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 1

Lasso Regression –

Optimal Value of Alpha: 0.001

If the optimal value is doubled (0.002), the train and test accuracy is reduced but the consistency is increased. It clearly indicates higher the alpha, higher is the consistency but lower is the accuracy. It is the bias variance trade-off.

Accuracy	Lasso Regression	
	Alpha: 0.001	Alpha: 0.002
Train Accuracy	0.853514473	0.816445119
Test Accuracy	0.850499604	0.81423903

Change in Predictors: The top 5 predictors remain the same. Only order of top 5 has changed.

Lasso Regression - Alpha 0.001			
Top 5	Feature	Coef	Absolute Coef
1	GrLivArea	0.243788	0.243788
2	OverallQual	0.132038	0.132038
3	GarageArea	0.054852	0.054852
4	BsmtFinSF1	0.051577	0.051577
5	KitchenQual	0.04846	0.04846

Lasso Regression - Alpha 0.002			
Top 5	Feature	Coef	Absolute Coef
1	GrLivArea	0.192256	0.192256
2	OverallQual	0.12511	0.12511
3	KitchenQual	0.053255	0.053255
4	BsmtFinSF1	0.039631	0.039631
5	GarageArea	0.039237	0.039237

Ridge Regression –

Optimal Value of Alpha: 5.0

If the optimal value is doubled (10.0), the train and test accuracy is reduced but the consistency is increased. It clearly indicates higher the alpha, higher is the consistency but lower is the accuracy. It is the bias variance trade-off.

Accuracy	Ridge Regression	
	Alpha: 5.0	Alpha: 10.0
Train Accuracy	0.909388382	0.897994362
Test Accuracy	0.878395686	0.878223312

Change in Predictors: The top 5 predictors remain the same. Only order of top 5 has changed.

Ridge Regression - Alpha 5.0			
Top 5	Feature	Coef	Absolute Coef
1	RoofMatl_WdShngl	0.083791	0.083791
2	OverallQual	0.076204	0.076204
3	GrLivArea	0.071524	0.071524
4	2ndFlrSF	0.067378	0.067378
5	1stFlrSF	0.063716	0.063716

Ridge Regression - Alpha 10.0			
Top 5	Feature	Coef	Absolute Coef
1	OverallQual	0.064864	0.064864
2	GrLivArea	0.059811	0.059811
3	RoofMatl_WdShngl	0.057261	0.057261
4	1stFlrSF	0.055137	0.055137
5	2ndFlrSF	0.053569	0.053569

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 2

The final model which is selected is the model based on Lasso Regression. There are 2 reasons for this decision.

I. R2_score are almost same for both. There is no case of overfitting and there is low variance
II. It will help in feature selection as there are coefficient with 0 value

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 five most important predictors in the final Lasso Regression model were.

Lasso Regression - Alpha 0.001			
Top 5	Feature	Coef	Absolute Coef
1	GrLivArea	0.243788	0.243788
2	OverallQual	0.132038	0.132038
3	GarageArea	0.054852	0.054852
4	BsmtFinSF1	0.051577	0.051577
5	KitchenQual	0.04846	0.04846

After removing these models and keeping the same alpha as 0.001, the new top 5 predictor variables are as follows –

Lasso Regression - Alpha 0.001 (After removing top 5)			
Top 5	Feature	Coef	Absolute Coef
1	1stFlrSF	0.170602	0.170602
2	2ndFlrSF	0.133317	0.133317
3	ExterQual	0.08376	0.08376
4	TotalBsmtSF	0.073775	0.073775
5	GarageCars	0.051895	0.051895

Question 4

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

Answer 4

There is a concept of bias variance trade-off which can help us understand this scenario. If we want our model to be more robust and generalizable, then we would need to have low variance in the model. This would mean that the model would have high consistency. To achieve this, we can increase the value of lambda to ensure that the model is more robust and generalizable.

On the contrary, by doing this we would take a hit on the bias. If we increase the value of lambda (or make the model more generalizable), the accuracy of the model would reduce.

This happens because the model is not complex enough to be very accurate.

It is clearly seen in Answer of Question 1 when increasing the value of lambda lead to high consistency but comparatively lower accuracy.