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:

Optimal alpha

- Ridge: 7.0
- Lasso: 0.0001

Most Important predictors (along with Coeff):

	Ridge	Lasso
With optimal alpha	('OverallQual', 0.085), ('2ndFlrSF', 0.07), ('GrLivArea', 0.069), ('Neighborhood_NoRidge', 0.06), ('RoofMatl_WdShngl', 0.056), ('1stFlrSF', 0.05), ('TotRmsAbvGrd', 0.041), ('GarageCars', 0.04), ('MasVnrArea', 0.039), ('KitchenQual', -0.038),	('GrLivArea', 0.335), ('Condition2_PosN', -0.287), ('RoofMatl_WdShngl', 0.144), ('OverallQual', 0.138), ('PoolQC', -0.075), ('Neighborhood_NoRidge', 0.058), ('GarageCars', 0.055), ('2ndFlrSF', 0.05), ('MasVnrArea', 0.048), ('PoolArea', -0.048),
With optimal alpha doubled	('OverallQual', 0.068), ('Neighborhood_NoRidge', 0.056), ('GrLivArea', 0.054), ('2ndFlrSF', 0.052), ('1stFlrSF', 0.04), ('KitchenQual', -0.039), ('TotRmsAbvGrd', 0.039), ('GarageCars', 0.036), ('RoofMatl_WdShngl', 0.036), ('FullBath', 0.035),	('GrLivArea', 0.309), ('Condition2_PosN', -0.159), ('OverallQual', 0.146), ('RoofMatl_WdShngl', 0.104), ('Neighborhood_NoRidge', 0.066), ('GarageCars', 0.051), ('Neighborhood_NridgHt', 0.047), ('2ndFlrSF', 0.038), ('BsmtFullBath', 0.038), ('MasVnrArea', 0.036),

In ridge most of the top 10 features are just shuffled but still present.

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:

There are two optimal lambdas/alphas: 7.0 for ridge and 0.0001 for lasso.

If I use 7.0 as lambda for both ridge and lasso, the lasso does really poorly as shown below. And if we use 0.0001 for both, the ridge performance drops resulting in slight overfitting but it is still good as shown below.

Thus, we choose lambda as 0.0001 for both.

Alpha/lambda=7.0	Alpha/lambda=0.0001
ALPHA/LAMBDA:7.0 ------RIDGE r2_train 0.8742309613934397 adj_r2_train 0.841035415887619 mse_train 0.001516035643145854 r2_test 0.8596493718442163 mse_test 0.0017421844541851948 ------LASSO r2_train 0.0 adj_r2_train -0.2639405204460967 mse_train 0.012054124448612707 r2_test -0.00016684167717184728 mse_test 0.012415157281856942	ALPHA/LAMBDA:0.0001 ------RIDGE r2_train 0.9365241689905095 adj_r2_train 0.9197703251181161 mse_train 0.000765145566467508 r2_test 0.8204366681523325 mse_test 0.002228935127668056 ------LASSO r2_train 0.8929271919300361 adj_r2_train 0.8646663392424249 mse_train 0.001290668953537767 r2_test 0.853101802913059 mse_test 0.0018234599921322474

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:

After removing the 5 most important predictors, we run grid search for best alpha and it comes same as before 0.0001. With this we build a new model and below are its top 5 predictors along with their coefficients.

```
('1stFlrSF', 0.279),
('2ndFlrSF', 0.198),
('Neighborhood_NoRidge', 0.071),
('GarageCars', 0.066),
('Neighborhood_NridgHt', 0.058),
```

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:

To make a model more generalizable:

1. Make it with as smaller number of features as possible to reduce its complexity.
    a. It can be achieved with VIF and p-score to reduce the number of features
    b. RFE to reduce the number of features
    c. Check r2_score and adjusted_r2_score and they should be close

2. Use Regularization such as Ridge and Lasso to reduce overfitting and make the model learn the general pattern
3. Cross validation can be used to make sure the model is trained well