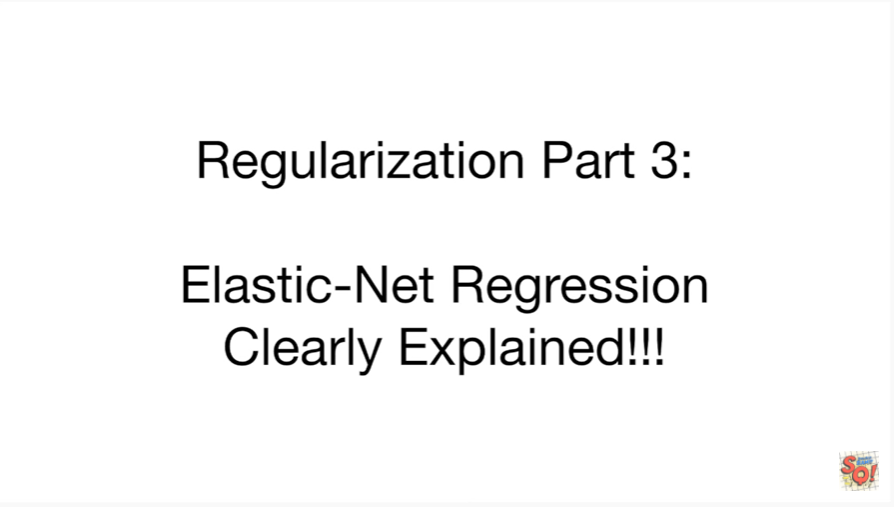
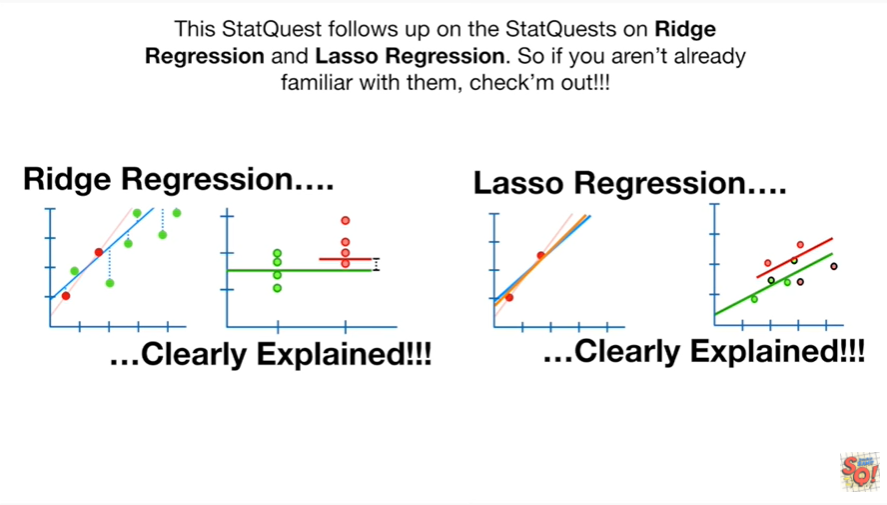
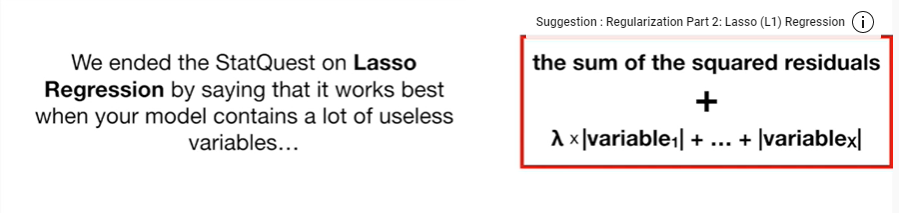
<https://www.youtube.com/watch?v=1dKRdX9bfIo&list=PLblh5JKOoLUICTaGLRoHQDuF_7q2GfuJF&index=22>



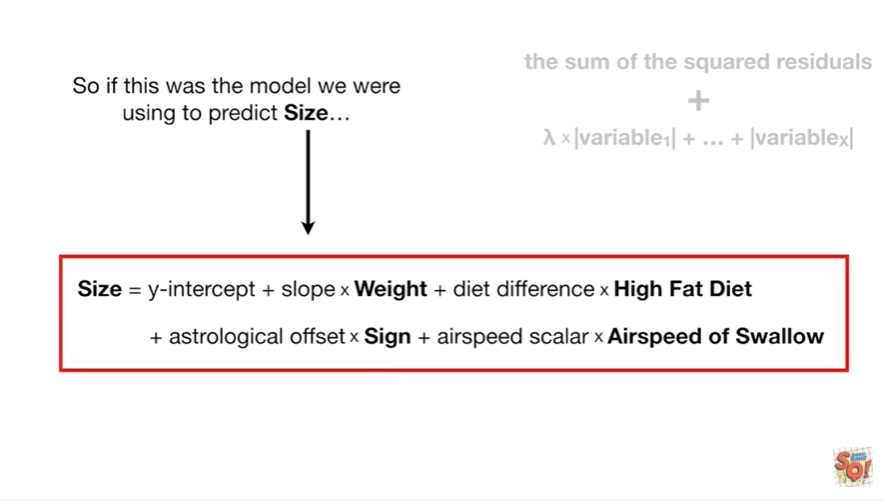
Today we're gonna do part 3 of our series on regularization.

We're gonna cover elastic net regression and it's going to be clearly explained.

This stat quest follows up on the stat quests on Ridge regression and lasso regression so if you aren't already familiar with them check them out.



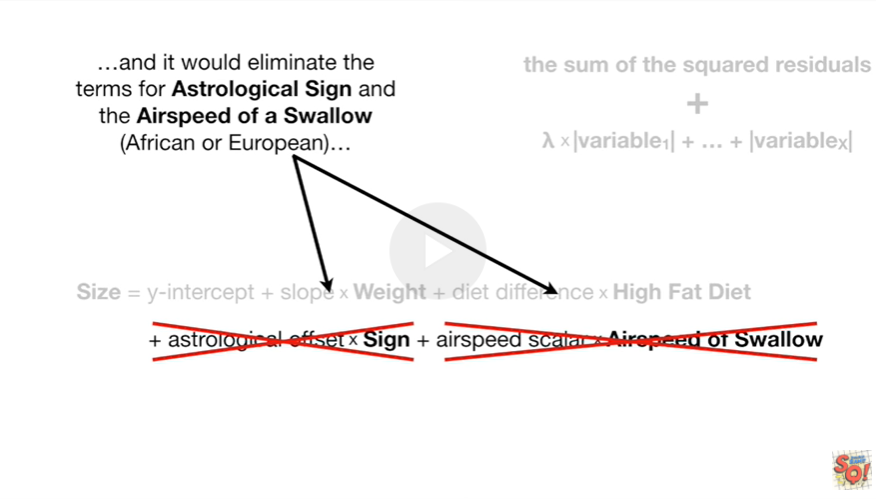
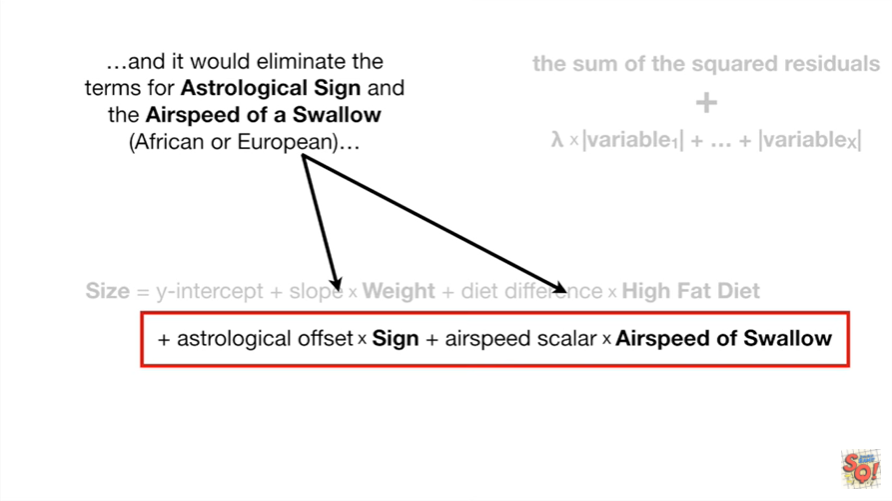
We ended the stat quest on lasso regression by saying that it works best when your model contains a lot of useless variables.



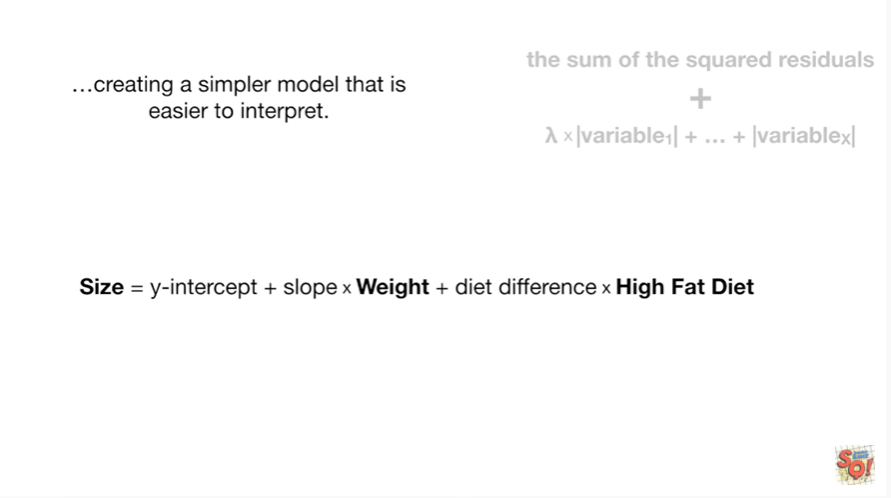
So if this was the model we were using to predict size



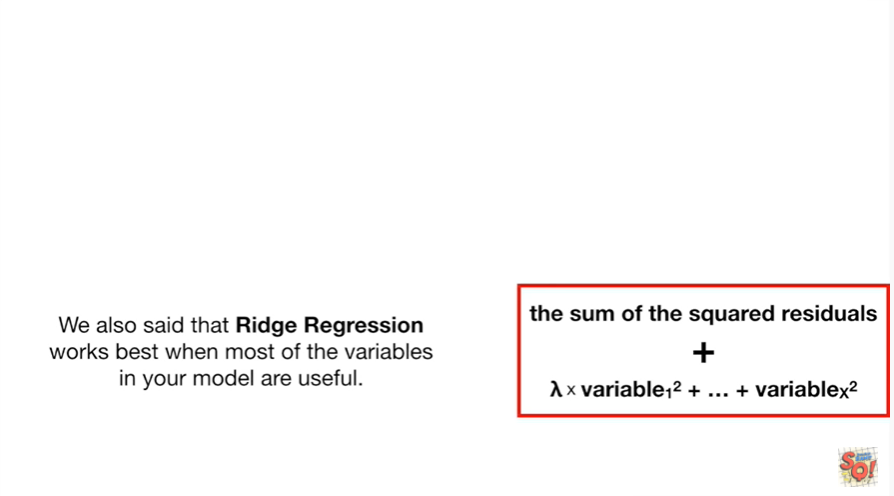
then lasso regression would keep the terms for weight and high fat diet



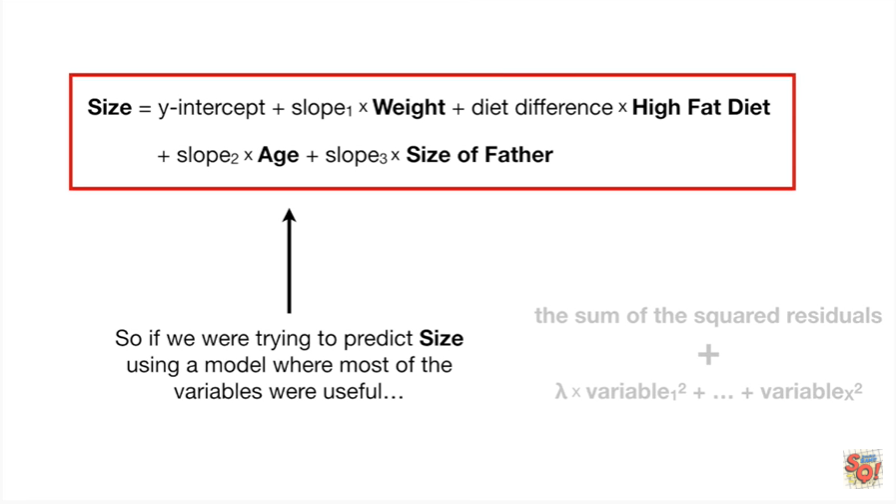
and it would eliminate the terms for astrological sign and the airspeed of a swallow African or European



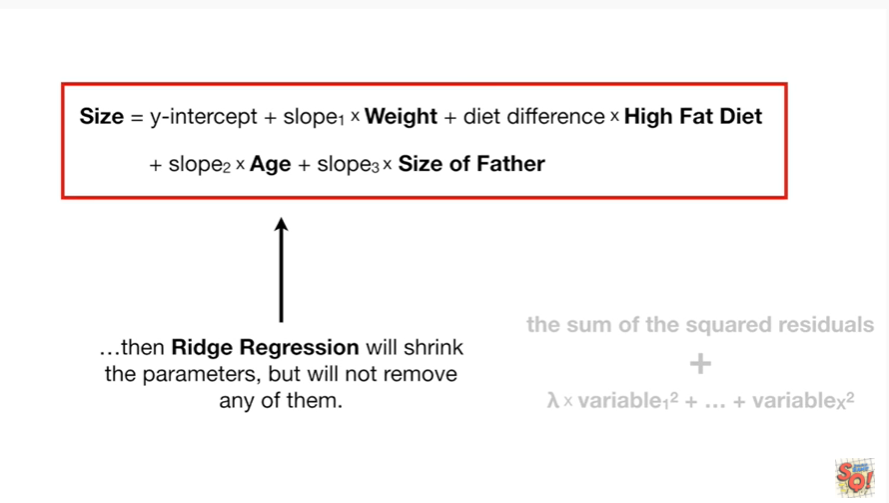
creating a simpler model that is easier to interpret.



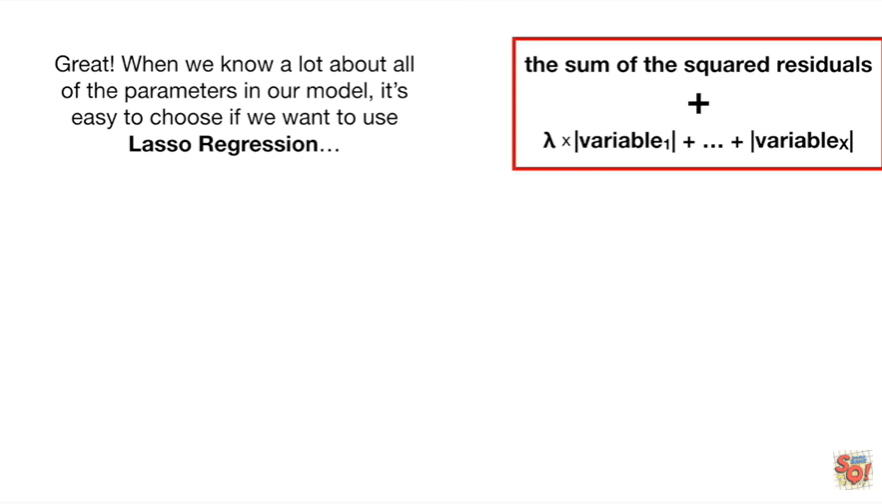
We also said that Ridge regression works best when most of the variables in your model are useful.



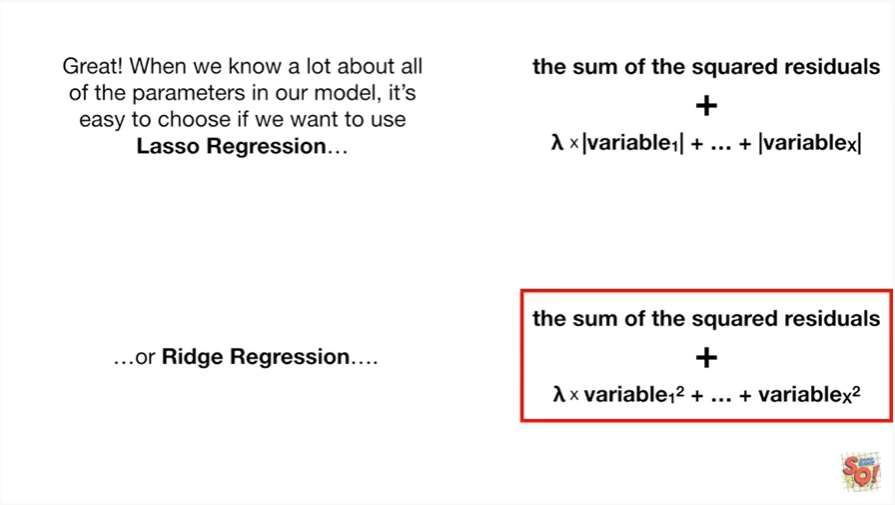
So if we were trying to predict size using a model where most of the variables were useful



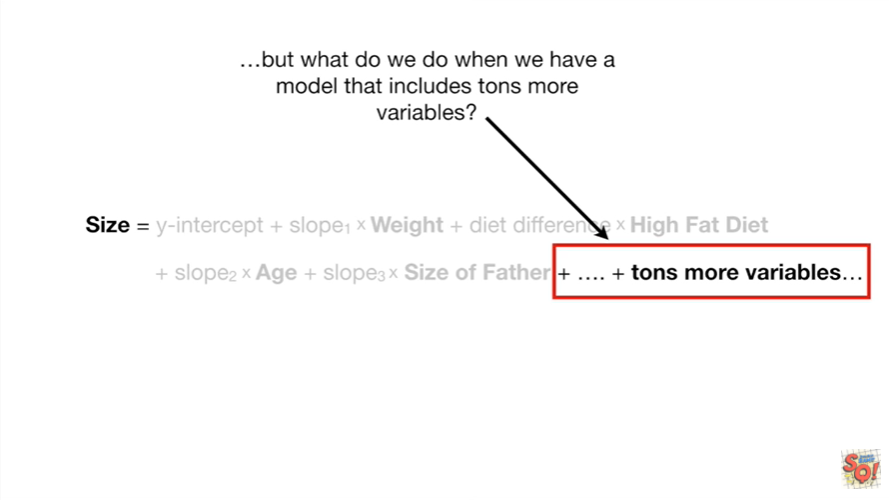
then Ridge regression will shrink the parameters but will not remove any of them.



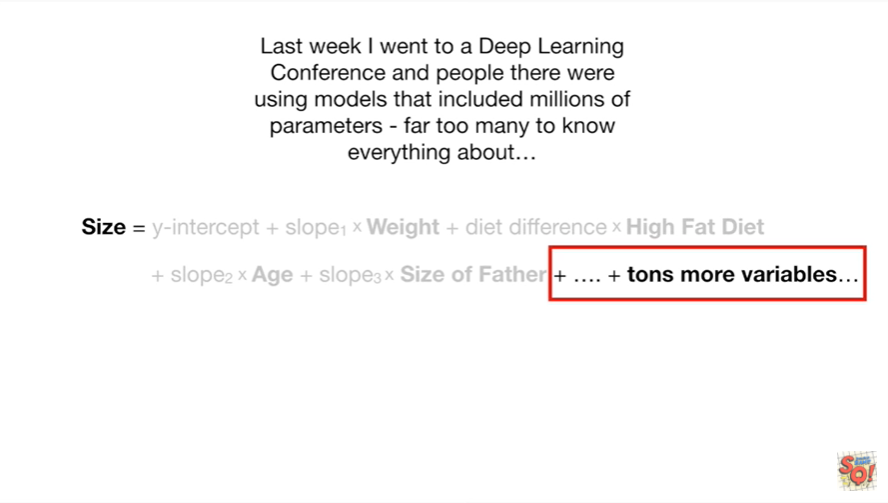
Great when we know a lot about all of the parameters in our model it's easy to choose if we want to use lasso regression



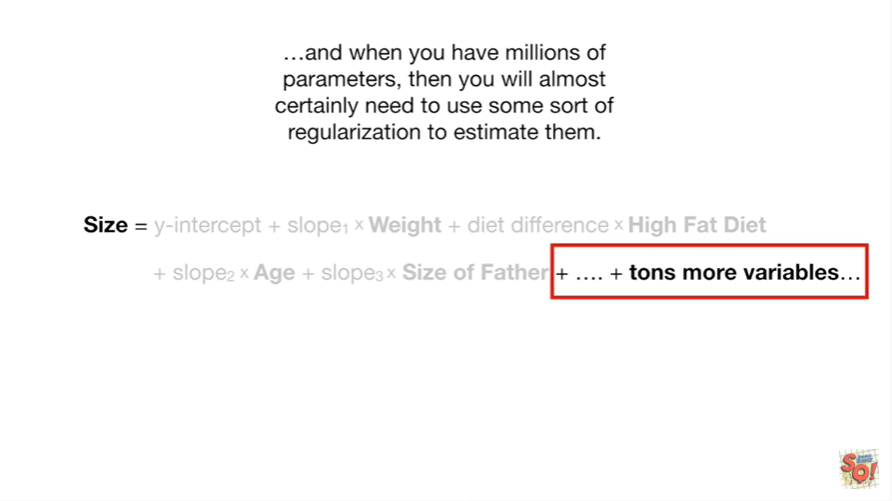
or Ridge regression



but what do we do when we have a model that includes tons more variables.



Last week I went to a deep learning conference and people there were using models that included millions of parameters far too many to know everything about



and when you have millions of parameters then you will almost certainly need to use some sort of regularization to estimate them.



However the variables in those models might be useful or useless, we don't know in advance.



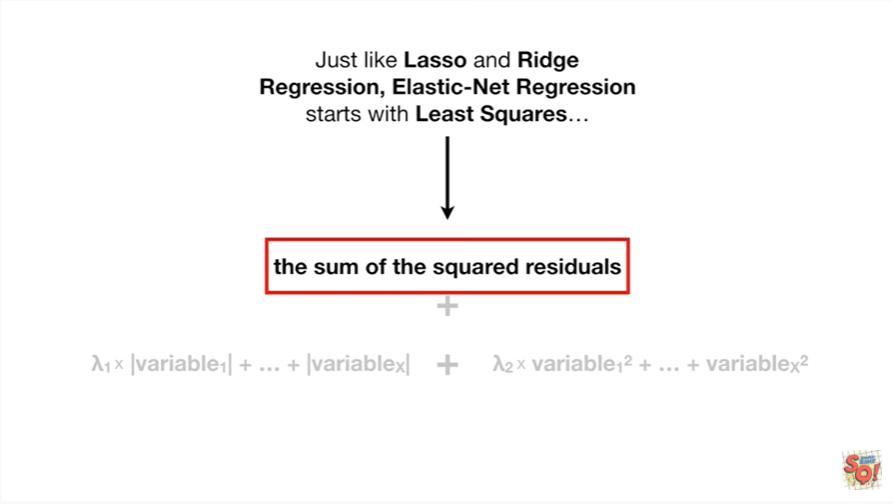
So how do you choose if you should use lasso or Ridge regression ?



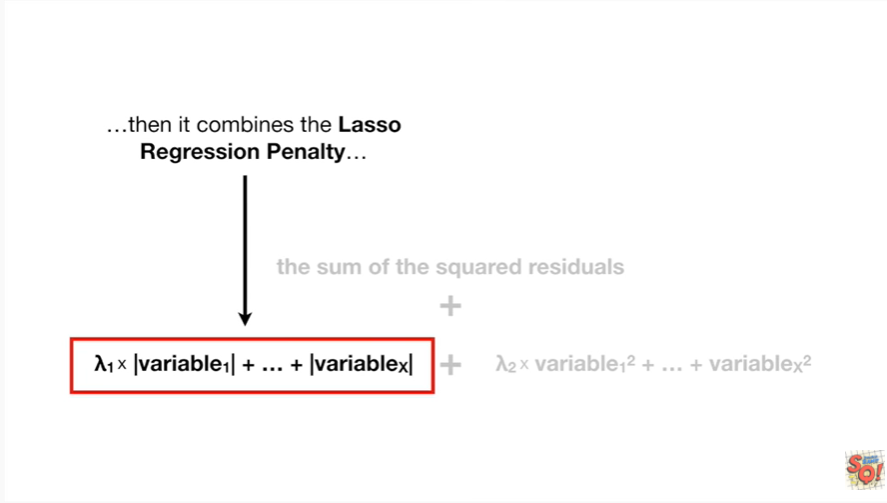
The good news is that you don't have to choose, instead, you use elastic net regression.



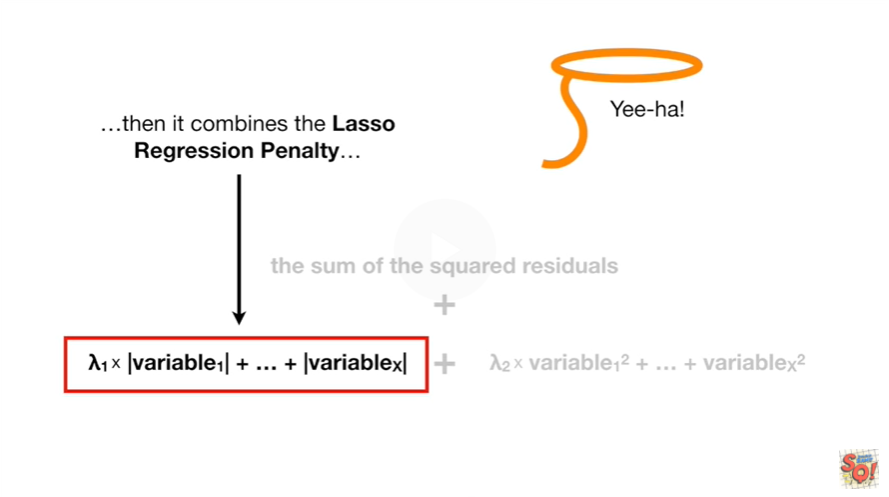
Elastic net regression sounds super fancy but if you already know about lasso and Ridge regression it's super simple.



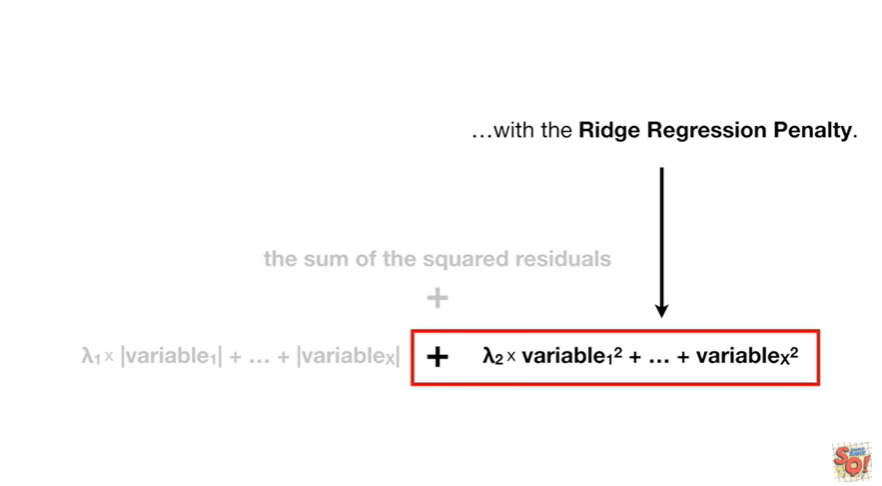
Just like lasso and Ridge regression elastic net regression starts with least squares



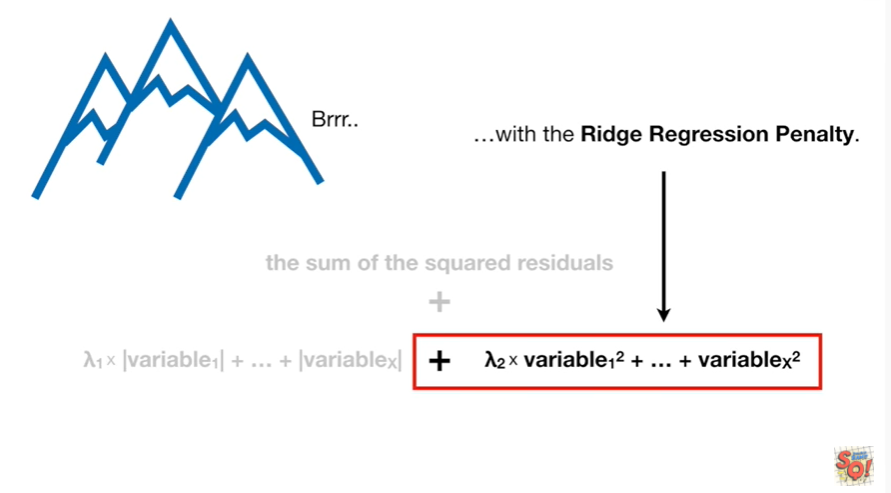
then it combines the lasso regression penalty



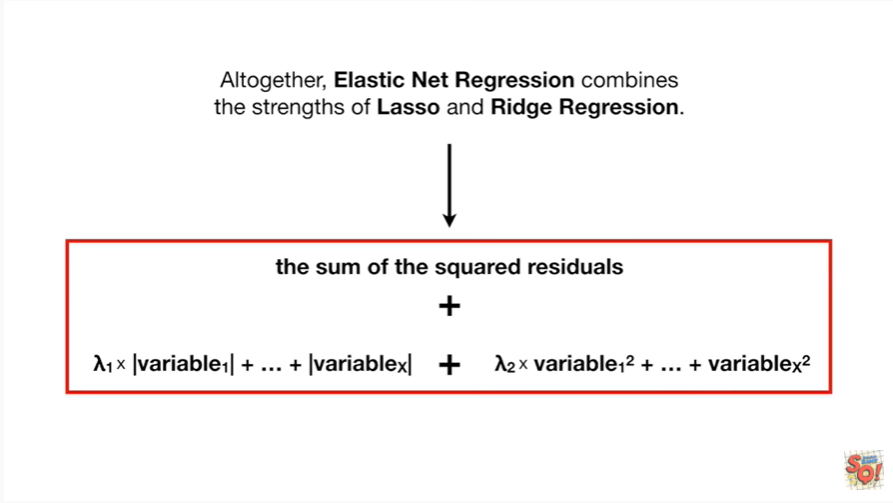
yeehaw



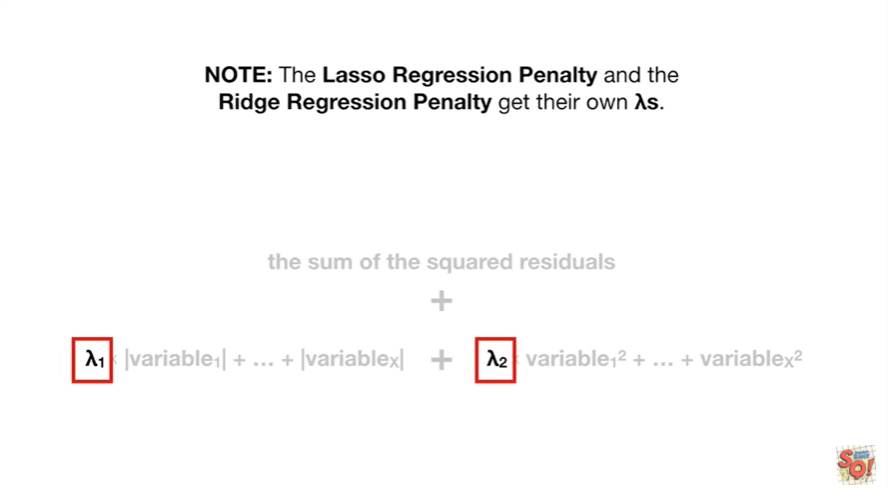
with the ridge regression penalty



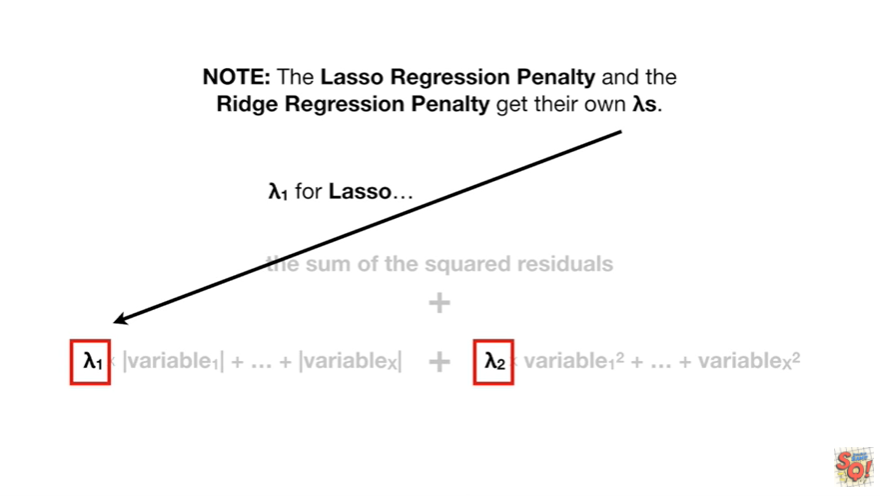
BRR



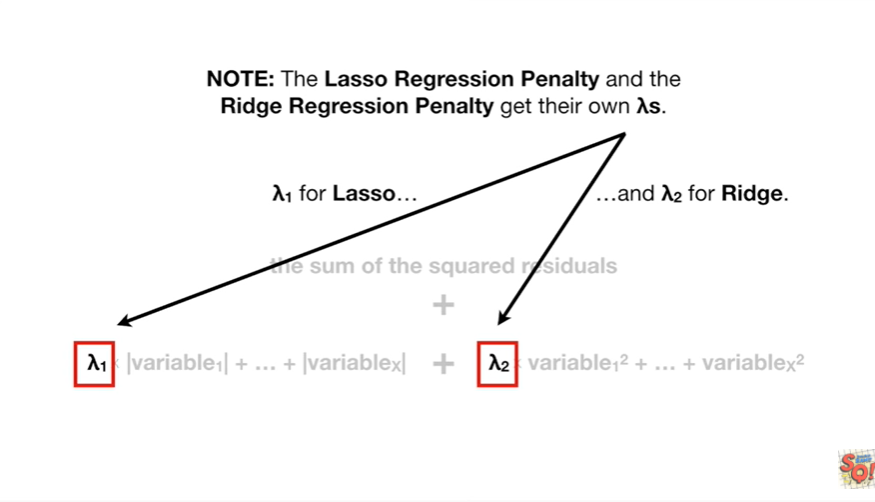
Altogether elastic net combines the strengths of lasso and Ridge regression.



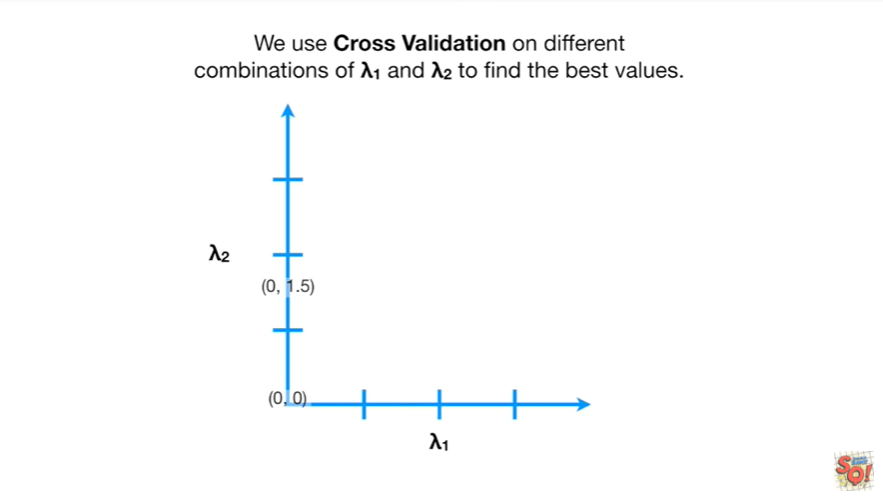
Note : the lasso regression penalty and the ridge regression penalty get their own lambdas.



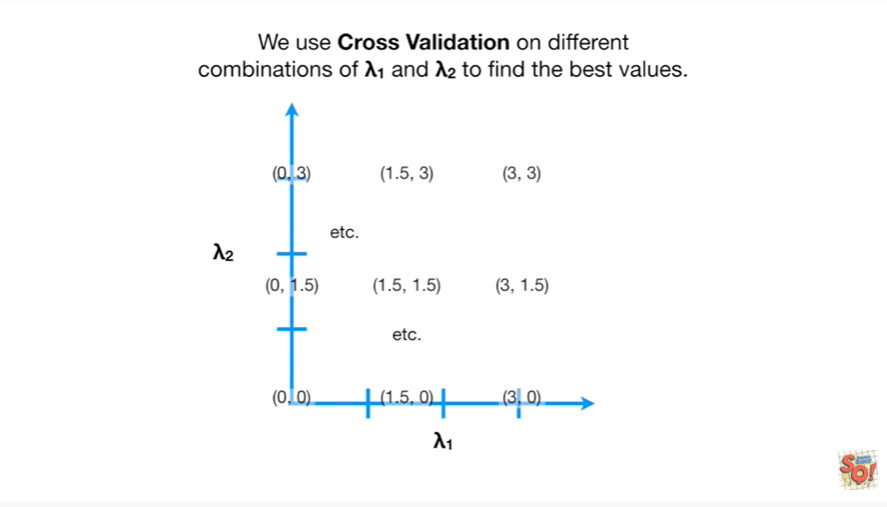
Lambda sub 1 for lasso



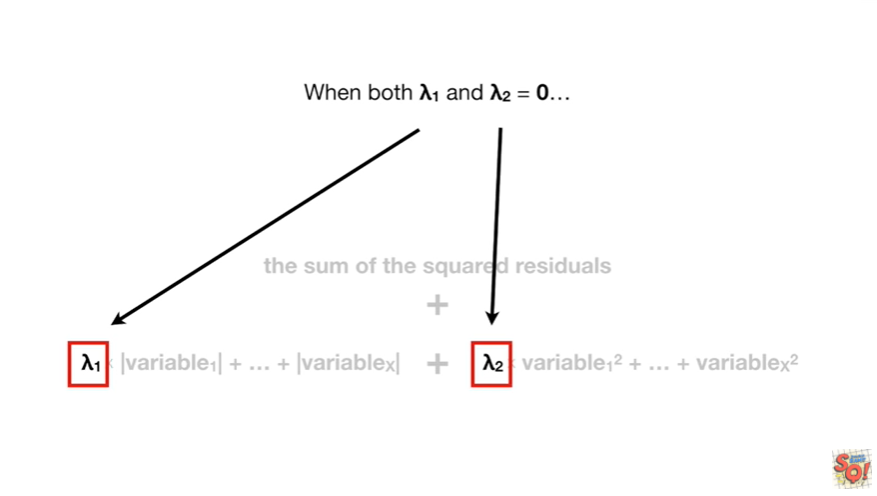
and lambda2 for Ridge.



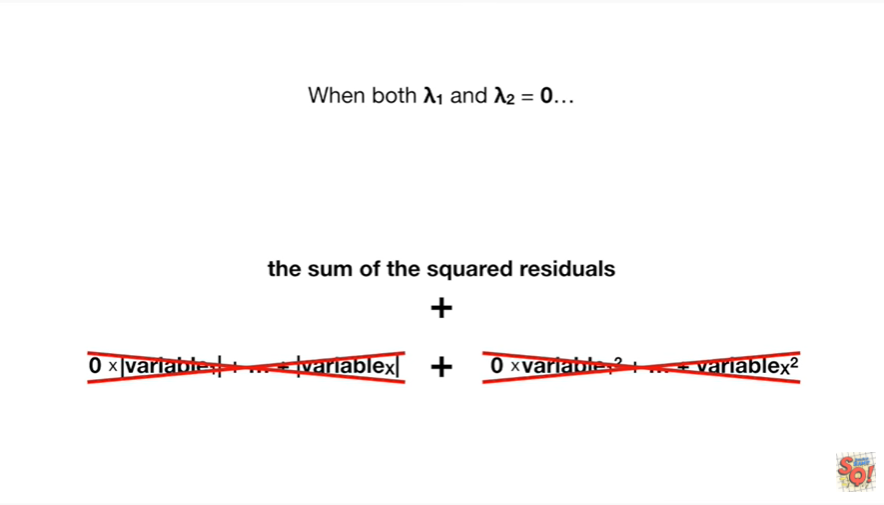
We use cross-validation on different combinations of lambda sub 1 and lambda sub 2



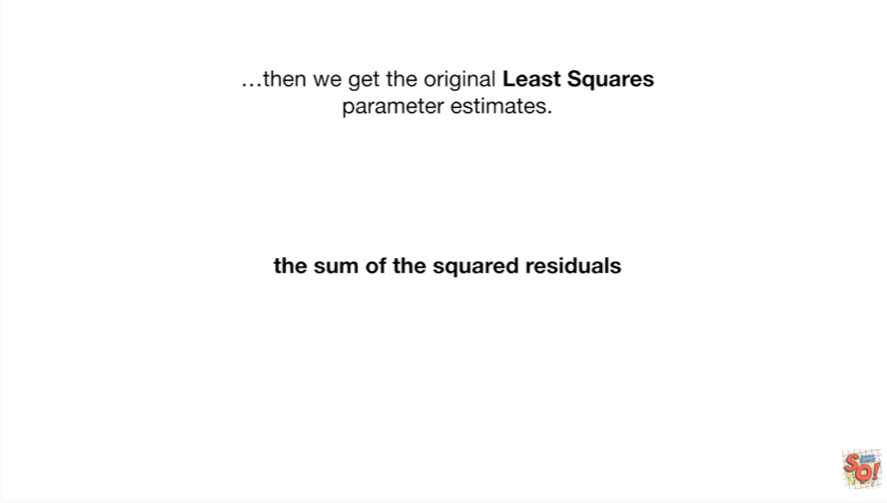
to find the best values.



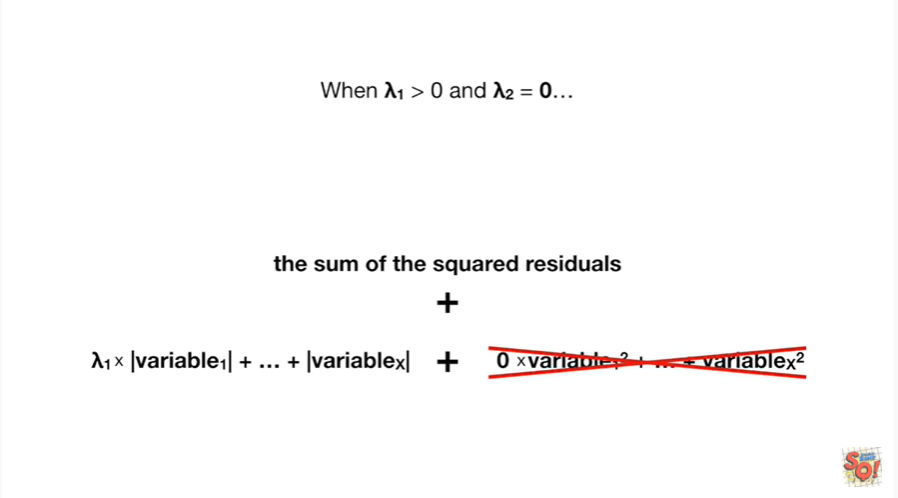
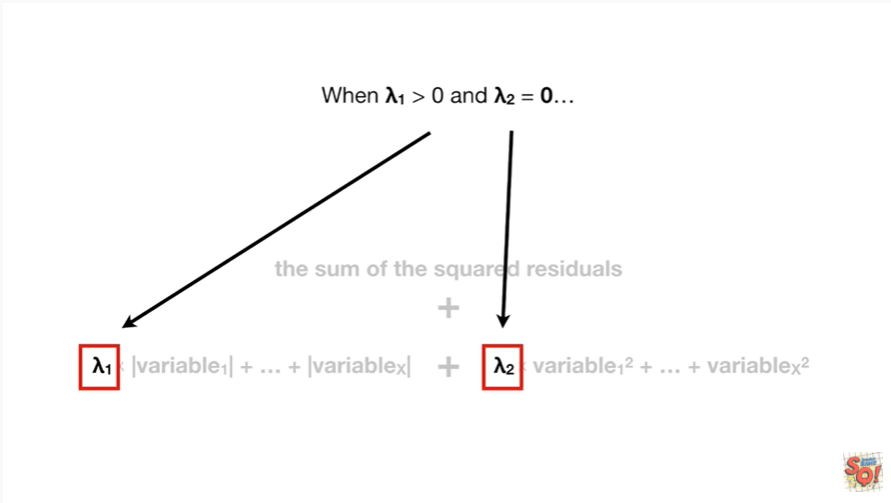
When both lambda sub 1 and lambda sub 2



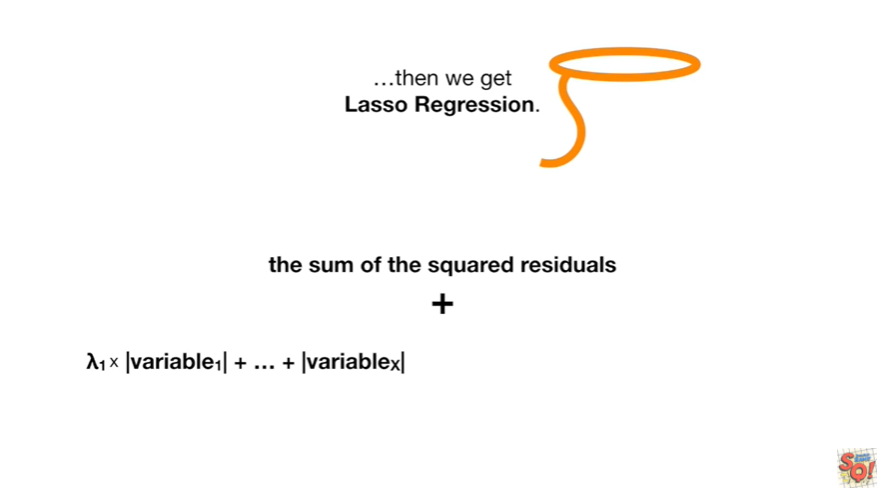
equals 0



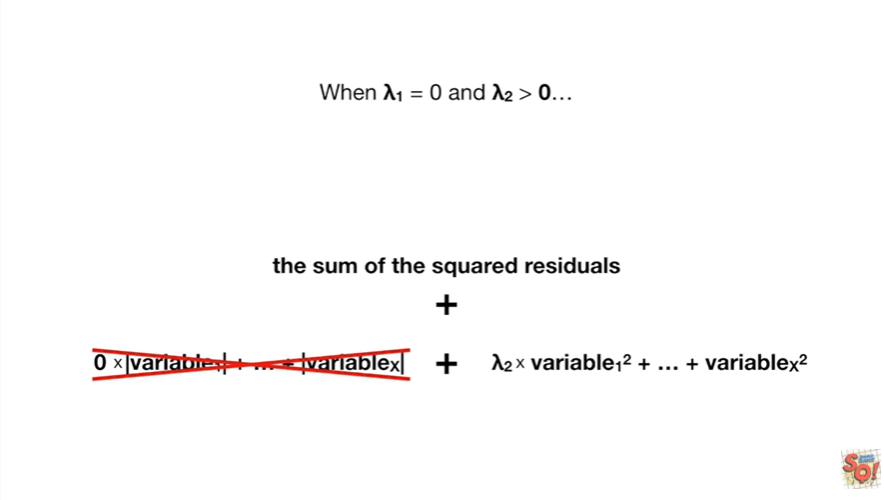
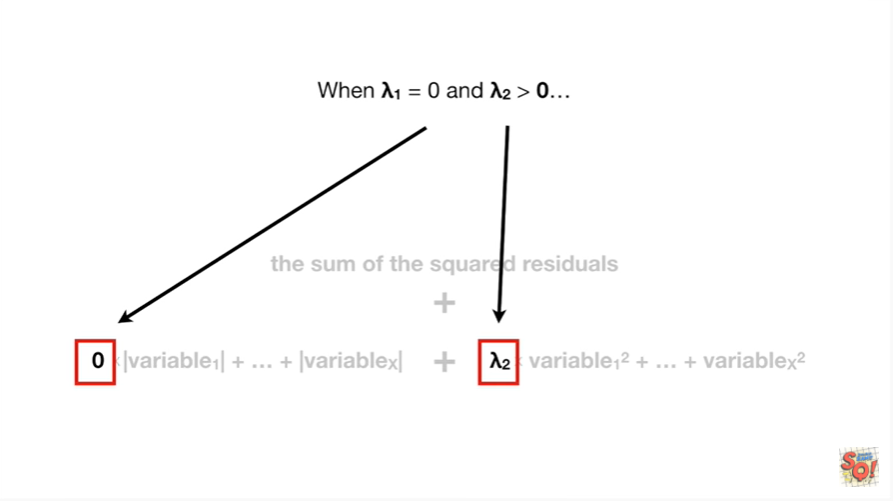
then we get the original least-squares parameter estimates.



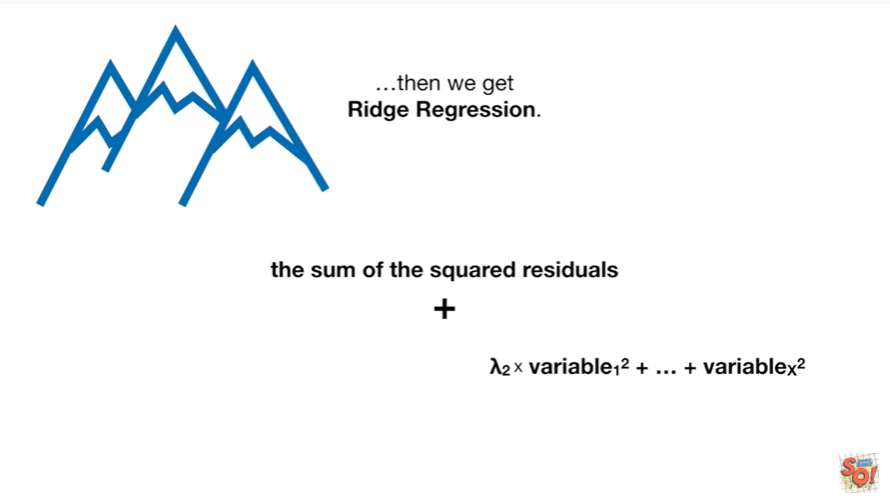
When lambda sub 1 is greater than zero and lambda sub two equals zero



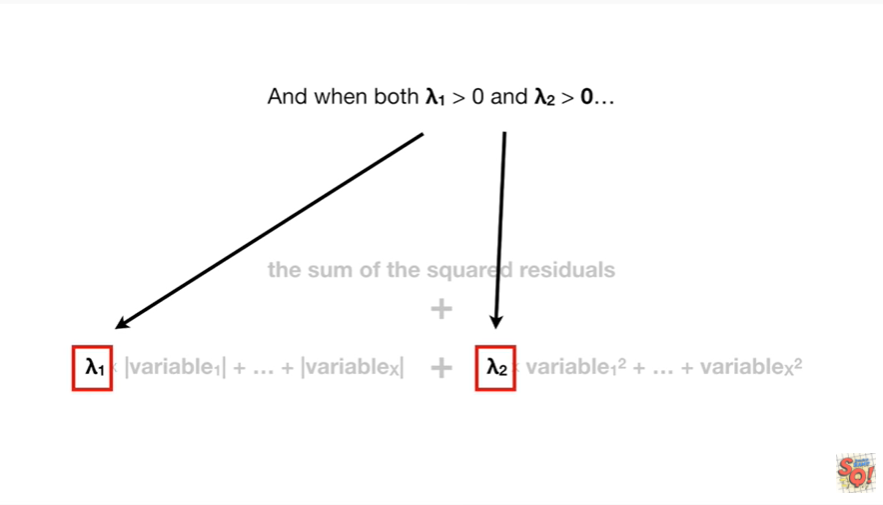
then we get lasso regression.



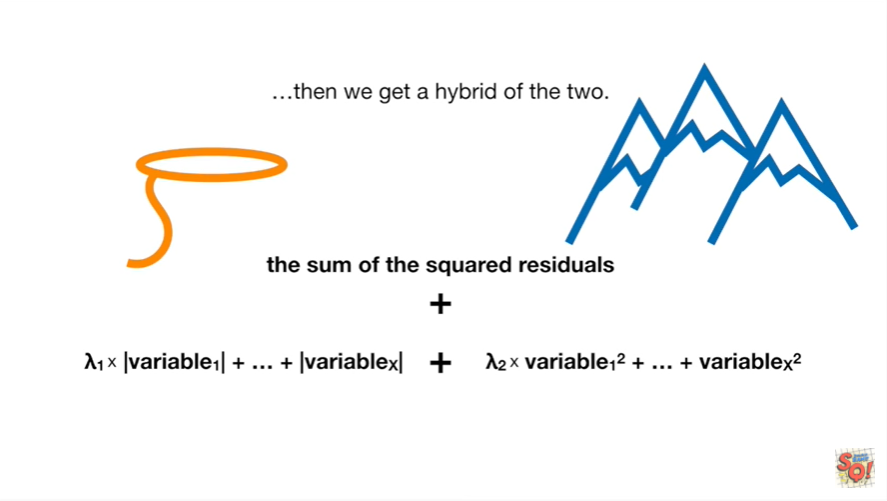
When lambda sub 1 equals zero and lambda sub two is greater than zero



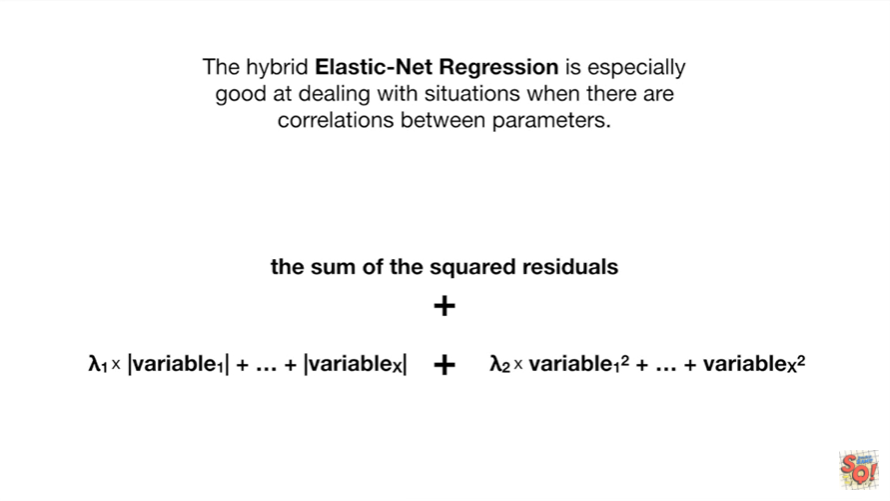
then we get ridge regression.



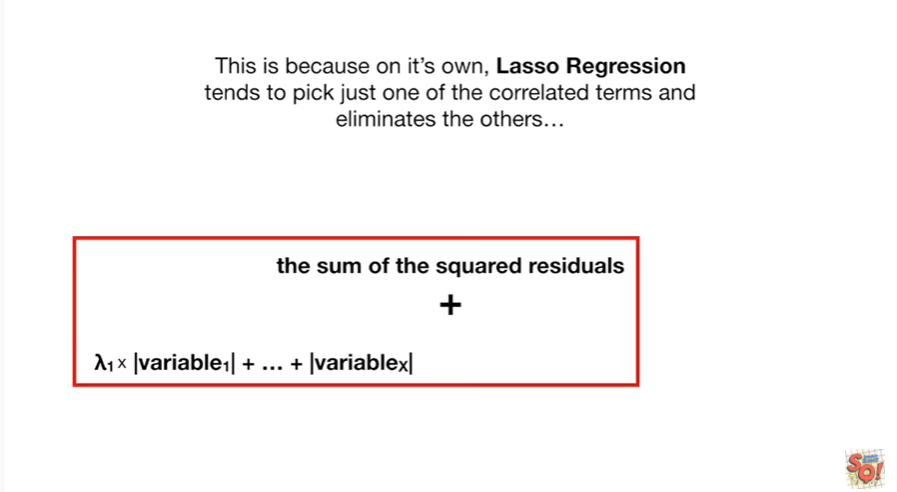
And when both lambda sub 1 is greater than zero and lambda sub two is greater than zero



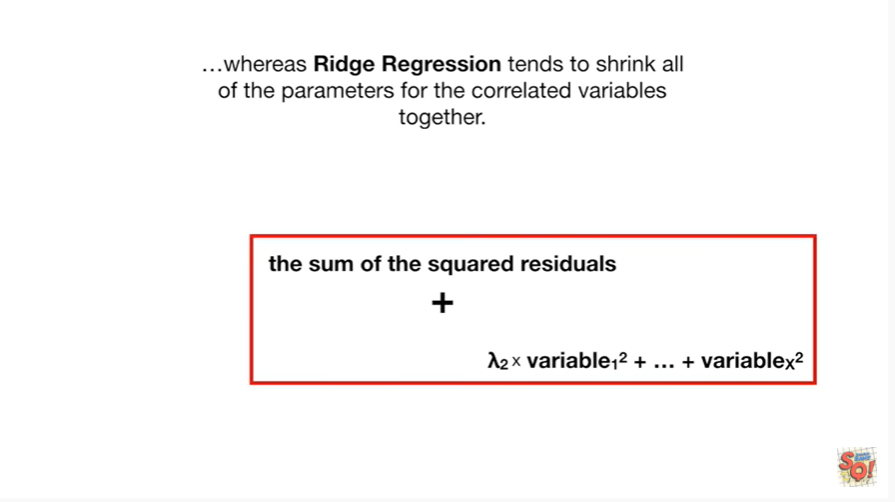
then we get a hybrid of the two.



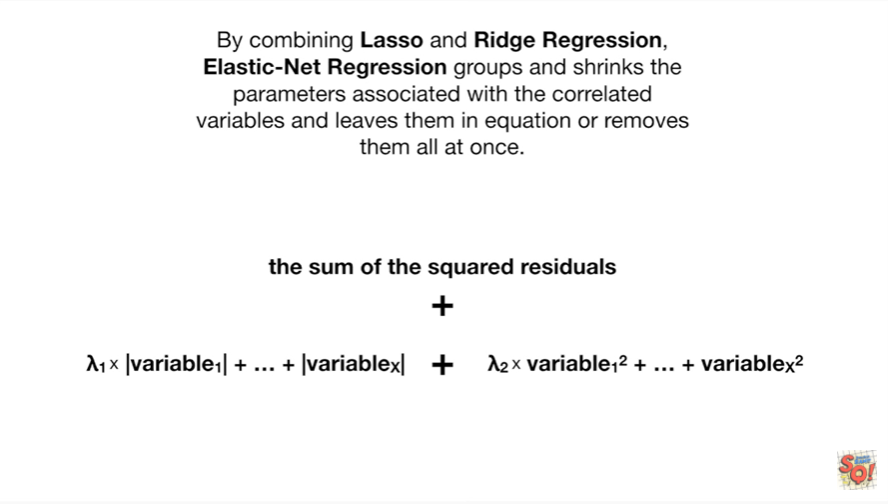
The hybrid elastic net regression is especially good at dealing with situations when there are correlations between parameters.



This is because on its own lasso regression tends to pick just one of the correlated terms and eliminate the others



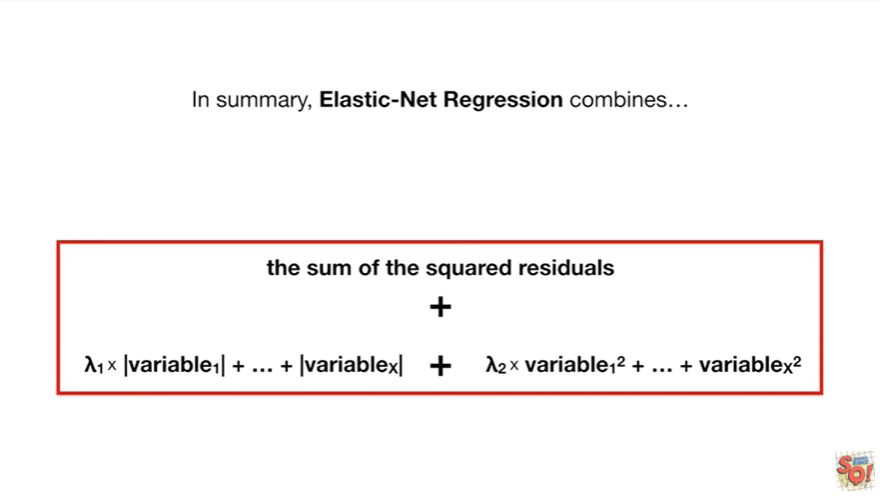
whereas Ridge regression tends to shrink all of the parameters for the correlated variables together.



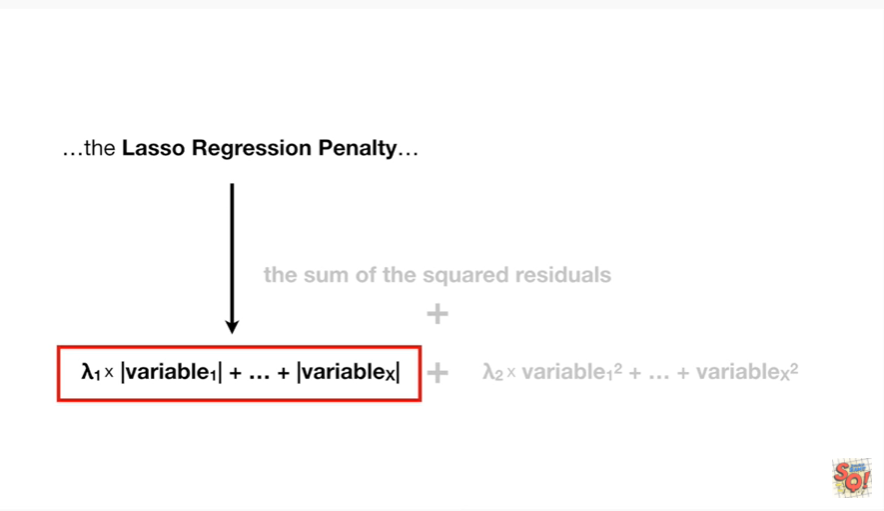
By combining lasso and Ridge regression elastic net regression groups and shrinks the parameters associated with the correlated variables and leaves them in the equation or removes them all at once.



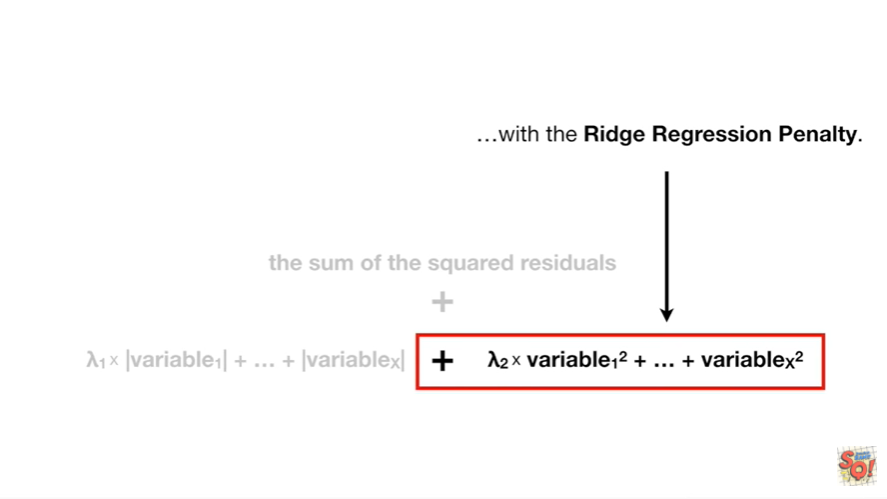
BAM !!!



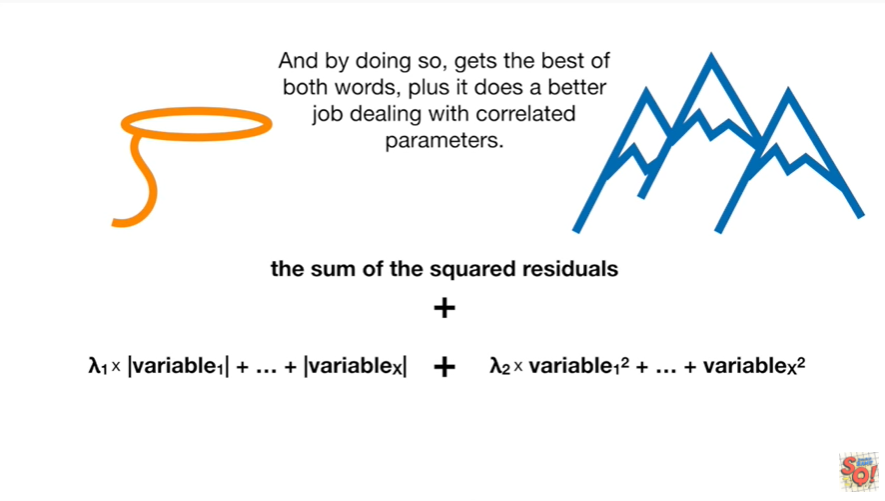
In summary, elastic net regression combines



the lasso regression penalty yeehaw



with the ridge regression penalty brr.



And by doing so gets the best of both worlds plus it does a better job dealing with correlated parameters.