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

The optimal value for Ridge regression is **4.0**

The top 5 predictors for Ridge regression are:

- OverallQual (Overall Quality)
- TotalSF (Total Square Feet)
- GrLivArea (Above Grade Living Area Square Feet)
- Neighborhood – Stone Brook
- OverallCond (Overall Condition)

The optimal value for Lasso regression is **50**

The top 5 predictors for Lasso regression are:

- OverallQual (Overall Quality)
- GrLivArea (Above Grade Living Area Square Feet)
- TotalSF (Total Square Feet)
- YearBuilt
- Neighborhood – Stone Brook

CHANGES WITH DOUBLED ALPHA VALUES

The doubled value for Ridge regression is **8**

Values with optimal Ridge regression:

- R2 score (train) = 0.9298796689075287
- RSS (train) = 229209858288.38065
- MSE (train) = 256961724.5385433
- R2 score (test) = 0.894740930363011
- RSS (test) = 144280991439.5193
- MSE (test) = 376712771.38255686

Values with doubled Ridge regression:

- R2 score (train) = 0.9233650522985375
- RSS (train) = 250504885372.33502
- MSE (train) = 280835073.2873711

- R2 score (test) = 0.8923057456401652
- RSS (test) = 147618954309.24887
- MSE (test) = 385428079.13642

So, when doubling alpha for Ridge regression, the R2 score for both train and test data decreases and the RSS and MSE increases

The doubled value for Lasso regression is **100**

Values with optimal Lasso regression:

- R2 score (train) = 0.9283793769633154
- RSS (train) = 234114023721.81015
- MSE (train) = 262459667.84956294
- R2 score (test) = 0.8977117349523894
- RSS (test) = 140208842284.04187
- MSE (test) = 366080528.1567673

Values with doubled Lasso regression:

- R2 score (train) = 0.9215155302783421
- RSS (train) = 256550616668.02945
- MSE (train) = 287612798.95518994
- R2 score (test) = 0.8977674320027893
- RSS (test) = 140132497075.2188
- MSE (test) = 365881193.40788203

So, when doubling alpha for Lasso regression, the R2 score decreases for train data but almost similar for test data. The RSS and MSE increases for train data but almost similar for test data

<u>TOP 5 PREDICTORS AFTER ALPHA IS DOUBLED</u>

The top 5 predictors for Ridge regression now become:

- OverallQual (Overall Quality)
- TotalSF (Total Square Feet)
- GrLivArea (Above Grade Living Area Square Feet)
- 1stFlrSF – First Floor Square Feet
- Neighborhood – Stone Brook

The top 5 predictors for Lasso regression now become:

- OverallQual (Overall Quality)
- GrLivArea (Above Grade Living Area Square Feet)
- TotalSF (Total Square Feet)
- OverallCond (Overall Condition)
- Neighborhood – Stone Brook

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?

Values with original Linear regression:

- R2 score (train) = 0.9191254846509997
- RSS (train) = 264363215539.29633
- MSE (train) = 296371317.86916625
- R2 score (test) = 0.874957466714972
- RSS (test) = 171398633264.50168
- MSE (test) = 447516013.7454352

Values with optimal Ridge regression:

- R2 score (train) = 0.9298796689075287
- RSS (train) = 229209858288.38065
- MSE (train) = 256961724.5385433
- R2 score (test) = 0.894740930363011
- RSS (test) = 144280991439.5193
- MSE (test) = 376712771.38255686

Values with optimal Lasso regression:

- R2 score (train) = 0.9283793769633154
- RSS (train) = 234114023721.81015
- MSE (train) = 262459667.84956294
- R2 score (test) = 0.8977117349523894
- RSS (test) = 140208842284.04187
- MSE (test) = 366080528.1567673

As we can see, both Ridge and Lasso increase the R2 score of train as well as test data. They also decrease the RSS and MSE error for both train and test data.

We'll choose to apply Lasso regression as:

- It reduces errors by slightly more margin than Ridge regression
- Lasso also performs feature selection by making coefficients of some parameters equal to zero

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?

The five most important variables after rebuilding are (calculated in the ipynb file) :

- 2ndFlrSF (2nd Floor Square Feet)
- 1stFlrSF (1st Floor Square Feet)
- OverallCond (Overall Condition)
- TotalBsmtSF (Total Basement Square Feet)
- BsmtFinSF1 (Basement Type 1 Finished Square Feet)

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?

The usefulness and accuracy of a model is decided by how it behaves when client tries it on unknown data. During model building, use Occam's razor. Occam's razor states that, when in dilemma, choose the simpler model.

Simple model means robust and generalisable model. They require fewer training data, make more errors in training data and perform well on unseen data.

The more complex a model becomes, the higher the chance of overfitting.

Overfitting is a phenomenon wherein a model becomes highly specific to the data on which it is trained and fails to generalise to other unseen data points in a larger domain. A model that has become highly specific to a training data set has 'learnt' not only the hidden patterns in the data but also the noise and the inconsistencies in it.

To make a model more simpler, robust and generalisable, we can take care of the following:

- Bias – Variance trade-off
    - bias = how much error the model is likely to make in test data = correctness
    - variance = how sensitive is the model to input data = consistency
    - complex models have a high variance
    - simple models have a high bias
    - strike a balance - model should be simple enough to be generalizable but complex enough to not make too many mistakes
- Regularization
    - helps with managing model complexity by essentially shrinking the model coefficient estimates towards 0
    - discourages the model from becoming too complex
    - with regularization, we compromise by allowing a little bias for a significant reduction in variance