In this video, we are going to learn about a new machine learning algorithm, which is called as Gradient Boosting Machine Learning algorithm. And with the help of this particular algorithm, we can solve both regression and classification use cases. Now, gradient boosting is a part of a boosting ensemble technique wherein we create decision trees sequentially when we combine all the weak learners together. And finally we basically get a strong learner. Already we have discussed about AdaBoost, and we have seen the internal working of how the decision tree is basically getting constructed in AdaBoost. So to start with, we'll take up this regression problem statement and we'll try to construct a gradient boosting decision trees on top of this particular data set. Now let me talk about this particular data set. This data set is a regression data set. Here I have three features. One is experience degree and salary. Experience and degree are my independent features okay this both are my independent features. And this is salary is my output or dependent feature. And here you can see that all my salary output is continuous value. So we are basically considering this as a regression data set okay. Now what are the steps to start with if you really want to create a gradient boosting machine learning algorithm. So let me note down all the steps over here. So this will be my steps. The first step will be to create a base model okay. To create a base model. And whenever we write we we create a base model. We have to make sure that this base model is not biased to any value. So it will be giving you some default value. So in order to find out what default value this base model will give you, just go have to go ahead and compute the average of all the salary output. Because this is my truth output y. Right. So when I compute all the salary uh, average of the salary output, you can see 50 plus 70 plus 80 plus 100. You will be getting somewhere around 75 K okay. So this is the first step. So in the first step what we are doing we are basically creating our base model and the output of any base model. If whenever I give any of my data set the output will be 75. Okay. Now let's go towards the second step. Now in the second step, what we have to basically do and how did we compute this base model. In short, we have to find out the average right now in order to find out the average I will just add up 50 plus 70 plus 80. So let me add over here 50 plus 70 plus 80 plus 100 divided by four. So in short I'm actually getting 75. So this is how we have basically constructed our base model which is giving one output that is 75. The next step that we are going to do is that we are going to compute. The residuals. Or errors. Okay. Now in order to compute the residual and error, what I'm actually going to do, let's say from my base model, I'm going to get my output as y hat. Let's say this is my predicted output. And you know what is my predicted output. It is nothing but 75 K, 75 K, 75 K and 75 K. Okay. So what we are going to do in order to find out the residuals and error, all you have to do I have to find out the difference between the salary and my predicted value. That is my true value and my predicted value. So in order to do this, let's consider that I'm going to write this as r one okay r one basically means the residual one. And here I'm just going to subtract both of this. So in short here what I'm actually going to do I'm going to subtract y and y hat okay. So if I subtract 50 -75 k I'm going to get -25 k. Similarly here I'm going to get five k minus five k here five k. And here 25 k okay. So this is my residual okay. Now let's go to the next step okay I have computed the residuals over here or errors okay. That is the difference between the truth value and the predicted value. This predicted value I have got it from the base models over here. Uh, which was the average of all my salary output. Now coming to the third step, which is a super important step we are going to construct. We are going to construct. A decision tree. See in AdaBoost, we were constructing what kind of tree. It was a stump. But here we are going to construct a decision tree. Considering. Inputs. X of I and output as r one or residual r of I. Okay, now in this particular case you can see that my inputs will be experience and degree. And I'm going to consider this r of one as my output feature. And I'm going to construct a decision tree. So if I really want to construct initially I have a base model. This base model will basically be taking 75 next. The next decision tree sequentially that I'm going to construct will be considering the input features as my independent feature and the output feature as R one. Okay. That is what we are going to do. So whatever decision tree we are going to construct over here, this decision tree one. Is going to take all my independent feature and my output feature is RFI. So here I'm going to take x of I comma r of I r of I basically means r of one okay for the first decision tree. So here I'm just going to write r of one. This is my residual one right. Now once we construct this decision tree, and I hope everybody knows how to create a regression decision tree, because we have already discussed that in the decision tree. So I'm not going to repeat that okay. Again. Mean squared error you'll be probably taking. You'll be basically splitting it based on different different features. And again you can use gain over there to find out which feature should be selected first along with the MSE. Uh, there we have also discussed about uh, variance reduction. Right. All those things have been basically discussed. Now this is super, super important till here okay. Now if we construct a decision tree considering x of I and R of R one as my output, obviously with respect to this, I will be getting one more output feature which is called as r two. Let's say when I pass any record to this particular model, it gives me some output. Let's let's say if the first record I'm just considering anything over here, it can be any value. The first record is giving -23 k. The second record probably it is giving minus three. The third record. Let's say it is giving three. And this is basically giving 20 okay. So what why did why did I write R to R two basically means after training right after training the model with independent features and r of one. Whenever I pass any record to this particular decision tree, I am getting this outputs okay. And obviously the output will be different over here. Okay. So again what I did after constructing a decision tree I will again go and compute the residual error okay. So that is what in short I've got the residual error for the once. This decision tree is basically getting trained with my previous input and output feature. Now this step is super, super important. Now we need to find out. Suppose let's say if I give any input what should be my predicted output. Because this is my residual value. This should not be definitely be my predicted output. So in order to compute the predicted output, how should I give it? Now let's say that I'm passing this particular record. And from this particular record I passed it over here. Right. So my first final predicted output will be nothing but predicted output for the first record, let's say here. First of all, the base model will go to the base model. So in the my base model my value is 75. Right. Then when it goes to the next decision tree one for the same record, you know what is the output. It is nothing but -23. So what I'm going to do, I'm just going to subtract -23 with 75. So 75 -23 is nothing but 52. Now here you can see that for the first record my predicted output is 52 okay. And the real output is 50 k. The difference is very very small. Then obviously I can think that okay, my model is performing well. So don't get mistaken by this guys here. The model is basically performing overfitting. Overfitting, right? So we will not calculated or calculate our predicted output like this. Instead, we will be introducing one more new term. So after my new record passes to 75. So let's say this is my new record that passes to 70 through this base model. And the base model output is 75. I am going to multiply with one learning rate parameter. So here alpha basically means it is called as learning rate. And this value ranges between 0 to 1. But let's consider that I'm going to consider alpha is equal to 0.1. So here what happens alpha I'm going to multiply with whatever output of the decision tree one I've actually got okay. Whatever decision tree output I have got along with that. So here you can definitely see that if I probably now multiply it is nothing but 0.1 multiplied by decision tree. That is -23 decision tree one for the same record. Right. For this particular record my output is 20 -23. So here you can definitely see that I'm going to subtract -2.3. And in short I'm actually going to get somewhere around 72.7. Now if I get this specific output obviously this output error is huge right. So let's say I got this particular output over here. Again I'm going to write my updated y hat right after passing through the decision tree one. So let's say I got over here as 72.7 okay. So here I got 72.7. Let's say after this I'm also going to get something like 60. Uh I'm also going to get somewhere around let's say 77.7. Because when I compare with this somewhat a larger value, then I may get some other values like, uh, you know, probably if I'm computing over here, I may get somewhere like 85. Okay. I may get 110 or 105, something like this. Right. So let's say after passing all the records, I'm getting this particular output. And if you really also want to compute, all I'll do is that I'll go ahead and compute the predicted output. Now let's say for the second record how do I compute 75 multiplied by point one. And what is the residual over here. Residual is nothing but minus three. Right over here you can see the residual is minus three. So for the second record this will be minus three. So in short this will be nothing be 75 plus. Um. -0.3. So here you can basically see that it will be 74.7. That basically is my second record. What it will be. It will be nothing but 74.7 right. 74.7 and similarly I will go ahead and compute this. Then after this, what will happen if I really want to construct my next decision tree? Again, I have to go and compute my r three. That is residual r three and residual r three will be giving from the difference of this 50 k and this 72.7. Right. So here I'm actually going to get -22.7 and some other values. So in this particular case I'm going to get minus four by seven. And some values will also be coming over here. Then what I'll do again I will repeat all this particular steps. And I will after computing the residual again I'll construct my next decision tree. Now my next decision tree will have first of all my base model 75 right along with my so. So here you will be able to see that I will be, first of all, having my base model. And obviously my base model is giving me 75. Then I will construct my first decision tree, which is already constructed, right? I already constructed and this will be to its complete depth. That will not be a problem. And after this I will again construct my next decision tree. Now in this next decision tree, what I am actually going to do, I'm going to take my inputs x of five comma. My output will be what in this particular case it will be R3. So when once I take R3 then again I will be getting my. You know, once this model will get trained, then again I will be getting my R4. That is my residual four, which is the output of that specific decision tree. Again this process will keep on continuing. So in final if I probably want to talk about it, uh, my final function of this particular gradient boost, let's say this will be my function that I'm going to basically write, I can definitely write that my function of f of x will be, uh, the base model H zero of x okay plus alpha 1H1 of x. Okay. Then similarly this is my first decision tree one okay then this will be my decision tree two h of x plus alpha three h of x okay. Like this up to alpha n h of x. Okay. So any number of decision tree that I keep on adding sequentially here, alpha one, alpha two, alpha three are all my learning rates okay. Learning rate. And if you say can we assign one simple value like point one to this. Yes. You can definitely apply uh apply to this okay. You can just use one learning rate for all the, all the decision tree that are available. So let's let's rewrite this something into a different notation here. Instead of writing alpha one here I'm going to write alpha zero also okay. So finally I'll be writing alpha zero. So at the end of the day I can write a summation of I is equal to zero to n. Here I'm going to write alpha of I h of I x. So this is the entire notation of my function f of x for gradient boost. Super super important. And this is how we construct all the decision tree sequentially. And we will be able to solve any kind of problem. Let it be with respect to a regression or classification in the case of gradient boosting. So this is the final function of the. Final function. Of gradient boosting. Again, guys don't get confused like how the decision tree will basically get constructed all the same steps that we have probably applied in the machine. Uh, in, in probably the, uh, decision tree with respect to the regression, right over here, we do multiple splits. We finally get a leaf node. And based on that, we can also do this in top of it. Here also we can apply pre-pruning. Uh, we can play with multiple parameters that are probably I'll show you in the help of coding. Okay. So yes, uh, this was it about gradient boosting regression? I hope you have understood it. Please just go through the video once again in order to understand it properly, because this will be super important when we are doing the practical implementation with Python and Scala. So yes, I will see you all in the next video. Thank you.