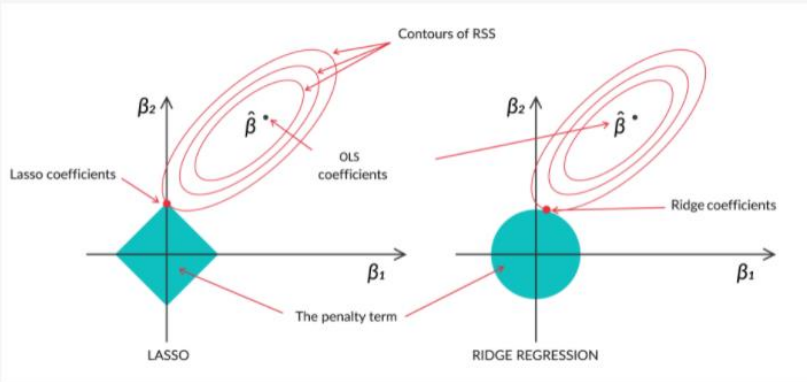
Question 1:

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?

Ans



As per the graphical representation of lasso and ridge regression the optimal value lies where the contour touches the diamond and circle shape. For our case study, Optimal values of Alpha for ridge and lasso regression are 100 and 0.001.

There is change in the model metrics after doubling the alpha values are below.

Before Alpha:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.412815e-01	0.922613	0.936884
1	R2 Score (Test)	-3.793819e+22	0.922613	0.872285
2	RSS (Train)	9.423858e+00	12.419997	10.129707
3	RSS (Test)	2.734151e+24	8.807348	9.204230
4	MSE (Train)	9.607303e-02	0.110293	0.099606
5	MSE (Test)	7.900857e+10	0.141803	0.144963

After Alpha Double:

	Metric	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.916262	0.928638
1	R2 Score (Test)	0.916262	0.876044
2	RSS (Train)	13.439316	11.453026
3	RSS (Test)	8.828678	8.933362
4	MSE (Train)	0.114730	0.105913
5	MSE (Test)	0.141975	0.142814

The change for the important predictor variables is following:

Aplha:

Ridge & Lasso:

	Linear	Ridge	Lasso
OverallQual	0.070679	0.065872	0.077257
1stFlrSF	0.126259	0.058917	0.101411
RoofMatl_CompShg	0.322801	0.042096	0.214506
TotRmsAbvGrd	0.023647	0.037286	0.032066
FullBath	0.023858	0.033582	0.028150

	Linear	Ridge	Lasso
RoofMatl_CompShg	0.322801	0.042096	0.214506
RoofMatl_Tar&Grv	0.210131	0.025885	0.142908
RoofMatl_WdShngl	0.166667	0.032148	0.114700
1stFlrSF	0.126259	0.058917	0.101411
RoofMatl_WdShake	0.131563	0.010957	0.086215

After doubling Alpha Values:

Ridge & Lasso Regression:

	Ridge2	Lasso2
OverallQual	0.059293	0.083970
1stFlrSF	0.049723	0.082936
TotRmsAbvGrd	0.036936	0.038993
GarageArea	0.032398	0.034668
FullBath	0.031797	0.031069

	Ridge2	Lasso2
RoofMatl_CompShg	0.021943	0.123888
OverallQual	0.059293	0.083970
1stFlrSF	0.049723	0.082936
RoofMatl_Tar&Grv	0.012701	0.081352
RoofMatl_WdShngl	0.022457	0.070378

The top 5 values changed slightly like in Ridge regression RoofMatl_Compshg is replaced by GarageArea whereas, in Lasso Regression RoofMatl_WdShake is changed by OverallQual and even the order of the variables is changed.

Question 2:

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?

Ans:

I will choose lasso regression because it will make the coefficients for few variables to 0 which will help in feature selection but, it will not be case every time sometimes ridge regression will provide better accuracy.

Comparing the accuracy and other variables for the given case study, I will choose **Lasso**.

Question 3:

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?

Ans: After excluding the 5 most important variables based on lasso regression model which are

```
['RoofMatl_CompShg', 'RoofMatl_Tar&Grv', 'RoofMatl_WdShngl', '1stFlrSF',
       'RoofMatl_WdShake']
```

Then after re-modelling with lasso regression model these are the 5 best features which are obtained in new model.

	Lasso
OverallQual	0.092793
TotRmsAbvGrd	0.059672
FullBath	0.044841
GarageArea	0.043949
Neighborhood_Somerst	0.033011

Question 4:

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?

Answer: A model is considered robust and generalizable when the dependent variable makes an accurate prediction even when one or more of its predictor variables are changed in any circumstances. In other words, we can say that a model is robust (i.e, strong in simple English) when nothing effects it and still performs the same way. It is considered generalisable when its accuracy is not significantly impacted by performing on a set of data it has not seen before.

To demonstrate and explain better, a model should not have high bias and low complexity or vice versa. We need a model which is centred on both bias and variance that will work with high accuracy. The implications will be that when the model is too complex there will be a chance of overfitting. This will cause the model to accurately predict well in training data but not in test data.