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 for ridge regression is 9.0 and for lasso regression is 0.05.

Increasing the alpha by double will slightly lower the r2 score for the models and there is a change in the coeff values of the predictors.

The features and their coeff are as follows:

Before changing the alpha value

	Ridge
YearBuilt	-0.180725
Neighborhood_Gilbert	-0.176522
Neighborhood_NWAmes	-0.160315
Neighborhood_Edwards	-0.132355
Neighborhood_MeadowV	-0.110426
...	...
TotalBsmtSF	0.151373
Neighborhood_BrkSide	0.155995
Neighborhood_Somerst	0.188324
Neighborhood_Crawfor	0.307521
GrLivArea	0.346659

147 rows × 1 columns

	Lasso
YearBuilt	-0.197759
Neighborhood_Gilbert	-0.171796
MSSubClass_2-STORY PUD - 1946 & NEWER	-0.138474
Neighborhood_NWAmes	-0.137572
Neighborhood_MeadowV	-0.114141
...	...
Neighborhood_BrkSide	0.224990
Neighborhood_StoneBr	0.252940
Neighborhood_Somerst	0.346874
GrLivArea	0.373403
Neighborhood_Crawfor	0.503117

88 rows × 1 columns

After doubling the alpha value

	Ridge
YearBuilt	-0.172702
Neighborhood_Gilbert	-0.138629
Neighborhood_NWAmes	-0.119192
Neighborhood_Edwards	-0.104552
Neighborhood_OldTown	-0.087078
...	...
Neighborhood_Somerst	0.141623
OverallQual	0.145068
TotalBsmtSF	0.146693
Neighborhood_Crawfor	0.230250
GrLivArea	0.330881

147 rows × 1 columns

	Lasso
YearBuilt	-0.208607
Neighborhood_Gilbert	-0.137090
Neighborhood_NWAmes	-0.111833
MSSubClass_2-STORY PUD - 1946 & NEWER	-0.088215
Neighborhood_Edwards	-0.075732
...	...
Exterior1st_BrkFace	0.185835
Neighborhood_BrkSide	0.212024
Neighborhood_Somerst	0.308895
GrLivArea	0.381207
Neighborhood_Crawfor	0.445897

68 rows × 1 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?

We will be choosing Lasso regression, as in this the model becomes much simpler as compared to ridge without losing a lot of accuracy value, using the feature selection, thus allowing us to give us better features to look at.

Simpler models are much more robust, and as we can see from our metrics derived there is no overfitting, thus the model is well generalized. Due to these reasons, it is slightly more better to use 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?

If these predictor variables are not available then we would have to create a new model. The top 5 important variables of the original models are:

1. Neighborhood_Crawfor

2. GrLivArea

3. Neighborhood_Somerst

4. Neighborhood_StoneBr

5. Neighborhood_BrkSide

	Lasso
YearBuilt	-0.197759
Neighborhood_Gilbert	-0.171796
MSSubClass_2-STORY PUD - 1946 & NEWER	-0.138474
Neighborhood_NWAmes	-0.137572
Neighborhood_MeadowV	-0.114141
...	...
Neighborhood_BrkSide	0.224990
Neighborhood_StoneBr	0.252940
Neighborhood_Somerst	0.346874
GrLivArea	0.373403
Neighborhood_Crawfor	0.503117

Now as we do not have these variables, we get

New Model with lasso regression:

The new top 5 variables are

1. BsmtFinType1_NA

2. Neighborhood_Gilbert

3. TotalBsmtSF

4. Neighborhood_NWAmes

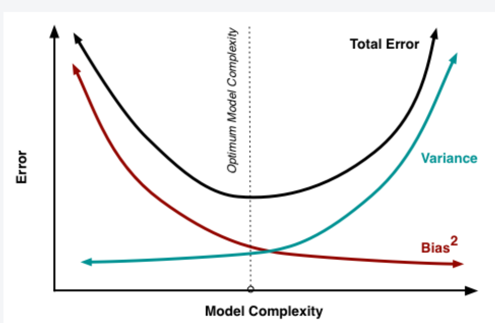
5. Neighborhood_MeadowV

	Lasso
Neighborhood_Gilbert	-0.398439
Neighborhood_NWAmes	-0.321974
Neighborhood_MeadowV	-0.315827
Neighborhood_Edwards	-0.276921
Neighborhood_Sawyer	-0.265583
...	...
MSZoning_FV	0.235862
Exterior2nd_CmentBd	0.269399
2ndFlrSF	0.284232
TotalBsmtSF	0.322083
BsmtFinType1_NA	0.521519

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?

Our aim is to minimize the total error, which can be kept in check considering the bias-variance tradeoff. This states as if we have high variance and low bias the model will be more complex and on other hands with high bias and low variance we will have simpler model.



As the complexity increase in variance cause the model to overfit and thus generalizability takes a hit, a more simple model will basically underfit and this would not be able to understand the patterns. We need a model which works on this tradeoff.

If our model is overfitting, i.e., the difference between train and test accuracies is very high, we need to reduce the complexity of our model, which can be done using different methods such as dropping predictor variables, regularization etc.

This will decrease the train accuracy but will also decrease the difference between the train and test and make the model more robust.

If our model is underfitting, it will not be able to perform well in detecting train data even, this can be solved by increasing the complexity of the model, which can be done by either increasing the number of predictor variables if available, using more data etc.

This will increase the accuracy of the model, thus making it more generalizable.