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:

Please find Optimal alpha value below,

Ridge – 50

r2 score on training set: 0.860263058800402

r2 score on test set: 0.8603786623276777

rss on training set: 142.67141696478956

rss on test set: 62.97382580230949

mse on training set: 0.139736941199598

mse on test set: 0.1437758579961404

Lasso - 0.001

r2 score on training set: 0.8986084414291955

r2 score on test set: 0.8500830388467437

rss score on training set: 103.52078130079144

rss score on test set: 67.61749137967365

mse score on training set: 0.10139155857080455

mse score on test set: 0.15437783420016815

Below is the output if we double the alpha on both Ridge and Lasso,

The r2 score on training set reduced by a small margin and increases on test set.

Ridge – alpha – 100

r2 score on training set: 0.848804673974609 r2 score on test set: 0.8556984102224094

rss on training set: 154.3704278719242

rss on test set: 65.08477378276623

mse on training set: 0.15119532602539099

mse on test set: 0.14859537393325625

Lasso – alpha – 0.002

The r2 score on training set reduced by a small margin and increases on test set.

r2 score on training set: 0.8843684383570555

r2 score on test set: 0.8595000319140403

rss score on training set: 118.05982443744631

rss score on test set: 63.37011708224885

mse score on training set: 0.11563156164294447

mse score on test set: 0.1446806326078741

Below are the important predictor after we double the alpha,

```
                      Feature  Coefficient
77       Conditions_PosN and PosN    -1.660031
57                BldgType_Twnhs    -0.210483
81     Exteriors_Stucco and Stucco    -0.183539
58               BldgType_TwnhsE    -0.136919
50          Neighborhood_Edwards    -0.111771
..                           ...          ...
15                     GrLivArea     0.342814
53          Neighborhood_NridgHt     0.461078
52          Neighborhood_NoRidge     0.499973
64              RoofMatl_WdShngl     0.688421
83   Exteriors_Wd Sdng and ImStucc     0.799759

[85 rows x 2 columns]
```

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?

I would use the Lasso Regression because it has almost similar or better prediction and it removes the in-significant variables coefficients from the analysis. Using this, Most of the times, We don't need to execute rfe or VIF to remove the variables that have high multi-collinearity. It takes care of it by itself.

Although Ridge reduces the coefficients to very near to zero but it still making the impact on the prediction.

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?

Below are the top 10 predictor variables,

	Feature	Coefficient
102	Conditions_PosN and PosN	-2.857577
70	BldgType_Twnhs	-0.279662
109	Exteriors_Stucco and Stucco	-0.234920
45	LotShape_IR3	-0.202534
71	BldgType_TwnhsE	-0.191413
...	...	...
16	GrLivArea	0.353188
63	Neighborhood_NoRidge	0.483362
64	Neighborhood_NridgHt	0.507359
79	RoofMatl_WdShngl	1.135937
112	Exteriors_Wd Sdng and ImStucc	1.424672

116 rows × 2 columns

If the top 5 are removed, the remaining 5 will be,
Neighborhood_NridgHt 0.507359

Neighborhood_NoRidge 0.483362

GrLivArea 0.353188

LotShape_IR3-0.202534

BldgType_TwnhsE -0.191413

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?

1. The model can be called as Robust if it's not having high variance on the unseen data – test data. The Model should not be trained with too many predictors so that it learns the underneath pattern of the data so that it shows good r2 score on train set but performs bad on the test set.

2. We should not have too many predictor variables that makes the model complex. This can be achieved by using various Feature elimination technuques, mainly RFE,VIF etc or we can use Lasso Regression.

   Lasso regression, like a good scout, uses a different penalty term that drives some of the feature coefficients all the way down to zero. This effectively removes irrelevant features from the model, leading to a simpler and more interpretable model .
3. There's a trade-off when using Ridge and Lasso. The model's accuracy on the training data might decrease slightly due to regularization. However, this is a small price to pay for a model that can handle variations in the real world and perform well on unseen data.
4. The model should not have multicollinear variables.

The downside of making a model robust and generalisable are,

1. The prediction performance on training set might reduce.
2. The regularization causes little high bias but less variance.
3. Regularization adds a penalty for overly complex models, and data augmentation introduces variations that the model has to learn to handle. Early stopping stops training before the model fully adapts to noise in the training data.
4. The focus shifts from achieving peak accuracy on a specific dataset to building a model that can adapt and perform well on a broader range of scenarios.
5. a slight decrease in training accuracy is a reasonable price to pay for a model that performs well in diverse real-world conditions.