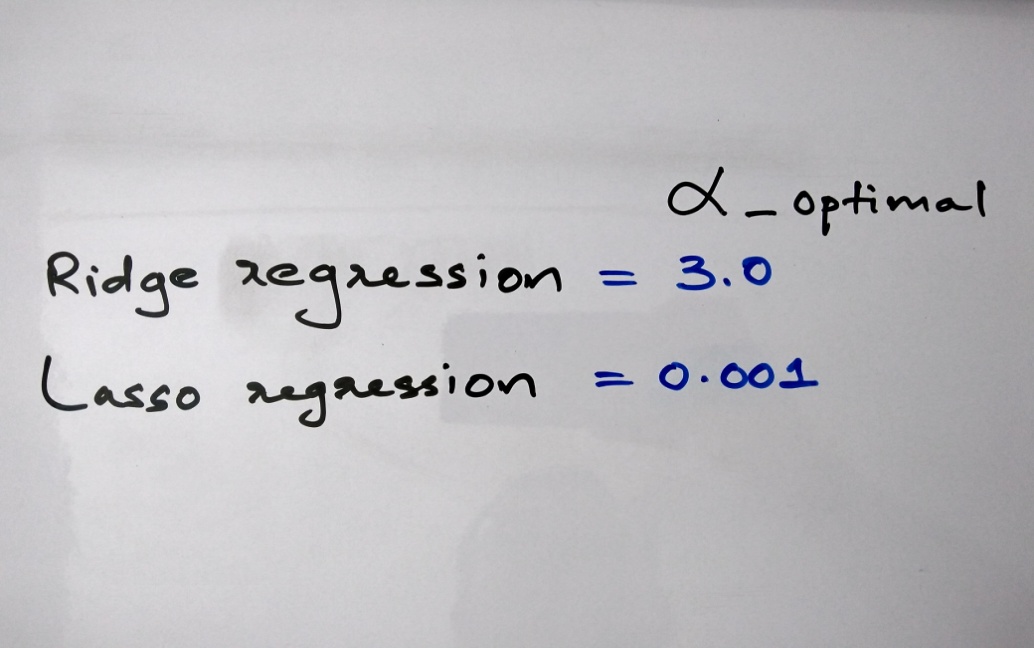
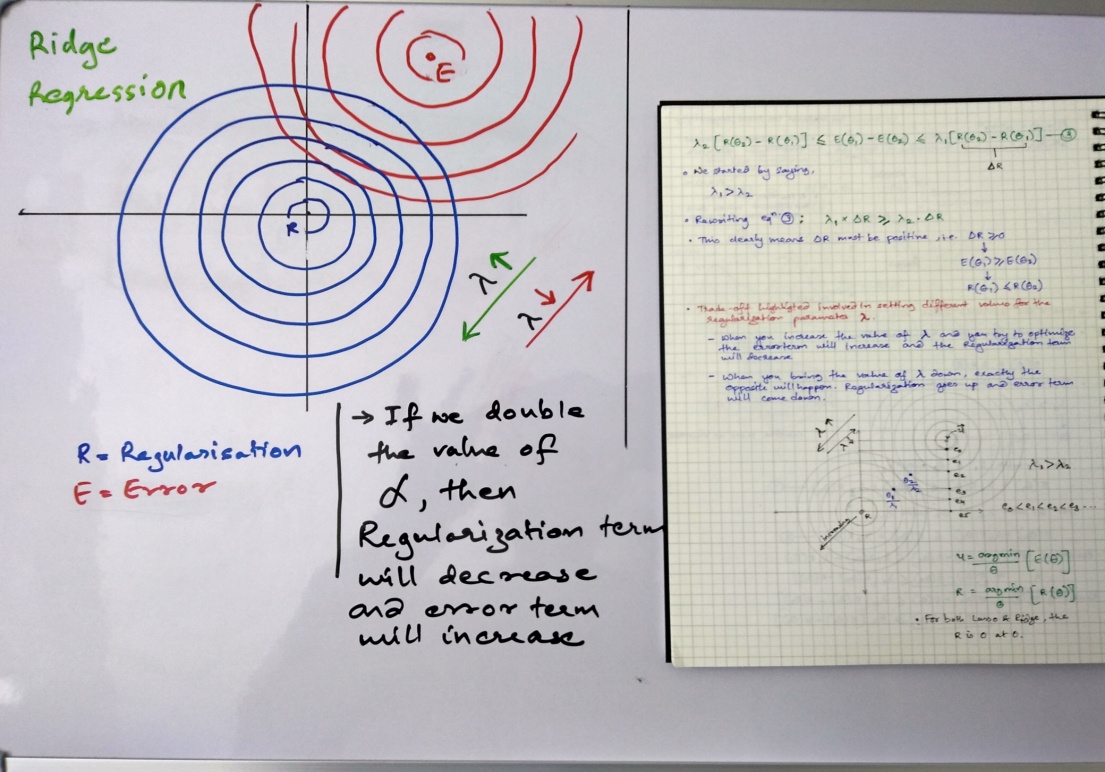
1. **What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?**





|  |
| --- |
| IMG_20200720_224142506_HDR.jpgIMG_20200720_224148773_HDR.jpg  -> After doubling the alpha value for Ridge :    -> After doubling the alpha value for Lasso: |

1. **You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?**

|  |
| --- |
| As the Business goal is to get important features that affect the House price ,and the model is supposed to handle the variation, lot of features with smaller coefficient would increase in bias. Hence I would go for Lasso with 0.001 as alpha value as it pulls the less information value variables to 0. Making model dependent upon a small set of features. This is a generalized model.  Also after doubling the alpha for lasso would make it more generalized, but then it might happen that we may loose some important variable. |

1. **After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?**

-> Now after dropping the top5 variables, the five most important predictor variables are: (some are derived features)

* 1stFlrSF
* 2ndFlrSF
* TotalBsmtSF
* GarageArea
* Functional

1. **How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?**

To make model robust and generalisable by:

* creating derived features
* handling outliers
* transformation of features
* selection of features based on business understanding
* Selection of minimalistic variables

To make model more generalisable, it is important to get a good balance between bias-variance trade off. If the model is over-fit , It will definitely increase the accuracy and will perform good on the data it is trained on.

But, In case scenarios such as Surprise Housing, A lot of features might go missing in the input dataset, Which might lead to a high variance for a good accuracy scoring model.

Hence A model build with generalisable approach may lead to poor accuracy but it is more powerful in handling variance.