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: The optimum alpha was 1.0 for Ridge and 0.001 for Lasso. With these alphas the R Squared value of the model were 0.89 and Lasso's was 0.90.

After doubling the alpha values in the Ridge and Lasso, the prediction accuracy slightly dips in Lasso's to 0.89 also found a small change in the co-efficient values. Reference can be found in Jupyter Notebook.

	Ridge (alpha=2.0)
1stFlrSF	0.509400
OverallQual	0.478703
2ndFlrSF	0.408197
BsmtFinSF1	0.202770
OverallCond	0.198382
Condition2_Norm	0.173957
LotArea	0.141485
GarageQual	0.127308
SaleType_New	0.117817
BsmtUnfSF	0.107042
Neighborhood_Somerst	0.103915
GarageArea	0.098622
CentralAir	0.092084
FullBath	0.087859
BsmtFullBath	0.086846
YearBuilt	0.082525
Neighborhood_NridgHt	0.079859
SaleCondition_Normal	0.078588
Street_Pave	0.076772
HalfBath	0.075156

	Lasso (alpha=0.002)
OverallQual	0.579751
1stFlrSF	0.564732
2ndFlrSF	0.407723
GarageArea	0.146847
OverallCond	0.140086
LotArea	0.137543
BsmtFinSF1	0.119548
CentralAir	0.077985
FireplaceQu	0.077030
MSZoning_RL	0.071671
SaleType_New	0.063288
Neighborhood_Somerst	0.061173
Condition1_Norm	0.057295
BsmtFinType1	0.046991
SaleCondition_Normal	0.044356
BsmtExposure	0.041944
Neighborhood_NridgHt	0.038392
HalfBath	0.033345
BsmtFullBath	0.015368
LotConfig_CulDSac	0.012280

Since the alpha values are small, there were no significant changes seen in the model 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?

▪ The optimum lambda value in case of Ridge and Lasso are:

- Ridge Regression – 1.0
- Lasso Regression – 0.001

▪ The R Squared value in case of Ridge and Lasso are:

- Ridge Regression - 0.8972
- Lasso Regression - 0.9008

▪ The Mean Squared Error in case of Ridge and Lasso are:

- Ridge Regression - 0.1272
- Lasso Regression - 0.1250

▪ The Mean Squared Error of both the models are almost same.

▪ Since Lasso has slightly higher R Squared value and also helps in feature elimination, from the example above we see the coefficient value of some of the features has become zero, Hence Lasso has a better advantage over Ridge and should be used as the final 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?

Ans: The five most important predictor variables in the current lasso model were

1stFlrSF, OverallQual, 2ndFlrSF, OverallCond, LotArea

After dropping these variables, We built the Lasso model again retrieved the R2 of the new model without the top 5 predictors drops to 0.83
The Mean Squared Error increases to 0.1615. The new Top 5 predictors are

	New Lasso Co-Efficient
FullBath	0.467214
BsmtFinSF1	0.462828
BsmtUnfSF	0.290161
GarageArea	0.274288
HalfBath	0.256788

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?

According to Occam's Razor, If given two models that show similar 'behaviour' in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons:-

1. Simpler models are also more robust in nature.
    a. Complex models tend to swing wildly when training data is changed.
    b. Complex models have low bias and high variance whereas Simple models have low variance, high bias.
    c. 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
    d. Therefore, to make the model more robust and generalizable, make the model simple but not so simple which will result in underfitting.
2. Simpler models require very less training samples for effective training than the more complex ones and hence are easier to train.
3. Simpler models are usually more **generic** in nature and are more widely applicable to different types of datasets

Regularization can be used to make the model simpler. Regularization helps to balance between keeping the model simple and not making it too naïve to be of any use. In other terms, it ensures the model doesn't under or overfit. For regression, regularization involves adding a regularization term to the cost function that adds up the absolute values or the squares of the parameters of the model.

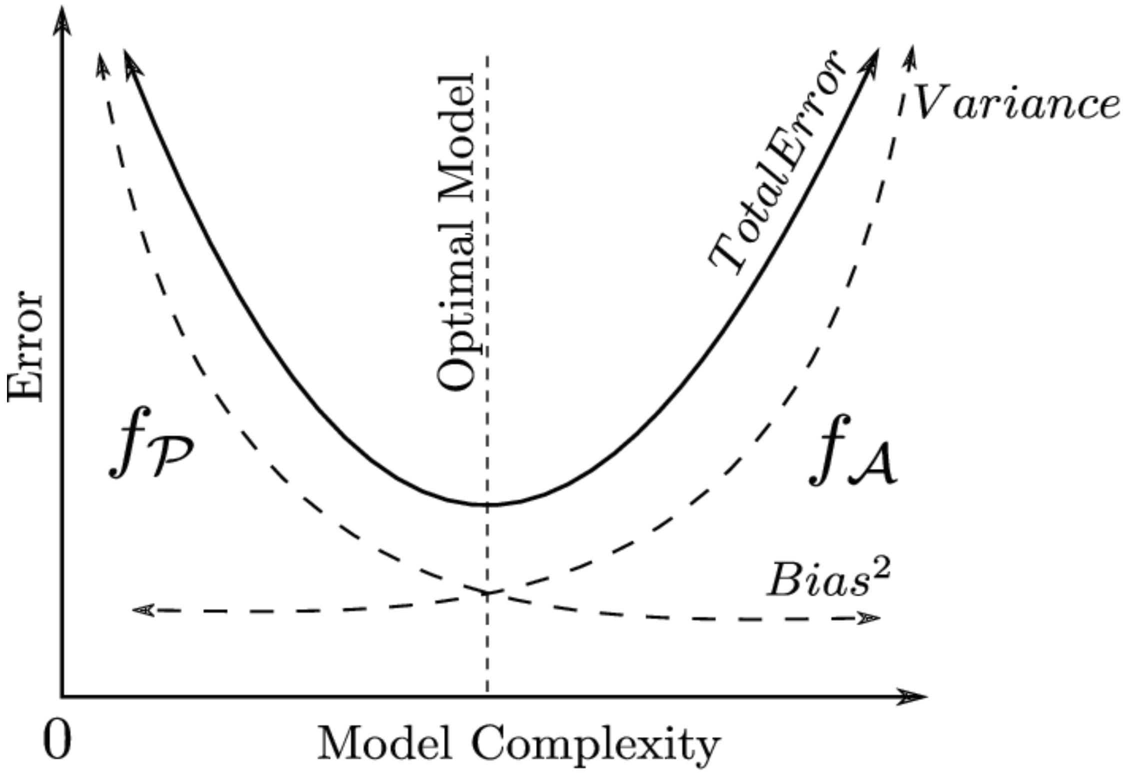
To create a simple model leads to Bias-Variance Trade-off:

- A complex model will need to change for every negligible change made in the dataset and is considered very unstable, unreliable and extremely sensitive to any changes in the training data.
- A simple model on the other hand, brings out some pattern followed by the data points given therefore making it highly unlikely to change 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.



Error
Optimal Model
$Total Error$
$Variance$
$f_{\mathcal{P}}$
$f_{\mathcal{A}}$
$Bias^2$
0
Model Complexity