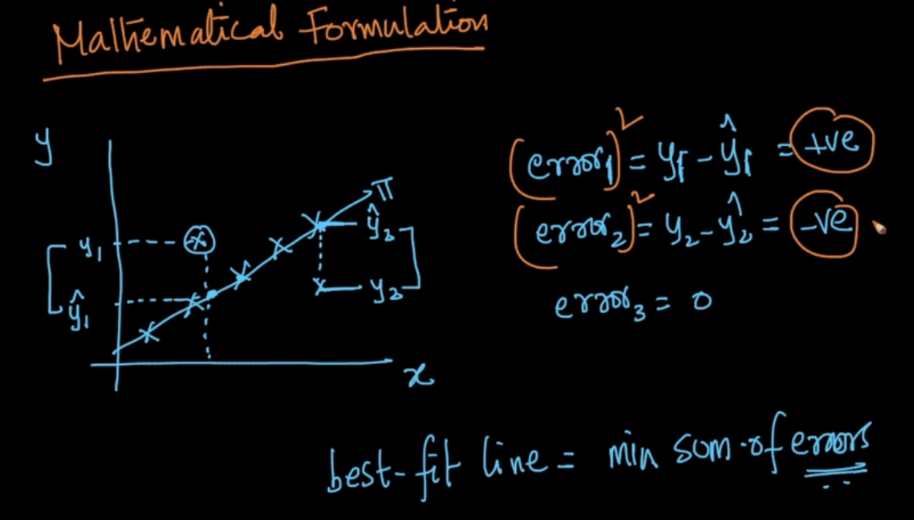
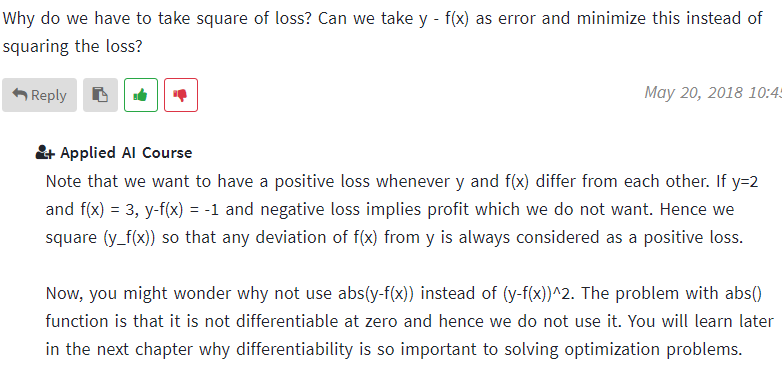
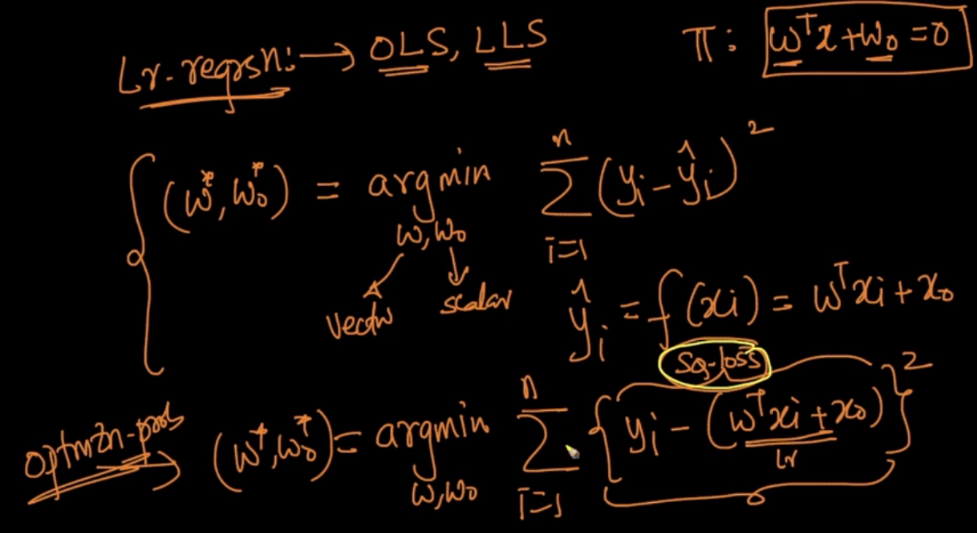
As we know that we want to min sum of errors, so here we are are squaring each error, to make –ve errors to be +ve.





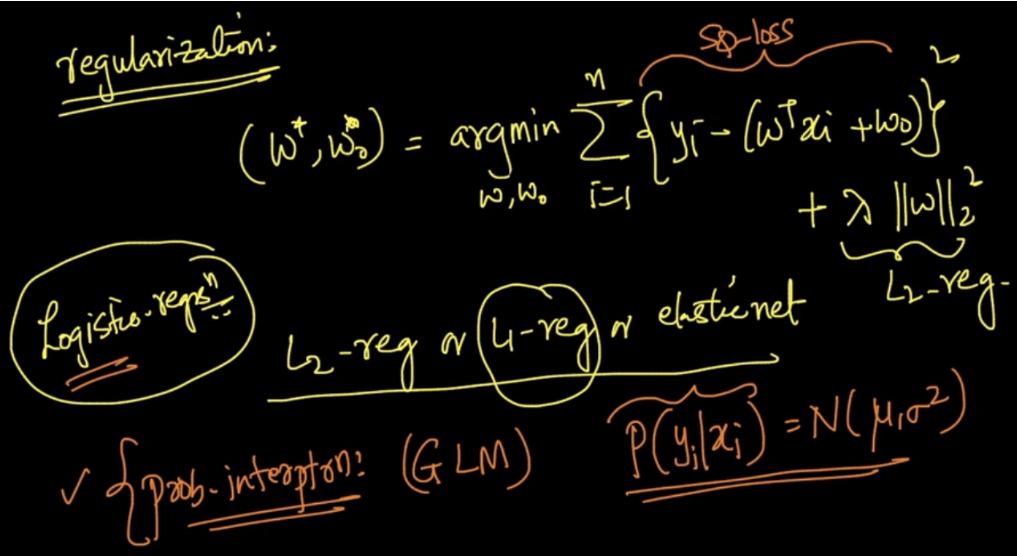
Linear Reg. is also known as ordinary least square and linear least square, because of this square loss function.

Our objective function is given below where we are finding value of w and w0 such that square loss function becomes minimum.

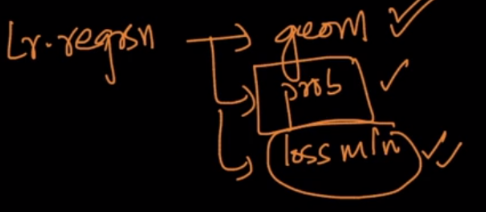


Here we add regularization to our objective.

Why do we add regularization? Explanation is given in end of notes.

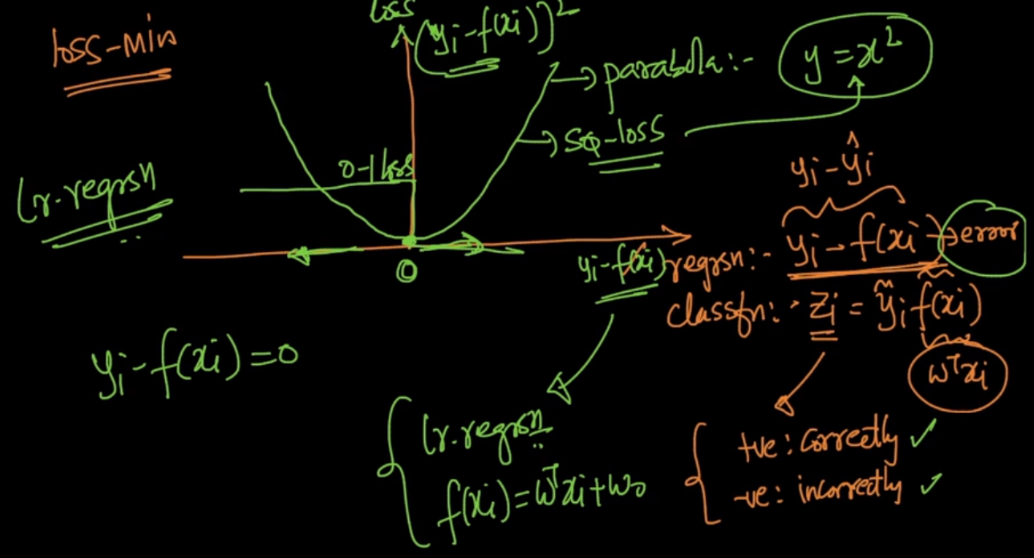


Similar to Logistic Reg, Linear Reg can also be represented in 3 forms, given in below image.



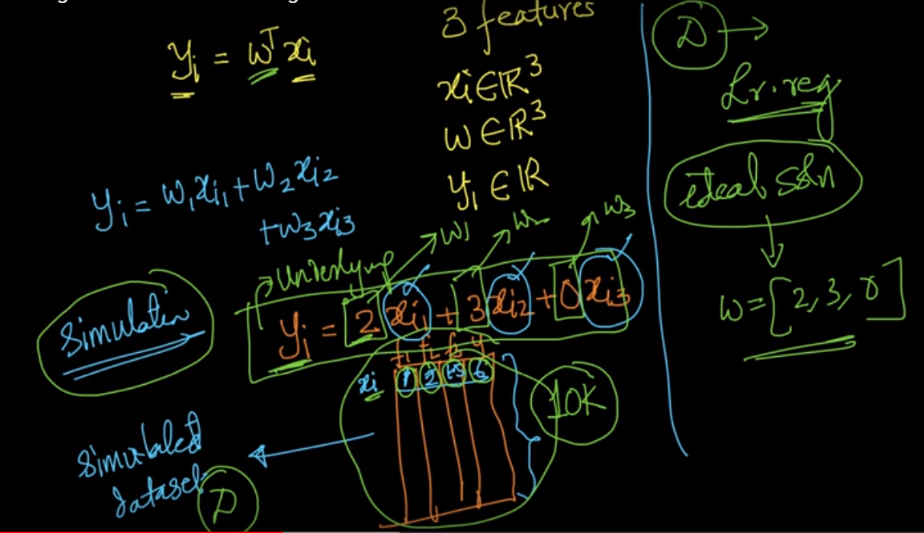
Loss minimum representation of Linear regression is given below,

Where our objective is min square loss function therefore it would be on y and this would occur when error will be 0.



**Why we are adding Regularization term?**

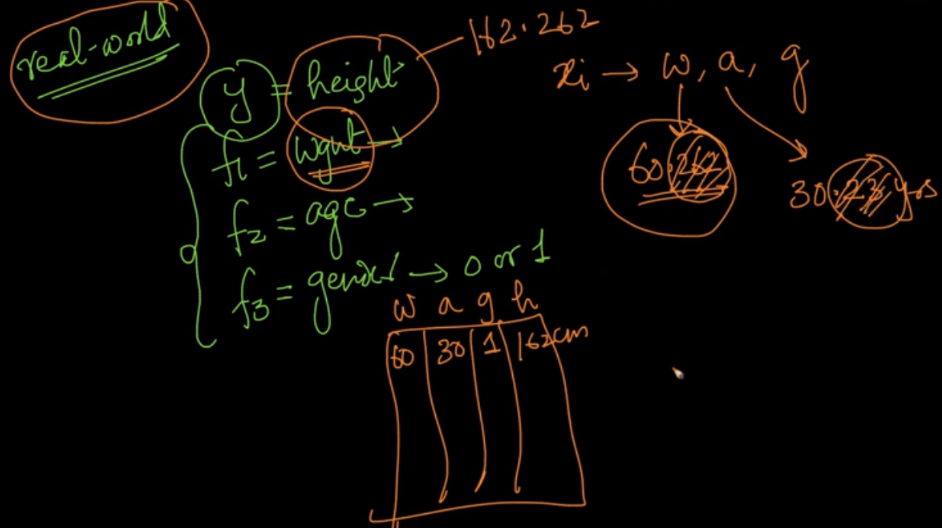
Let suppose with given equation we will simulate dataset. So now we know w already and using this w we are simulating dataset.



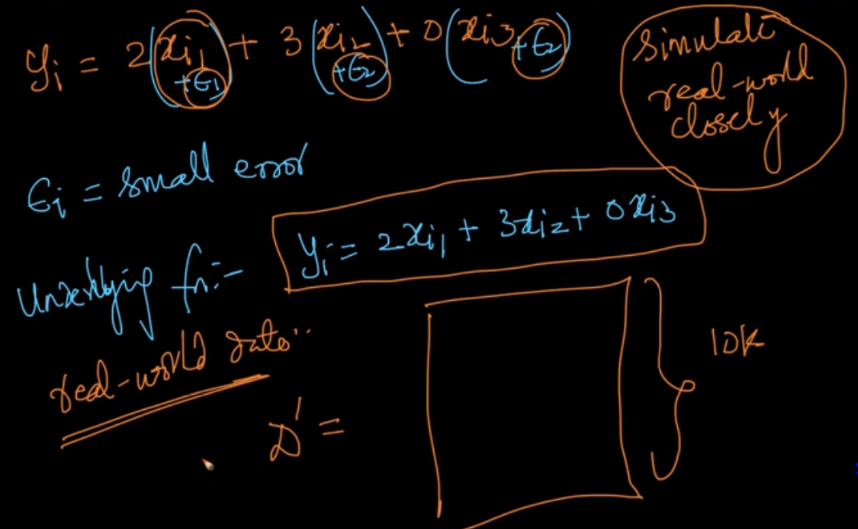
But in real world, error is always there with dataset

For example if ht is 162.23, it would be noted as 162, weight is 45.67 it would be noted as 45.

So there exist error with each features in real world.

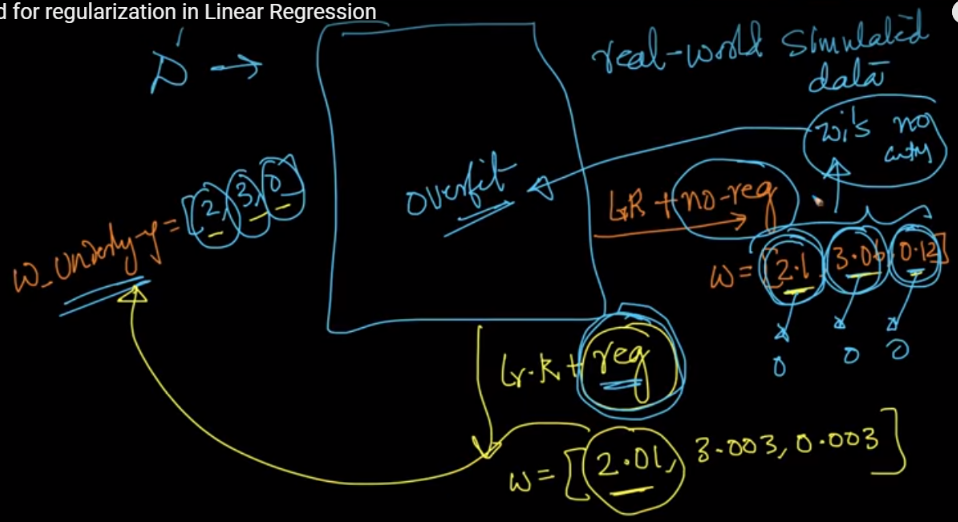


So to simulate real world kind of data we’ve added error to value of each feature.



Now when we try to apply linear reg we will get w = [2.1, 3.06, 0.12], but actually w was [2,3,0].

So why this is coming, because of error present in value of features, As linear reg tries to best fit the line means minimize the error, even for outliers we would get higher w values than actual w values.

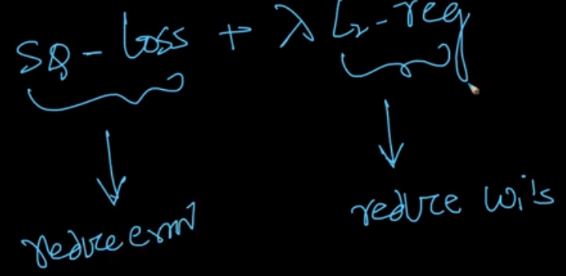
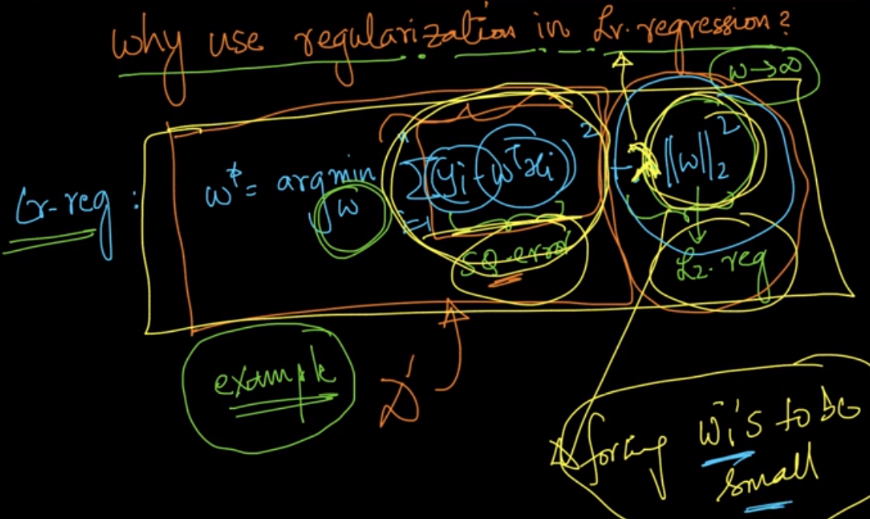


So in order to decrease obtained w to bring it near to actual w, we will add regularization term.

So our objective function will try to minimize this also.

So will w comes to 0?

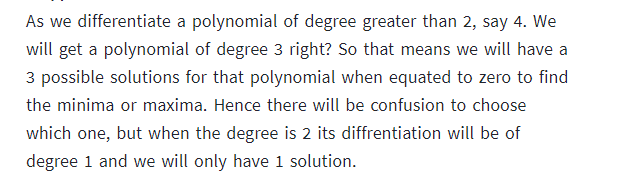
No because if w comes to 0, square error become very high and this value is not minimum which our function is finding, so eventually we would get a value of w which will neither be too high and neither be too low.



**Comments:**

2)





3)

