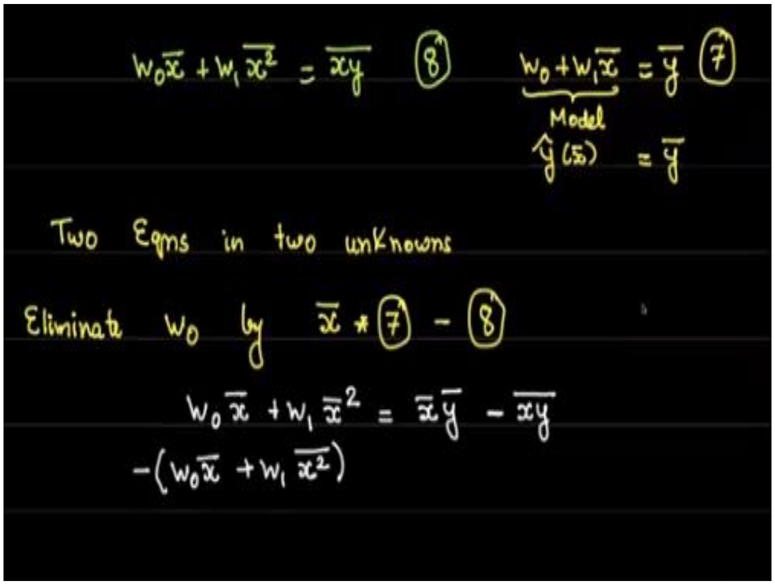
So, the square of the location is what we are averaging at this point so this is $\overline{x^2}$. Now it is important to note something here this is a mistake that students often make or people often make. $\overline{x^2}$. is not the same as $\bar{x}^2$ what is the difference? $\overline{x^2}$ is the average of the squares and $\bar{x}^2$ is the square of the average. So, for example if you have the numbers 3 and 5 then $\overline{x^2}$ would be 3^2 + 52 by 2 so that is 9 + 25 by 2, 34 by 2, so 17.

Now $\bar{x}^2$ is, what is $\bar{x}$? $\bar{x}$ is 4, the average of this is 4. So, this is 4^2, so this is 16. So, you can notice $\overline{x^2}$ is not the same as $\bar{x}^2$. So, what do we have in this term? This is x^2 average so that is $\overline{x^2}$. so please do notice that it is easy to get confused by this term quite easily. So, you see this term here this is average of x_i y_i so this is $\overline{xy}$,

$$w_0\bar{x} + w_1\overline{x^2} - \overline{xy} = 0$$

$$w_0\bar{x} + w_1\overline{x^2} = \overline{xy}$$

(Refer Slide Time: 41:39)

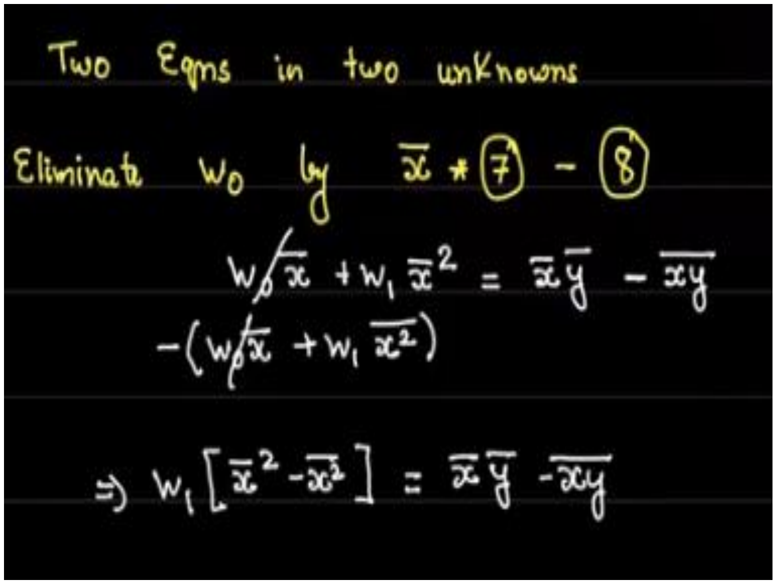


So again, the same thing holds true here too also note that $\overline{xy}$is not the same as $\bar{x}\bar{y}$, this is the average of the product and this is the product of the average. These 2 are not the same and we will see this in greater detail when we come to the probability portions of this course. So we can put this together and write an equation $w_0\bar{x} + w_1\overline{x^2} = \overline{xy}$. So this will call equation 8 just for your recollection equation 7 was $w_0 + w_1\bar{x} = \bar{y}$, this was equation 7.

I am just rewriting here so that it sits on the same page for you for easy reference. Now notice this again, this is a useful thing to remember. this is called taking moments unfortunately. we would not have the time to discuss this in great details in the course. This basic equation if you remember you can generate this equation fairly easily so this equation is just $w_0 + w_1\bar{x} = \bar{y}$. All I am saying is, this is coming from the model remember.

So, this comes from the model. So, what it is saying is $\hat{y}$ at an average location should be equal to the ground truth averaged. So, from there if you just multiply each term notionally by $\bar{x}$ or you first multiply each term by x and take an average, you get this equation. So, you can see $\bar{y}$ becomes $\overline{xy}$, $\bar{x}$ becomes $\overline{x^2}$ and 1 becomes $\bar{x}$. Now we have 2 equations and two unknowns and we can solve this to obtain w_0 and w_1, it is a fairly easy to do that so the way to solve this is. Let us say we multiply we eliminate w_0 by multiplying equation 7 by $\bar{x}$ and subtracting out equation 8. So, $\bar{x}$ multiplied by equation 7 - equation 8.

(Refer Slide Time: 43:51)



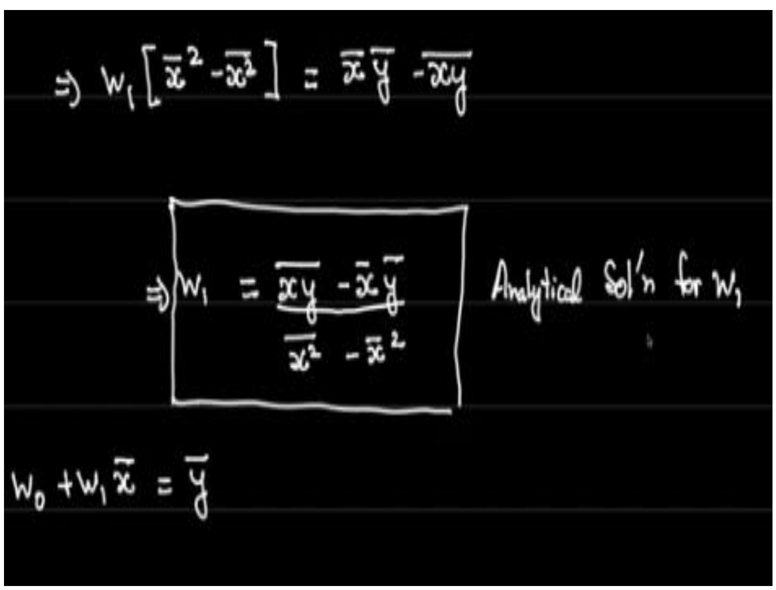
So, what that gives you is,

$$w_0\bar{x} + w_1\bar{x}^2 = \bar{x}\bar{y}$$

This is the right-hand side and now you subtract out this equation terms, which are $w_0\bar{x} + w_1\bar{x}^2$. So, this is where the distinction between $\bar{x}^2$ and $\overline{x^2}$ becomes important. On the right-hand side, you subtract out $\overline{xy}$. So, these 2 terms cancel out. So, you get,

$$w_1[\bar{x}^2 - \overline{x^2}] = \bar{x}\bar{y} - \overline{xy}$$

(Refer Slide Time: 45:29)



So, you can now write w_1 equal to, I am going to reverse the signs, because that is the usual way in which it is written,

$$w_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$$

So, this is the analytical solution for w_1. of course, we need w_0 also and we know that $w_0 + w_1\,\bar{x} = \bar{y}$. so, this gives us,

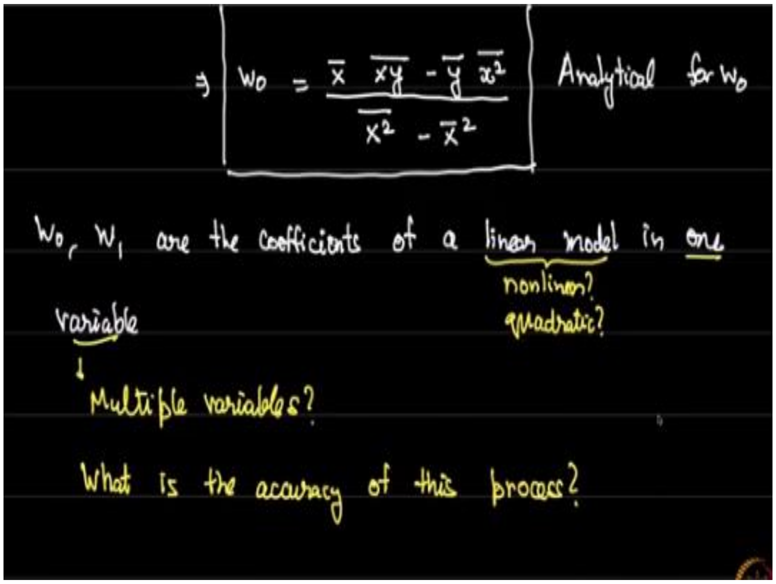
$$w_0 = \bar{y} - w_1\bar{x}$$

and you can write this in a convenient way, w_0 becomes $\bar{x}\,\overline{xy} - \bar{y}\,\overline{x^2}$, I will leave the derivation up to you so fairly straightforward derivation, divided by $\overline{x^2} - \bar{x}^2$.

$$w_0 = \frac{\bar{x}\,\overline{xy} - \bar{y}\,\overline{x^2}}{\overline{x^2} - \bar{x}^2}$$

So, this is the analytical expression for w_0. so put together, these are the coefficients of the linear fit of a linear model in one variable.

(Refer Slide Time: 47:25)



So, this very specification here tells you a few things. we have automatic questions that will come from here. What happens if my model is non-linear or let us say it is quadratic? So, when would it be quadratic for example in case, I have something with heat addition. So, you could have instead of a linear model, you could have a quadratic model there what happens in that case? What happens if it is cubic, quartet etc.?

What happens in general if it is non-linear. we were able to find out this formula in case of w0 and w1 what if it is non-linear and it is very messy then what do we do? The next question is instead of one variable because I was doing in only one dimension what happens if there are multiple variables? so what we do for such cases? Final question that we have is what is the accuracy of this model or of this process?

How do we know is this the best that we can do? can we do something else maybe we fit a line here, just because physics said it was a line what if we fit a slightly different model maybe physics is something else. So, all these questions we will answer in the future videos within this week. So, in this video all we saw was 2 simple expressions for w0 and w1 usually these equations are written a slightly different form which I will discuss in the next video.

But more importantly in the next video we will also start with the quadratic model. we will also look at how you can generalize the same process, if you have multiple variables so thank you that is it for this video.