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?

By hyperparameter tuning for alpha using grid search and cross validation, we found the following optimum alpha values.

Ridge regression optimum alpha : **0.1**

Lasso regression optimum alpha : **0.0001**

Effect of doubling alpha values

i) Ridge model

Model with original alpha values

Ridge Test R2-Score : 0.81

Mean square error : 0.00754

Model with double of alpha value

Ridge Test R2-Score 0.812

Mean square error : 0.00744

ii) Lasso model

Model with original alpha values

Lasso Test R2-Score : 0.82

Mean square error: 0.00716

Model with double of alpha value

Lasso Test R2-Score : 0.825

Mean square error : 0.00695

We found that R2 score remained almost the same when alpha was doubled for both ridge and lasso models. Also there is a small decrease in mean square error in both as alpha is doubled.

Ideally we see that as regularisation increases, bias is compromised for lower variance. Here since the alpha values were small(0.1 and 0.0001) hence doubling did not have much effect on the model.

Now the top 5 predictors before and after doubling alpha values for both ridge and lasso models are given below.

Ridge model

Predictors with alpha 0.1 and their coefficients	**Predictors with alpha 0.2 and their coefficients**
GrLivArea : 0.616300	GrLivArea :0.600293
OverallQual : 0.377091	OverallQual : 0.376115
LotShape_IR3 : -0.242975	LotShape_IR3 : -0.234518
LotArea : 0.229578	LotArea : 0.221179
GarageCars : 0.214225	GarageCars : 0.213538

Lasso model

Predictors with alpha 0.0001 and their coefficients	**Predictors with alpha 0.0002 and their coefficients**
GrLivArea : 0.619604	GrLivArea : 0.602335
OverallQual : 0.381542	OverallQual : 0.379474
LotShape_IR3 : -0.230938	LotShape_IR3 : -0.208333
GarageCars : 0.202804	GarageCars : 0.194200
LotArea : 0.184071	LotArea : 0.133809

We see that though the model coefficients have changed slightly, the top predictors remain same even after doubling alpha value for both lasso and ridge models. This may be due to small alpha values and hence doubling did not have much effect on predictors.

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?

The R2 score and mean square error of the LR model and ridge and lasso model with optimal alpha values are as shown below.

Model	**Adjusted R2 train**	**Adjusted R2 test**	**Mean square error**
LR model	0.835193	0.780388	0.007659
Ridge model	0.831001	0.783088	0.007545
Lasso model	0.831001	0.794504	0.007160

We see that the R2 scores for test unseen data are slightly better for Lasso model (0.79) compared to normal LR and ridge model. Also the mean square error is slightly lesser.

The Lasso model has additional advantage that it has inbuilt feature selection by regularising some model coefficients to zero and hence is more a simpler model than other two models.

Hence as per Occam Razor principle,when all else remains same we should choose a simpler model over a complex model. Hence the Lasso model is selected.

Question 3

After building the model, you realized 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?

For the lasso model , the five most important predictor variables are:GrLivArea,OverallQual, LotShape_IR3, GarageCars, and LotArea.

Now these variables are dropped and the lasso model is created again. The top five most important predictor variables for the new model are as shown below along with model coefficients.

Predictors	**Coefficients**
Full Bath	0.328894
MSZoning_FV	0.154273
Neighborhood_NoRidge	0.153299
Exterior1st_Stone	0.143980
Neighborhood_StoneBr	0.130568

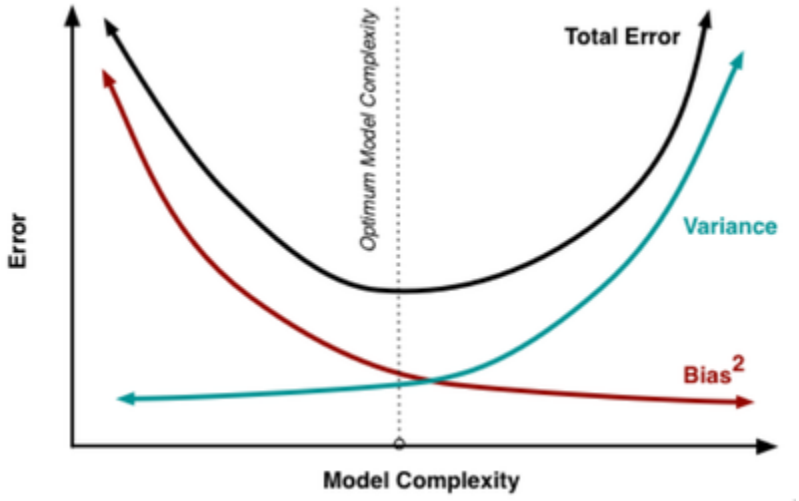
Question 4

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?

A model is robust and generalisable if the model can learn the underlying general pattern in the input data without learning noise as a pattern. That is the model should neither overfit nor underfit. So the model should do well both on train data as well as unseen test data. This is achieved using bias-variance tradeoff.

A very simple model will have high bias and low variance and hence will underfit.
A complex model will have low bias but high variance and hence overfit.
Hence we should find a tradeoff between bias and variance such that both bias and variance are low for the model and total error is lowest.

The bias variance tradeoff is as shown below.



One way to achieve the tradeoff is to use regulariation. Regularisation penalises the higher model coefficients thus making the model simpler. Regularisation compromises on bias for lower variance thus achieving the required bias-variance tradeoff.

Too much regularisation will reduce accuracy of model as bias error wil increase steadily and model become too simple. Also not enough regularisation will reduce bias but as the cost of generalisability. Hence optimum value of alpha should be used in regularisation so that accuracy on train data is high enough with low variance as well.