Question1

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?

Answer1:

Ridge Regression

For ridge regression optimal value of alpha=3. If we chose double the value of alpha for ridge i.e 4 below were observed:

a) Value of model coefficients reduced
b) Ridge imposes more aggressive penalty as it uses sum of square of all beta coefficients (L2 norm) as shrinking penalty.

At Alpha2

	Params	Coef
7	1stFlrSF	0.211
8	2ndFlrSF	0.171
4	BsmtFinSF1	0.125
17	GarageCars	0.099
2	LotArea	0.088
31	Neighborhood_Crawfor	0.079
6	BsmtUnfSF	0.071
56	OverallCond_8	0.070

At Alpha4

	Params	Coef
7	1stFlrSF	0.191
8	2ndFlrSF	0.147
4	BsmtFinSF1	0.117
17	GarageCars	0.093
2	LotArea	0.087
31	Neighborhood_Crawfor	0.074
6	BsmtUnfSF	0.066

Lasso Regression

Due to the fact that coefficients will be shrunk towards a mean of zero, less important features in a dataset are eliminated when penalized. The shrinkage of these coefficients based on the alpha value provided leads to some form of automatic feature selection, as input variables are removed in an effective approach.

Alpha 0.0001

	Params	Coef
0	constant	-0.024
1	LotFrontage	0.001
2	LotArea	0.087
3	MasVnrArea	0.029
4	BsmtFinSF1	0.135
5	BsmtFinSF2	0.041
6	BsmtUnfSF	0.079
7	1stFlrSF	0.236

Alpha 0.0002

If we increase the value of alpha then shrinking penalty will be higher, so Ridge and Lasso both will try to shrink values of beta coefficients towards zero, so our model will be simpler. That means it will increase the bias where variance will be reduced. If we increase the value of alpha to a very large number, then all coefficients of Lasso become 0 and for Ridge coefficients become close to zero (as they cannot be exact 0 in Ridge). That means the model will have very high bias and low variance and it may result in underfitting. That means model will fail to learn the underlying data pattern in training dataset.
If we reduce the value of alpha then shrinking penalty will be lower, so model bias will reduce, and variance will increase. Now if we put value of alpha as 0, then the cost function of both Ridge and Lasso become OLS cost function (i.e., RSS) and we will get exact same model as we get using OLS. So, reducing value of alpha will reduce the effect of shrinking penalty may lead to possible overfitting for very low or close to zero value of alpha.
So, we need to find the optimal value of alpha by performing hyperparameter tunning.

Question2

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?

As per Occam's Razor a model should not be unnecessary complex.
Model complexity depends on two main things: No. of features or independent variables and Magnitude of beta coefficients. Normalization (Ridge and Lasso) already shrinks beta coefficients towards zero.

Now, Lasso and Ridge both have similar r2 score and MAE on test dataset. No of features in Lasso Regression model is 88. Where Ridge has all 100 features. So, the Lasso model is simpler than Ridge with having similar r2 score and MAE.
Ridge:
r2 score on testing dataset: 0.88
MSE on testing dataset: 0.008
Lasso:
r2 score on testing dataset: 0.88
MSE on testing dataset: 0.008
As these two models shows almost similar performance on test dataset, we should choose the simpler model. So, I will choose Lasso as my final model

Question3

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 variable for lasso are

7	1stFlrSF	0.236
8	2ndFlrSF	0.200
4	BsmtFinSF1	0.135
17	GarageCars	0.100
2	LotArea	0.087

After dropping five most important variable Below are the 5 most important predictors now

12	TotRmsAbvGrd	0.180
66	BsmtExposure_Gd	0.131
0	constant	0.122
26	Neighborhood_Crawfor	0.098
63	BsmtQual_Ex	0.082
47	OverallQual_8	0.071

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?

The model should be generalized so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training. Too much importance should not given to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, the outliers analysis needs to be done and only those which are relevant to the dataset need to be retained. Those outliers which it does not make sense to keep must be removed from the dataset. If the model is not robust, It cannot be trusted for predictive analysis.