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
The optimal value of alpha for Ridge is 10 and Lasso is 0.0004. Using these alphas, the R2 Score for the test set in both Ridge and Lasso was almost similar.

After doubling the values of alpha for both ridge and lasso, there were negligible changes in the mean squared error and R2 Score. Though some major changes were seen in the coefficients and the features. Some new features were now in the significant list and some previous features were not.

- *Significant features before doubling alpha in Lasso model*

Out[48]:

	Features	rfe_support	rfe_ranking	Coefficient
36	SaleCondition_Partial	True	1	0.0984
41	Neighborhood_Crawfor	True	1	0.0875
55	Exterior1st_BrkFace	True	1	0.0808
72	Foundation_PConc	True	1	0.0719
37	MSZoning_FV	True	1	0.0685
35	SaleCondition_Normal	True	1	0.0605
0	OverallQual	True	1	0.0566
1	OverallCond	True	1	0.0408
69	Exterior2nd_Wd Sdng	True	1	0.0364
11	BsmtExposure_Gd	True	1	0.0362

- *Significant features after doubling alpha in Lasso model*

	Features	rfe_support	rfe_ranking	Coefficient
36	SaleCondition_Partial	True	1	0.0909
55	Exterior1st_BrkFace	True	1	0.0780
41	Neighborhood_Crawfor	True	1	0.0655
72	Foundation_PConc	True	1	0.0642
0	OverallQual	True	1	0.0631
35	SaleCondition_Normal	True	1	0.0556
1	OverallCond	True	1	0.0421
6	GarageCars	True	1	0.0307
11	BsmtExposure_Gd	True	1	0.0274
37	MSZoning_FV	True	1	0.0254

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?

Answer
The optimal value of lambda determined for Ridge and Lasso Regression are as follows.

- Ridge - 10
- Lasso - 0.0004

Using these lambdas, models with the following stats were generated

- Ridge
    - *R2 Score for Train set*: 0.9262
    - *R2 Score for Test set*: 0.8832
    - *Mean Square error* : 0.01548
- Lasso
    - *R2 Score for Train set*: 0.9254
    - *R2 Score for Test set*: 0.8820
    - *Mean Square error* : 0.01566

So, both the techniques generated models with very similar stats. So, I have chosen the Lasso model as my final model because it helps in feature reduction (coefficients of some features are zero).

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?

Answer
The top 5 most significant features of the chosen lasso model are

1. *SaleCondition_Partial*
2. *Neighborhood_Crawfor*
3. *Exterior1st_BrkFace*
4. *Foundation_PConc*
5. *MSZoning_FV*

On removing the top 5 features, the new model generated the following stats

```
[45]  ✓ 0.1s

...   R2score for Train set: 0.9042544065025516
      R2score for Test set: 0.8698455204962544
      Mean_squared_error:  0.017265435625841485
```

It is observed that the R2 Score reduced and the Mean Squared error increased. With this, the significant features list also changed.

```
[48]  ✓ 0.7s
```

	Features	rfe_support	rfe_ranking	Coefficient
0	OverallQual	True	1	0.0641
43	Exterior2nd_BrkFace	True	1	0.0555
33	LotConfig_CulDSac	True	1	0.0302
1	BsmtFullBath	True	1	0.0294
12	BsmtExposure_Gd	True	1	0.0267
36	MasVnrType_Stone	True	1	0.0260
5	GarageCars	True	1	0.0257
3	HalfBath	True	1	0.0255
4	Fireplaces	True	1	0.0219
2	FullBath	True	1	0.0166

The top 5 most significant features with the new model are

1. *OverallQual*
2. *Exterior2nd_BrkFace*
3. *LotConfig_CulDSac*
4. *BsmtFullBath*
5. *BsmtExposure_Gd*

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?

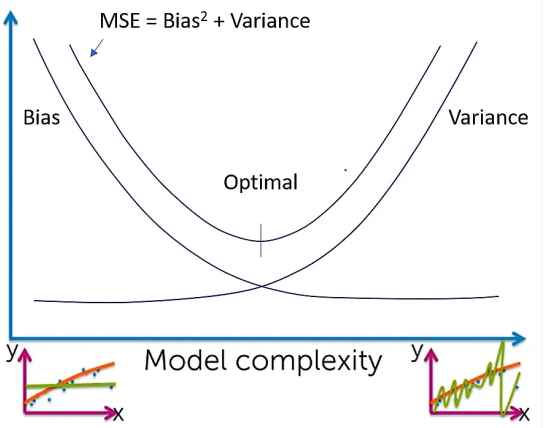
Answer

Models that are simple are more robust and generalisable. As per Occam's Razor, a simple model

- Is usually more 'generic' and thus can be easily applicable in a wide range of test data.
- Can be easily trained with less training samples when compared to a complex model
- Is less susceptible to changes than a complex model.
- Also avoids overfitting which is usually very common with complex model

In order to make sure that the model we generate does not become a complex model, Regularization techniques can be used. This involves adding a regularising term to the cost equation. Thus, making sure that there is an additional cost associated with complexity.

When the accuracy of the model is considered, complex models usually perform better with the train data as they are better capable of explaining the variance but fail mostly on the test data (high variance, low bias) whereas simple models usually aren't too accurate with the train data. (high bias, low variance)



So, to improve the accuracy of the model, a model complexity where the total error is the lowest should be considered.