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?

Values for optimal alpha and corresponding r2 score for ridge and lasso regression is given below:

- Optimal value of alpha of lasso regression is 50 and r2 score for optimal value of alpha is given below
  R2 score for train : 0.9372405328256925
  R2 score for test : 0.9254664123086983

- Optimal value of alpha of ridge regression is 4 and r2 score for optimal value of alpha is given below
  R2 score for train : 0.9371096095852764
  R2 score for test : 0.9253982765709686

- Optimal value of alpha is 1 for ridge regression on variables selected by lasso regression and r2 score for optimal value of alpha is given below
  R2 score for train : 0.9392476999525955
  R2 score for test : 0.9231948226312643

From above r2 score for optimal value of alpha, we can see that the model ridge has slightly lesser than lasso give almost very nearly same r2 score.
Hence, we will choose lasso regression which will do the variable selection as well with r2 score

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?

1. We need to reduce the variance to increase the prediction accuracy.
2. If the requirement is feature selection and we have many variables then we can use Lasso.
3. Lasso reduces the coefficient value to zero as the lambda value increases where lambda is the tuning parameter.
4. If the requirement is to reduce the coefficient and we do not want the large coefficients than we can use Ridge Regression.

Question 3
After building the model, you realized 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?

1. GrLivArea
2. OverallQual_8
3. OverallQual_9
4. Functional_Typ
5. Neighborhood_Crawfor

6. Exterior1st_BrkFace
7. TotalBsmSF

Question 4
How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Per, Occam's Razor— given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons:
Simpler models are usually more 'generic' and are more widely applicable. Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.

- Simpler models are more robust.
    - Complex models tend to change wildly with changes in the training data set
    - Simple models have low variance, high bias and complex models have low bias, high variance.
- Simpler models make more errors in the training set. Complex models lead to overfitting — they work very well for the training samples, fail miserably when applied to other test samples

Therefore, to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use.

Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use. For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model.

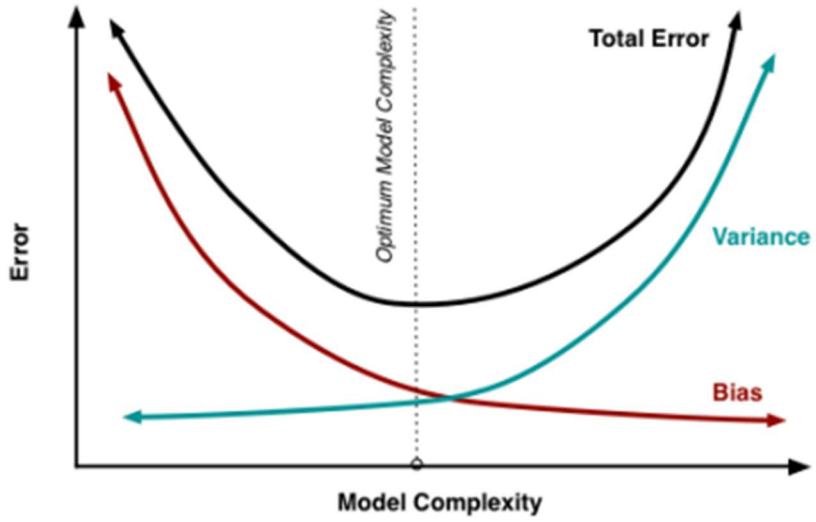
Also, Making a model simple leads to Bias-Variance Trade-off:

- A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.
- A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed.

Bias quantifies how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high.

Variance refers to the degree of changes in the model itself with respect to changes in the training data.

Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph


Total Error
Optimum Model Complexity
Variance
Bias
Error
Model Complexity