

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 value for ridge to minimise RSS and for lasso also minimises RSS by adds penalty which can bring few coefficients to 0. Doubling the Alpha in ridge increases the penalty on large coefficients and in lasso pushes more coefficients to 0 removing less important predictors . In Ridge primary predictors remains same but in lasso it is non zero coefficients after doubling the alpha

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 - Add the screenshot fr0m my own python code attached

```
- Lasso is better than Ridge  
  
- Optimal lambda value :  
- Ridge - 10  
- Lasso - 0.0004  
- Mean Squared error :  
- Ridge - 0.0140819  
- Lasso - 0.0138783
```

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 :

1. GrLivArea,
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. GarageArea

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 : Use crossvalidation, regularisation & various training dataset. This can balance complexity, overfitting and can improve performance. It might reduce training accuracy as the model prioritises generalisation overfitting the training data perfectly.