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?

Optimal value of alpha:

Ridge – alpha = 1, with 41 features with a R^2(Test) 0.82499

```
Alpha: 1, R^2(Train): 0.8206308053471565, R^2(Test):
0.8249970901945431,f-c:41
Alpha: 2, R^2(Train): 0.8198421576782735, R^2(Test):
0.8251560361589055,f-c:41 (slightly better)
```

Lasso – alpha = 100 with 41 features with a R^2(Test) 0.8256

```
Alpha: 100, R^2(Train): 0.8192046912182973, R^2(Test):
0.8256270539877 88,f-c:41
Alpha: 200, R^2(Train): 0.8146813023705348, R^2(Test):
0.8226684383827453,f-c:41 (slightly reduced)
```

Type	lambda		comment
Ridge	1	Alpha: 0, R^2(Train): 0.8425064244819005, R^2(Test): 0.835510172277372,f-c:66 Alpha: 1, R^2(Train): 0.8409559218803891, R^2(Test): 0.8411386207120621,f-c:66 Alpha: 10, R^2(Train): 0.8280221665098821, R^2(Test): 0.8386500143969056,f-c:66 Alpha: 100, R^2(Train): 0.7450018404187573, R^2(Test): 0.7627674565385987,f-c:66 Alpha: 1000, R^2(Train): 0.4072993928877502, R^2(Test): 0.4199215364551364,f-c:66	1 for test set (0.8411) Features:66
Ridge	1	Alpha: 0, R^2(Train): 0.8210111931846006, R^2(Test): 0.8245578757165712,f-c:41 Alpha: 1, R^2(Train): 0.8206308053471565, R^2(Test): 0.8249970901945431,f-c:41 Alpha: 10, R^2(Train): 0.8097941955405731, R^2(Test): 0.8192760899748077,f-c:41 Alpha: 100, R^2(Train): 0.711627154739354, R^2(Test): 0.7189158844129373,f-c:41 Alpha: 1000, R^2(Train):	1for test set(0.82499) Features:41

		0.3459454186094152, R^2(Test): 0.3490819182038827,f-c:41	
Ridge	1	Alpha: 0, R^2(Train): 0.8182539318338804, R^2(Test): 0.8196781924337424,f-c:38 Alpha: 1, R^2(Train): 0.8178006127894363, R^2(Test): 0.819855514018333,f-c:38 Alpha: 10, R^2(Train): 0.8055165763611126, R^2(Test): 0.812546839856581,f-c:38 Alpha: 100, R^2(Train): 0.7039201077196066, R^2(Test): 0.7093659418754893,f-c:38 Alpha: 1000, R^2(Train): 0.336904376825254, R^2(Test): 0.3388632746668187,f-c:38	1for test set(0.81985) Features:38
Lasso	100	Alpha: 0.01, R^2(Train): 0.8425064239905273, R^2(Test): 0.8355107040464946,f-c:66 Alpha: 0.1, R^2(Train): 0.842506375480773, R^2(Test): 0.8355154613667504,f-c:66 Alpha: 1, R^2(Train): 0.842501950765217, R^2(Test): 0.8355848102076249,f-c:66 Alpha: 10, R^2(Train): 0.8423534930295526, R^2(Test): 0.8374522744707262,f-c:66 Alpha: 100, R^2(Train): 0.8356776733849647, R^2(Test): 0.8425359199977582,f-c:66 Alpha: 500, R^2(Train): 0.8142381427559354, R^2(Test): 0.8338386552858467,f-c:66 Alpha: 1000, R^2(Train): 0.789562735053901, R^2(Test): 0.8105746942775707,f-c:66	100 for test set(0.8425) Feature: 66
Lasso	100	Alpha: 0.01, R^2(Train): 0.8210111931561865, R^2(Test): 0.824558241365909,f-c:41 Alpha: 0.1, R^2(Train): 0.8210111903517205, R^2(Test): 0.8245614706078289,f-c:41 Alpha: 1, R^2(Train): 0.821010909171606 7, R^2(Test): 0.8245940868498021,f-c:41 Alpha: 10, R^2(Train): 0.8209857689830239, R^2(Test): 0.824913447584807,f-c:41 Alpha: 100, R^2(Train): 0.8192046912182973, R^2(Test): 0.825627053987788,f-c:41 Alpha: 500, R^2(Train):	100 for test set (0.8256) Feature: 41

		0.7981703589505361, R^2(Test): 0.8138508161153258,f-c:41 Alpha: 1000, R^2(Train): 0.7734406945376675, R^2(Test): 0.7945822187295016,f-c:41	
Lasso	10	Alpha: 0.01, R^2(Train): 0.8182539318168278, R^2(Test): 0.8196783355424202,f-c:38 Alpha: 0.1, R^2(Train): 0.8182539301142487, R^2(Test): 0.819679663372767,f-c:38 Alpha: 1, R^2(Train): 0.8182537609812434, R^2(Test): 0.8196924128841312,f-c:38 Alpha: 10, R^2(Train): 0.8182367616216548, R^2(Test): 0.8198091922698039,f-c:38 Alpha: 100, R^2(Train): 0.8165389077212369, R^2(Test): 0.8195848789366852,f-c:38 Alpha: 500, R^2(Train): 0.7951638161125933, R^2(Test): 0.807495454735629,f-c:38 Alpha: 1000, R^2(Train): 0.7669486856059752, R^2(Test): 0.7866618796322293,f-c:38	10 for test set(0.8198) Feature: 38

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?

Ridge – alpha = 1, with 41 features with a R^2(Test) 0.82499

```
Alpha: 1, R^2(Train): 0.8206308053471565, R^2(Test):
0.8249970901945431,f-c:41
Alpha: 2, R^2(Train): 0.8198421576782735, R^2(Test):
0.8251560361589055,f-c:41 (slightly better)
```

Lasso – alpha = 100 with 41 features with a R^2(Test) 0.8256

```
Alpha: 100, R^2(Train): 0.8192046912182973, R^2(Test):
0.825627053987788,f-c:41
Alpha: 200, R^2(Train): 0.8146813023705348, R^2(Test):
0.8226684383827453,f-c:41 (slightly reduced)
```

Looking at above data both methods are very close - I would go with Ridge, with alpha value of 2, since it turns out to be better.

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?

The five most important variables are:

```
13            YrSold  1.066262
31   BsmtFinType2_enc  1.099073
40       SaleType_enc  1.109702
16      LotConfig_enc  1.118267
36     Functional_enc  1.123404
```

After removing above features – below are the top five features:

```
36  SaleCondition_enc  1.115492
20       RoofMatl_enc  1.134775
26       BsmtCond_enc  1.159018
16   Neighborhood_enc  1.169231
13       LotShape_enc  1.173644
```

Ridge:

```
Alpha: 2, R^2(Train): 0.8187824573686472, R^2(Test):
0.824026926828994,f-c:36
```

Lasso:

```
Alpha: 100, R^2(Train): 0.818399697268194, R^2(Test):
0.825052838278836,f-c:36

This shows that there is further scope to reduce the complexity without
impacting the performance.
```

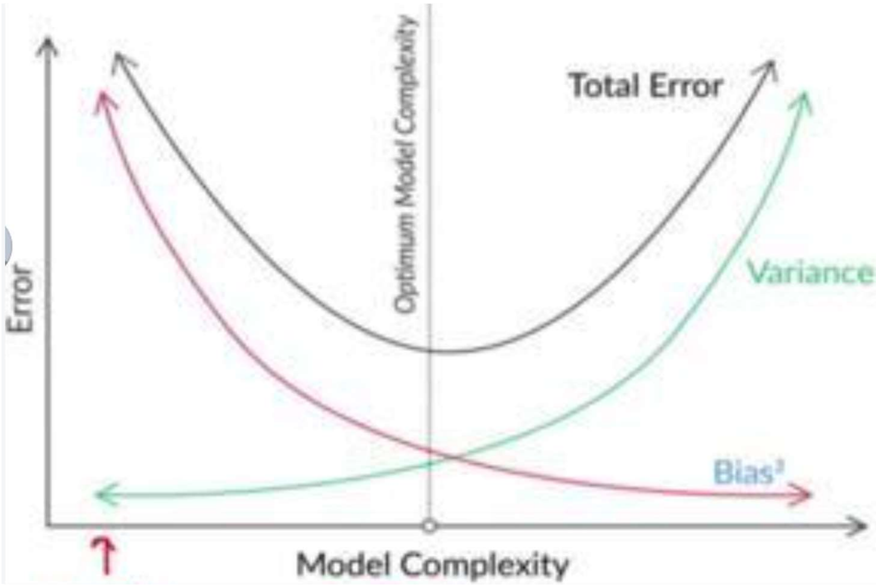
Question 4

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?

To make sure the model is generalisable, we need to reduce the complexity of the model by avoiding overfitting, overfitting will reduce the accuracy on the test data. We can improve the model by:

1. reducing the number of independent variables by removing insignificant features.
2. By compromising on the Bias a bit and bringing the variance drastically low we can improve the accuracy of the model by balancing Bias and Variance of the model.

We do this by finding the optimum model complexity where model performs at its best on un-seen data as well.


Optimum Model Complexity
Total Error
Variance
Error
Bias²
Model Complexity