Question 1

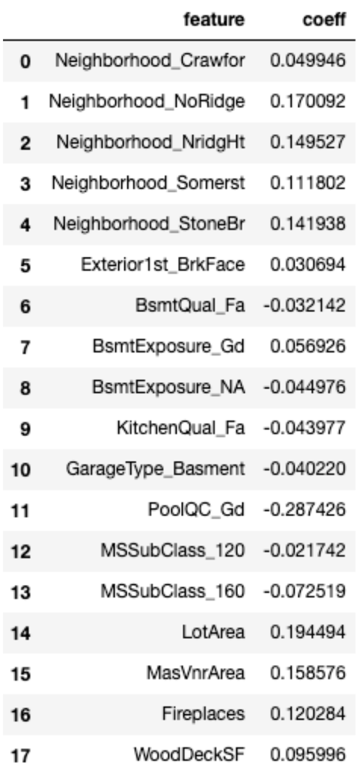
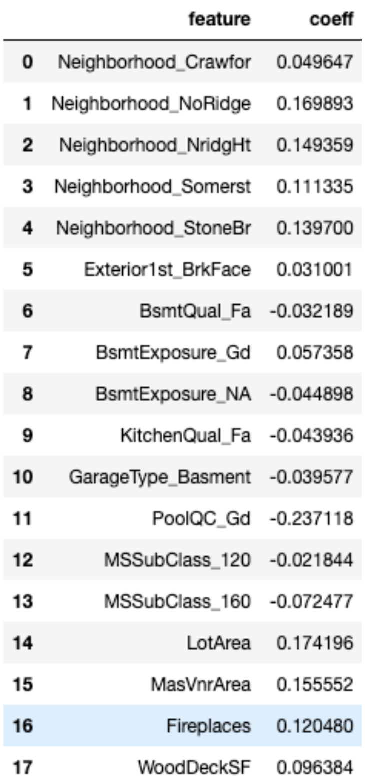
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

The optimal values of alpha are
Ridge Regression: 0.01
Lasso Regression: 0.0001

Using the above alpha values the results are

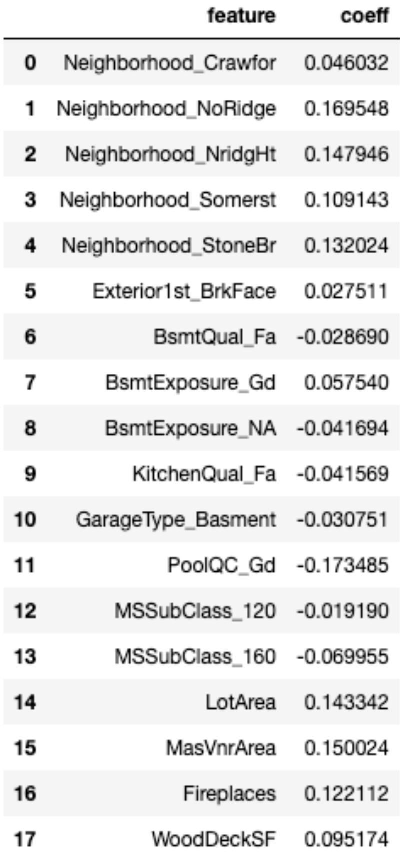
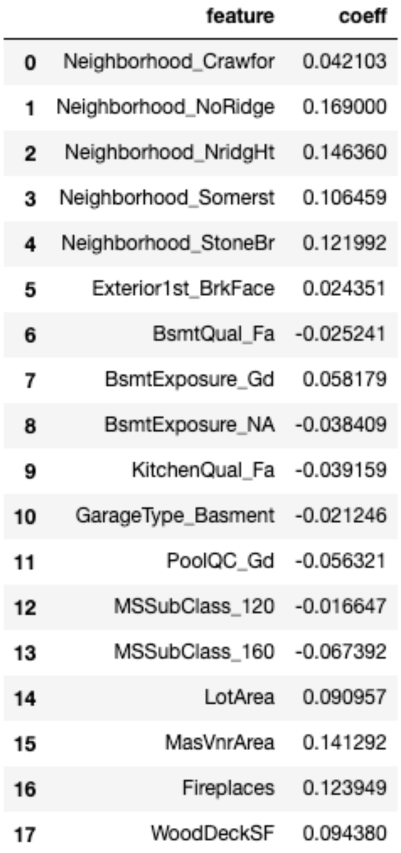
Regression	Train R2	Test R2	Train RMSE	Test RMSE
Ridge (alpha =0.01)	0.64	0.58	0.07	0.07
Ridge (alpha=0.01*2)	0.64	0.59	0.07	0.07
Lasso (alpha=0.0001)	0.64	0.6	0.07	0.07
Lasso (alpha=0.0001 * 2)	0.63	0.6	0.07	0.07

If the alpha values are doubled then there is no significant difference in the RMSE and R-squared values. But the values of the coefficients are slightly smaller after doubling the alpha values.

Ridge Coefficients	Ridge doubled alpha coefficients
	

	feature	coeff
0	Neighborhood_Crawfor	0.049946
1	Neighborhood_NoRidge	0.170092
2	Neighborhood_NridgHt	0.149527
3	Neighborhood_Somerst	0.111802
4	Neighborhood_StoneBr	0.141938
5	Exterior1st_BrkFace	0.030694
6	BsmtQual_Fa	-0.032142
7	BsmtExposure_Gd	0.056926
8	BsmtExposure_NA	-0.044976
9	KitchenQual_Fa	-0.043977
10	GarageType_Basment	-0.040220
11	PoolQC_Gd	-0.287426
12	MSSubClass_120	-0.021742
13	MSSubClass_160	-0.072519
14	LotArea	0.194494
15	MasVnrArea	0.158576
16	Fireplaces	0.120284
17	WoodDeckSF	0.095996

	feature	coeff
0	Neighborhood_Crawfor	0.049647
1	Neighborhood_NoRidge	0.169893
2	Neighborhood_NridgHt	0.149359
3	Neighborhood_Somerst	0.111335
4	Neighborhood_StoneBr	0.139700
5	Exterior1st_BrkFace	0.031001
6	BsmtQual_Fa	-0.032189
7	BsmtExposure_Gd	0.057358
8	BsmtExposure_NA	-0.044898
9	KitchenQual_Fa	-0.043936
10	GarageType_Basment	-0.039577
11	PoolQC_Gd	-0.237118
12	MSSubClass_120	-0.021844
13	MSSubClass_160	-0.072477
14	LotArea	0.174196
15	MasVnrArea	0.155552
16	Fireplaces	0.120480
17	WoodDeckSF	0.096384

Lasso coefficients	Lasso Doubled Alpha coefficients
	

	feature	coeff
0	Neighborhood_Crawfor	0.046032
1	Neighborhood_NoRidge	0.169548
2	Neighborhood_NridgHt	0.147946
3	Neighborhood_Somerst	0.109143
4	Neighborhood_StoneBr	0.132024
5	Exterior1st_BrkFace	0.027511
6	BsmtQual_Fa	-0.028690
7	BsmtExposure_Gd	0.057540
8	BsmtExposure_NA	-0.041694
9	KitchenQual_Fa	-0.041569
10	GarageType_Basment	-0.030751
11	PoolQC_Gd	-0.173485
12	MSSubClass_120	-0.019190
13	MSSubClass_160	-0.069955
14	LotArea	0.143342
15	MasVnrArea	0.150024
16	Fireplaces	0.122112
17	WoodDeckSF	0.095174

	feature	coeff
0	Neighborhood_Crawfor	0.042103
1	Neighborhood_NoRidge	0.169000
2	Neighborhood_NridgHt	0.146360
3	Neighborhood_Somerst	0.106459
4	Neighborhood_StoneBr	0.121992
5	Exterior1st_BrkFace	0.024351
6	BsmtQual_Fa	-0.025241
7	BsmtExposure_Gd	0.058179
8	BsmtExposure_NA	-0.038409
9	KitchenQual_Fa	-0.039159
10	GarageType_Basment	-0.021246
11	PoolQC_Gd	-0.056321
12	MSSubClass_120	-0.016647
13	MSSubClass_160	-0.067392
14	LotArea	0.090957
15	MasVnrArea	0.141292
16	Fireplaces	0.123949
17	WoodDeckSF	0.094380

The important predictors still remain the same even after doubling the alpha values

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 choose the Lasso regression model since it has a very less difference between the train R-squared and test R-squared. This shows a very less chance of overfitting to the training data. Regularized models which can generalize more can give better accuracy on unseen data.
The RMSE of both the models is almost same. But, some of the coefficients in lasso can become zero hence this can also reduce features while maintaining same R2 and RMSE.

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 import 5 features are

1. Neighborhood_NoRidge
2. Neighborhood_NridgHt
3. MasVnrArea
4. Fireplaces
5. Neighborhood_StoneBr

Excluding these features has the following affect

	Train R2	Test R2	Train RMSE	Test RMSE
Before dropping	0.64	0.6	0.07	0.07
After dropping	0.29	0.26	0.09	0.1

After dropping these features there is a significant drop in the R sqaured value and also an increase in RMSE value.
The available best predictors are:

1. LotArea
2. WoodDeckSF
3. Neighborhood_Somerst
4. BsmtExposure_Gd
5. Neighborhood_Crawfor

Question 4

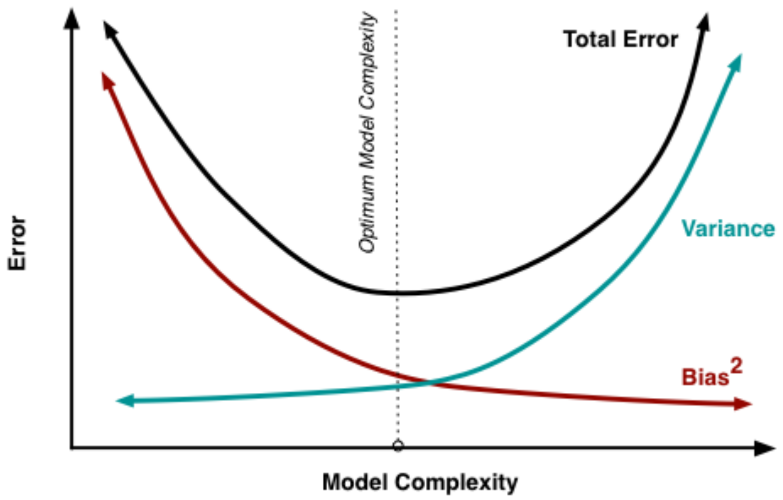
How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

According to Occam's Razor, given any 2 models that show similar performance on the training set and test set. A simpler model has to be chosen among them. A simpler model will have lesser predictors and smaller coefficients.

1. Simpler models are more robust and generalisable. The models must be simple but not very simple such that they are not usable
2. Simple models can be trained on smaller training datasets effectively than complex models.
3. Complex models tends to vary with small changes in the training datasets. They can have very good accuracy on training set but will perform very badly on test sets
4. In other words complex models have high variance and low bias. This is a typical overfitting condition. Similarly a model with very less variance and high bias is an underfitted model. Both the underfitted and overfitted models are not reliable
5. Hence, a model with a balance between the bias and variance is more reliable model. This is also known as bias variance trade-off

6. Regularization can help us to reduce the coefficients of the model and make the model more simple. Also, Lasso regression can make few coefficient as 0 further reducing the total number of predictors

The below picture shows the bias variance curves and the model complexity. We should choose a model with optimal bias and variance while keeping the total error low.


Optimum Model Complexity
Total Error
Variance
Bias^2
Error
Model Complexity