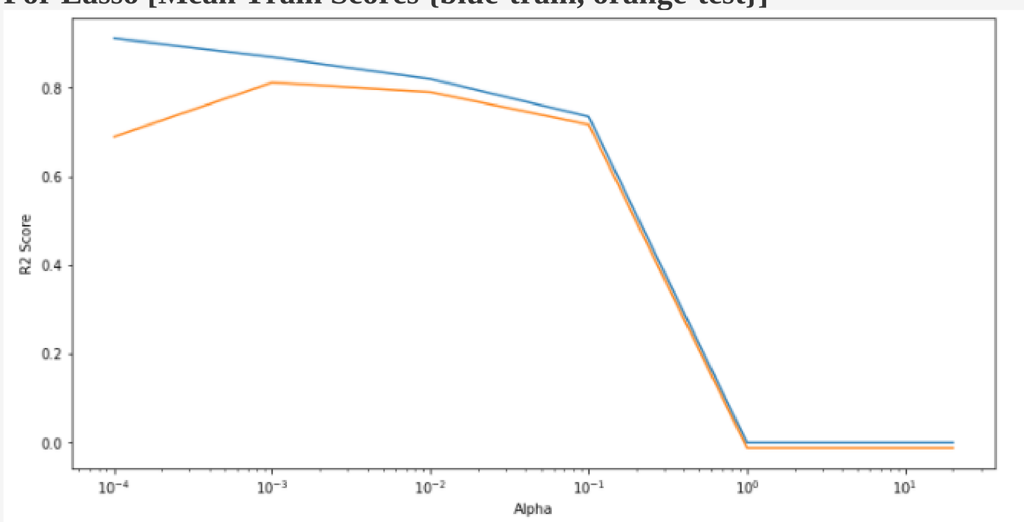
Question 1

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?

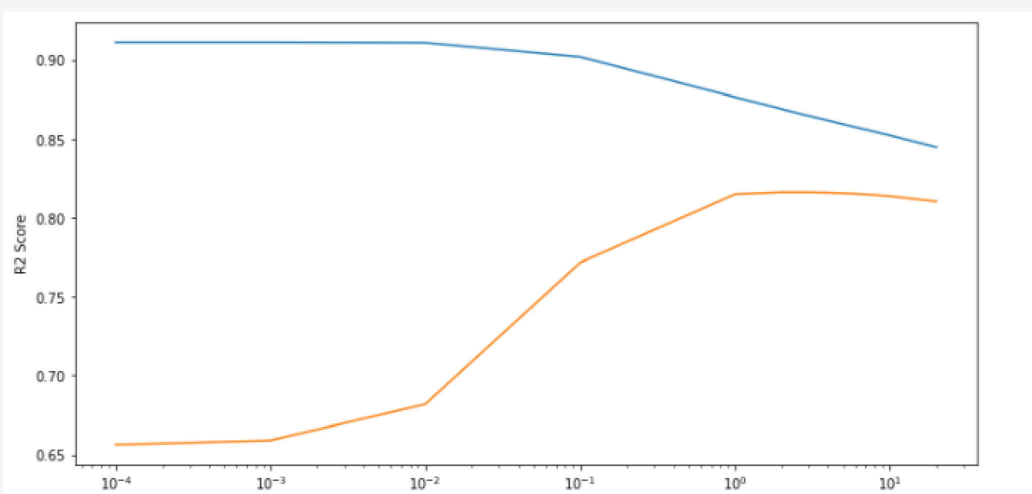
Answer:- For Lasso the best value obtained is **[0.001]** and for ridge it has come to **[2]**.

1) **Effect on R squared score for various alpha (Gridsearch operation)**

For Lasso [Mean Train Scores {blue-train, orange-test}]



For Ridge [Mean Train Scores {blue-train, orange-test}]



➔ Train and Test score for Lasso [0.001]

```
0.8653191073528599
0.8332269985340148
```

➔ Train and Test score for Lasso [0.002](doubled)

```
0.8914719437487669
0.8678950366993137
```

➔ Train and Test score of Ridge [2]

```
0.8679708993068388
0.8434980074415632
```

➔ Train and Test score of Ridge [4](doubled)

```
0.860855087785482
0.8462099182001803
```

The increase in alpha is more visible in Lasso as the train and test scores have changes considerably compared to the Ridge regularization.

2) Effect on number of coefficients Lasso minimized to 0

```
print(lasso_df[lasso_df['Alpha: 0.001'] == 0][['feature', 'Alpha: 0.001']].shape)
print(lasso_df[lasso_df['Alpha: 0.002'] == 0][['feature', 'Alpha: 0.002']].shape)
```

```
(19, 2)
(23, 2)
```

We can see that the number of features, which has been ignored, has increased when we double the alpha for lasso.

The selection effect on (important predictor) variable is consistent. **However, the importance (rank) of previously selected predictor variable has not changed, but the effect can be seen**

in their coefficients which has decreased substantially. Even after increasing the alpha the most important variable remains same which is **"Condition2" for both Ridge and Lasso**

To support the inference I have attached the screen shot of the difference for some of the features

→For Lasso

	feature	Alpha: 0.001	Alpha: 0.002	predictors
24	Condition2_PosN	-2.604454	-1.529683	Condition2
5	CentralAir	-0.465423	-0.450719	CentralAir
0	OverallQual	0.399210	0.411491	OverallQual
16	Neighborhood_Somerst	0.379891	0.378207	Neighborhood
13	Neighborhood_ClearCr	0.456038	0.368530	Neighborhood
6	1stFlrSF	0.326116	0.316658	1stFlrSF
10	MSZoning_RL	0.356535	0.294453	MSZoning
17	Neighborhood_Veenker	0.371549	0.290061	Neighborhood
40	Exterior2nd_Stucco	-0.377252	-0.284384	Exterior2nd

→ For Ridge

	feature	Alpha: 2	Alpha: 4	predictors
24	Condition2_PosN	-1.100863	-0.665187	Condition2
35	RoofMatl_WdShngl	0.578487	0.347626	RoofMatl
10	MSZoning_RL	0.551487	0.410052	MSZoning
29	RoofMatl_CompShg	0.531794	0.336471	RoofMatl
42	Heating_GasW	0.474618	0.337298	Heating
5	CentralAir	-0.452579	-0.448628	CentralAir
13	Neighborhood_ClearCr	0.443381	0.402299	Neighborhood
37	Exterior1st_BrkComm	-0.405699	-0.267622	Exterior1st
17	Neighborhood_Veenker	0.403724	0.347278	Neighborhood

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?

Answer:- In case of no constraint provided by the business or the requester regarding the retention of any particular features, it is advisable to stick with the lasso analysis. LASSO produces simpler and more interpretable models with reduced set of features. Lasso should not be used when there is greater coefficient difference for a set of beta compared to the others, but in this problem, the betas are comparable, and though there is bias – variance trade off in Lasso , the model needs simplification and Lasso is right candidate for it.

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?

Answer:-

Lasso Regression before and After

1) Before removing 5 top most predictors, please refer to Alpha:0.001

	feature	Alpha: 0.001	Alpha: 0.002	predictors
24	Condition2_PosN	-2.604454	-1.529683	Condition2
5	CentralAir	-0.465423	-0.450719	CentralAir
0	OverallQual	0.399210	0.411491	OverallQual
16	Neighborhood_Somerst	0.379891	0.378207	Neighborhood
13	Neighborhood_ClearCr	0.456038	0.368530	Neighborhood

2) After removing 5 topmost predictors and recreating the model again

	feature	Alpha: 0.001	predictors
47	MSZoning_FV	0.559524	MSZoning
116	RoofMatl_WdShngl	0.375777	RoofMatl
66	Neighborhood_MeadowV	-0.360308	Neighborhood
49	MSZoning_RL	0.343851	MSZoning
72	Neighborhood_NridgHt	0.342507	Neighborhood

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:- Generalization is a term used to describe a model's ability to react to new data. That is, after being trained on a training set, a model can digest new data and make accurate predictions. So to make a model generalized we need to design the test and train sets in a way so that the model is exposed to maximum variation of the data in question.

Generalization error could be measured by MSE.

To achieve this first step is to

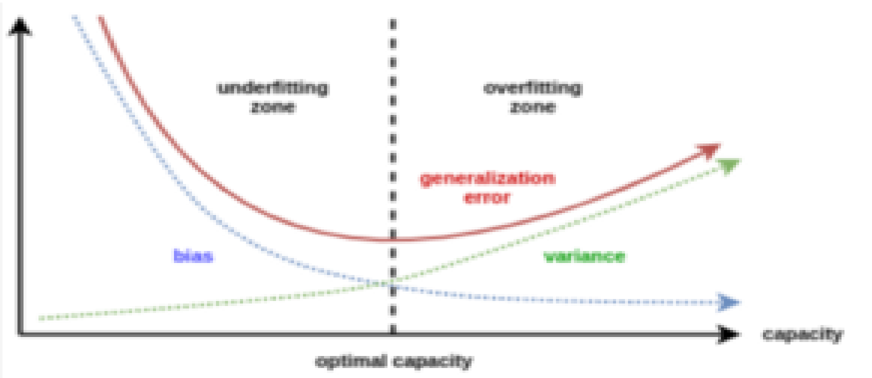
1) Imbibe more data (more data to train and test to)

2) Proper data preparation using outlier treatment, imputations

3) Feature engineering with derived features and feature selection which increases the metrics like (R2) which explains the data well.

4) Modelling followed by hyper parameter tuning/regularization

The model should not underfit or overfit the data and only that way generalization can be achieved.



As the model capacity increases, the bias decreases as the model fits the training datasets better. However, the variance increases, as your model become sophisticated to fit more patterns of the current dataset.

Bias is how much deviation the model has from the real values, as the model capacity increase it decreases the bias but the risk of overfitting also increases

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-y_i)^2$$

Small MSE produces fitted values closer to real values, increasing accuracy but with the risk of overfitting

High MSE → High Accuracy (risk of overfitting) and MSE is relative to outputs of other model results.

Low MSE →Low Accuracy (risk of underfitting).