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?

Ans:

Optimal hyperparameter value (alpha) for Ridge regression = 4.727272

Optimal hyperparameter value (alpha) for Lasso regression = 0.0003418

Usually, if we double the value of alpha, the model penalizes the features more than necessary and so the accuracy(R^2 score) of the models decreases and the same can be observed from the table below.

Scores\Models	Ridge ($\alpha_{optimal}$)	Lasso ($\alpha_{optimal}$)	Ridge(2× $\alpha_{optimal}$)	Lasso(2× $\alpha_{optimal}$)
R^2 score (Train)	0.922864	0.911628	0.911724	0.893258
R^2 score (Test)	0.871653	0.886120	0.861704	0.972871
RSS (Train)	1.321074	1.513499	1.511840	1.828106
RSS (Test)	0.940510	0.834501	1.013413	0.931581
RMSE (Train)	0.035971	0.038502	0.038480	0.042314
RMSE (Test)	0.046339	0.043649	0.048101	0.046118

Note: RSS and RMSE values have increased when the values are doubled, because lower the values of RSS & RMSE means better is the model performance.

After doubling the alpha, the most important **Top-4** predictor variables for both lasso & ridge models are **GrLivArea, OverallQual, OverallCond, LotArea** (In decreasing order of importance)

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?

Ans:

Choosing the regularization technique depends on each scenario. But for the use-case of this assignment, since it has vast features for modelling, it would be better if we opt for model working with less features i.e., Lasso Regression since it eliminates features, making it less prone to overfitting.

Also, we can observe the difference between Train and Test R2 scores for Lasso (0.0255) is least compared to Ridge (0.051). Although the difference is not that much, but it supports our decision of choosing the Lasso Regression, because we not only want the R2 scores of training dataset to be higher, but also Test R2 score to be similar to have confidence that the model might work better for the unseen data as well.

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?

Ans:

The top 5 most important predictor variables in Lasso regression are **GrLivArea, OverallQual, OverallCond, LotArea & GarageCars**.

After creating a new model with excluding the top 5 most important predictor variables above, the five most important predictor variables are

Features (predictor variables)	Beta (Co-efficients)
FullBath	0.119419
GarageArea	0.100814
TotalBsmtSF	0.093516
BedroomAbvGr	0.077671
Neighborhood_Crawfor	0.056297

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?

Ans:

An ideal model should have low bias and low variance.

Bias is the difference between Predicted Value and the Expected value. And the closer their values are low is the bias of the model. For a real-world data, simple models tend to have high bias (underfitting) and complicated models have low bias (overfitting).

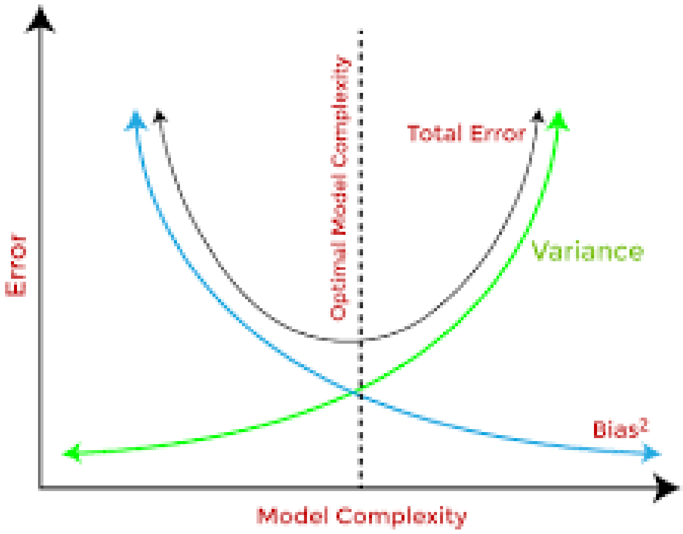
Variance refers to the changes in model when using different portion of training data. For a real-world data, simple models tend to have low variance and complicated models have high variance.

A model is robust and generalised.

Robust means that on changing the data, the model performance should only change minimally. In other terms, **variance** should be **low**.

Generalised means that the model should be able to perform well on the test data and on the unseen data as well. That means, **bias** should be low, so the model shouldn't overfit the data.

Usually, a **simple model** will have **high bias** & **low variance** and as the complexity increases i.e., a more **complex model** will have **low bias** & **high variance**. The same thing is represented in the graph below.



As said earlier, since an ideal model with low bias and low variance is not possible, there should be tradeoff between them so that we get minimal **Total Error**.

This is achieved by multiple methods and one of them is **Regularization**. We penalize the features with large coefficients, and thereby decreasing the dependency of the model on those features and preventing overfitting.

<u>Implications</u>:

To get to a model with minimum Total Error, we shouldn't just concentrate on the training data accuracy. We have to make sure that the selected model should not be overfitting the training dataset. As a result, the we will decrease the complexity of the model in turn a **reduction(decrease)** in the model accuracy.

So, to obtain an optimal model, we compromise on the model accuracy upto a certain extent so that the model doesn't overfit.