

### 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?

#### Optimal Alpha Values are

**Lasso Alpha:** 25.60

**Ridge Alpha:** 568.9

#### Effects of Doubling Alpha on

- **Ridge Regression:**
  - By Increasing alpha in Ridge regression increases the amount of regularization.
- **Lasso Regression:**
  - Similar to Ridge, doubling the alpha will also increase the penalty on coefficients.

#### Changes in Important Variables

- **With Doubled Alpha:**
  - The doubling of alpha seems to have affected the coefficients, likely leading to more features being shrunk or eliminated .
  - **Overall Qual** remains critical variable, while others, like MSSubClass, show negative relationship.

### 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 - Comparison and Decision:

Lasso Regression:

Useful for variable selection, it can shrink some coefficients to zero, which is beneficial if you suspect many features are irrelevant.

Ridge Regression:

Tends to perform better when all predictors are relevant but may not have strong predictive power if you have a large number of features.

If interpretability and feature selection are important, choose Lasso.  
expect many predictors to contribute to the model and want to maintain them all, Ridge will be good.

### 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 –

Five most important variables:

```
PoolArea      104619.863329
PoolQC_unknown 88722.313543
GrLivArea     20115.520153
GarageCond_TA 18093.338788
TotalBsmtSF   10430.485571
dtype: float64
```

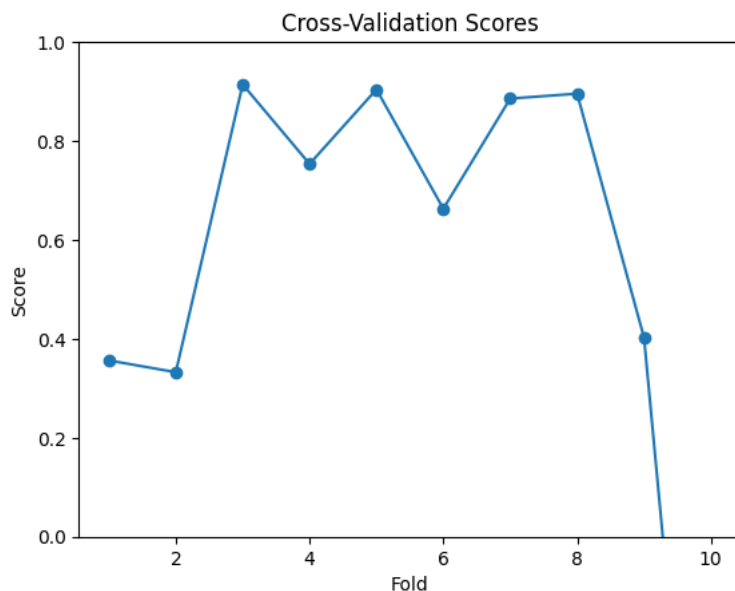
### Question 4

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 –

By using k-fold cross-validation to assess how well the model performs on different subsets of the data. This helps prevent overfitting.

A Plot that validate scores against different values of k to visualize model performance:



A model that is robust and generalizes well tends to have a lower RMSE and a higher  $R^2$  on unseen data compared to a model that might have learned noise in the training data.

A plot comparing training and validation errors for models with different complexities (e.g., by plotting RMSE for varying alpha values) to illustrate overfitting and underfitting:

