

## Subjective Questions - Assignment

### 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 regression is 10 and for Lasso regression is 0.001.

If we choose to double the value of alpha for both Ridge and Lasso regression, it will increase the regularization strength, resulting in more shrinkage of the coefficients towards zero. This means that the model will become more parsimonious and less complex.

After implementing this change, the most important predictor variables may change depending on the specific dataset used for analysis. However, in general, the most important predictor variables will be those with the highest absolute values of the coefficients in the model.

---

## 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:

Based on the results of the assignment, the optimal value of lambda for Ridge regression was found to be 1.0, while for Lasso regression it was 0.001.

If the goal is to select a model that can predict the house prices accurately and reliably, then Lasso regression would be a better choice. This is because Lasso regression performs both variable selection and regularization, and can remove the redundant and irrelevant features from the model. This results in a more interpretable and simpler model, which is less prone to overfitting, and can be more generalizable to new data.

So, the choice between Ridge and Lasso regression depends on the nature of the problem, the number of variables, and the need for interpretability of the model.

---

### 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:

1. OverallQual
2. GrLivArea
3. TotalBsmtSF
4. 2ndFlrSF
5. GarageCars

To create another model excluding these variables, we can simply remove them from the list of features used in the model and retrain it using the remaining variables. The new set of most important predictor variables will depend on the feature selection method used and the specific data used for training.

---

#### **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:**

1. Cross-validation: This technique involves dividing the data into k-folds and training the model on k-1 folds and testing it on the remaining fold. This is done iteratively k times, with each fold serving as the test set once. Cross-validation helps to ensure that the model is not overfitting to the training data and can generalize well to new data.
2. Regularization: Regularization techniques such as L1 and L2 regularization can help to prevent overfitting and improve the model's ability to generalize to new data.
3. Feature selection: Selecting the most relevant features to include in the model can help to reduce the risk of overfitting and improve the model's generalization ability.
4. Data augmentation: Generating additional training data by augmenting the existing data through techniques such as rotation, scaling, and flipping can help to improve the model's ability to generalize to new data.

-----