Subjective Questions

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 500 and for Lasso is 0.01.

If we were to double the alpha values, we will force more regularization onto the data. Which means, more and more reduction will happen to the coefficients' values and model will get simpler. However, that may lead to underfitting and reduce the power of the model to explain the variance in the test data.

Below are the summarized metrics from the original models –

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.560146e-01	0.894869	0.906853
1	R2 Score (Test)	-7.803589e+25	0.861109	0.843343
2	RSS (Train)	4.490909e+01	107.338832	95.103238
3	RSS (Test)	3.519676e+28	62.644443	70.657395
4	RMSE (Train)	2.097270e-01	0.324239	0.305200
5	RMSE (Test)	8.964257e+12	0.378185	0.401644

Below are the summarized metrics from the new models (with double alphas) –

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.560146e-01	0.871748	0.886103
1	R2 Score (Test)	-7.803589e+25	0.850843	0.838385
2	RSS (Train)	4.490909e+01	130.944937	116.288924
3	RSS (Test)	3.519676e+28	67.274888	72.893779
4	RMSE (Train)	2.097270e-01	0.358122	0.337486
5	RMSE (Test)	8.964257e+12	0.391913	0.407951

We can see that the R2 scores have fallen for both train & test for both Ridge & Lasso models. The errors have gone up as the model is now underfitting.

The 5-most important predictor variables for Ridge (new model) are –

	Ridge
GrLivArea	0.083817
OverallQual_10	0.074443
OverallQual_9	0.070602
1stFlrSF	0.061885
Neighborhood_NoRidge	0.060966

The 5-most important predictor variables for Lasso (new model) are –

	Lasso
GrLivArea	0.366185
OverallQual_9	0.163350
OverallQual_10	0.158847
OverallQual_8	0.127686
GarageCars	0.094736

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:

Both Ridge and Lasso are performing good and are pretty comparable in terms of Test R2, RSS, and RMSE values.

However, since Lasso has given a much simpler model by reducing a large number of coefficients to zero, we will select this as the final model.

Among the various values of alpha, we will stick to the optimal values we got from the original grid search method (alpha = 0.01 for Lasso).

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:

These were the top-five predictors in the original Lasso model –

	Lasso
GrLivArea	0.375302
OverallQual_10	0.162418
OverallQual_9	0.158653
OverallQual_8	0.133709
RoofMatl_WdShngl	0.083598

After dropping them, we fit a new model which gave these features as top-five predictors (see the jupyter notebook for the code).

	Lasso
2ndFlrSF	0.309217
1stFlrSF	0.267655
Neighborhood_NoRidge	0.103489
Neighborhood_NridgHt	0.089982
GarageCars	0.089755

Also, the model has not really suffered and has only slightly reduced performance metrics -

```
R2 on train dataset =  0.8932386152727777
R2 on test dataset =  0.8342541890042399
RSS on train dataset =  109.00337380649398
RSS on test dataset =  74.75682444473966
RMSE on train dataset =  0.3267436070181363
RMSE on test dataset =  0.41313155456910355
```

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:

We make a model robust and generalizable by making it simple. The golden rule of model selection is the Occam's Razor which says to make a model as simple as we can but no simpler. Simple models are more versatile and applicable to more general new data.

In the process of making it simple, we sacrifice some accuracy. This is because of the bias-variance trade-off. When we are making the model simpler, we are essentially reducing its variance. We try to reduce a lot of variance by taking on some extra bias (hence, reduced accuracy).