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?

From the experimentation:

- Optimal value of lambda for Ridge regression is 0.2
- Optimal value of lambda for Lasso regression is 0.0001

Doubling the values of lambda for Ridge and Lasso, we get 0.4 and 0.0002 lambdas respectively. When we double the lambda values, we get the following metrics

RIDGE:

R2-Score (Train set) for lambda 0.2 is = 0.894

R2-Score (Test set) for lambda 0.2 is = 0.876

R2-Score (Train set) for lambda 0.4 is = 0.894

R2-Score (Test set) for lambda 0.4 is = 0.877

LASSO:

R2-Score (Train set) for lambda 0.0001 is = 0.890

R2-Score (Test set) for lambda 0.0001 is = 0.874

R2-Score (Train set) for lambda 0.0002 is = 0.882

R2-Score (Test set) for lambda 0.0002 is = 0.869

When the lambda values are doubled, we can see slight decrease in R2 values (Lasso model), no difference observed in Ridge model.

Variables that are significant after doubling lambda values are:

- Ridge:

Feature	Co-efficient

GrLivArea	0.414788
Exterior1st_BrkComm	-0.173694
TotalBsmtSF	0.132509
Functional_Sev	-0.129541
ExterQual_Fa	-0.113016
MSZoning_FV	0.110684
KitchenAbvGr	-0.10528
Functional_Maj2	-0.104835
MSZoning_RL	0.099128
OverallQual_Very Poor	-0.095711

- Lasso:

Feature	Co-efficient
GrLivArea	0.425322
TotalBsmtSF	0.133118
ExterQual_Fa	-0.112015
KitchenAbvGr	-0.090115
GarageCars	0.086012
BsmtFinSF1	0.078438
Functional_Maj2	-0.072659
LotArea	0.071207
ExterQual_TA	-0.069912
AgeYearBuilt	-0.069085

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?

From the experimentation:

RIDGE:

R2-Score (Train set) for lambda 0.2 is = 0.894

R2-Score (Test set) for lambda 0.2 is = 0.876

LASSO:

R2-Score (Train set) for lambda 0.0001 is = 0.890

R2-Score (Test set) for lambda 0.0001 is = 0.874

The R2 Score of Ridge regression is observed to be slightly greater than that of Lasso regression.

Ridge regression shrinks the coefficients towards 0 but not exactly 0, which implies that all the 30 (RFE) features are included in the Ridge model. When the dataset has a greater number of features, it is tedious and time consuming to use Ridge regression as it includes all the features in the model. As the number of features increase, the model becomes too complex. As the complexity increases, the variance also increases but the bias compromise is reduced.

On the other hand, Lasso regression pushes some of the coefficients to exactly 0, which implies it does feature selection. Here, with the optimal value of lambda 0.0001, Lasso regression selected 21 significant features while the Ridge regression contains all the (30) features. As the number of features are mitigated, the model is not as complex as Ridge and the variance is also shrinked with reasonable bias compromise.

Hence, better to use Lasso regression as it does have feature selection and thereby reducing the complexity 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?

The 5 top-most variables that are significant for lambda 0.0001 in Lasso regression are

- GrLivArea (0.427307)
- Exterior1st_BrkComm (-0.142601)

- TotalBsmtSF (0.132262)
- ExterQual_Fa (-0.114256)
- KitchenAbvGr (-0.100688)

Suppose these predictor variables are not present in the incoming data. Now, we will create another model using Lasso regression with lambda 0.0001 and predict the new significant features.

- BsmtFinSF1 (0.341147)
- BsmtUnfSF (0.283386)
- 2ndFlrSF (0.195174)
- BsmtFinSF2 (0.194476)
- GarageCars (0.130873)

Slight reduction in R2 values observed:

LASSO (without feature removal):

R2-Score (Train set) for lambda 0.0001 is = 0.890

R2-Score (Test set) for lambda 0.0001 is = 0.874

LASSO (top-5 feature removal):

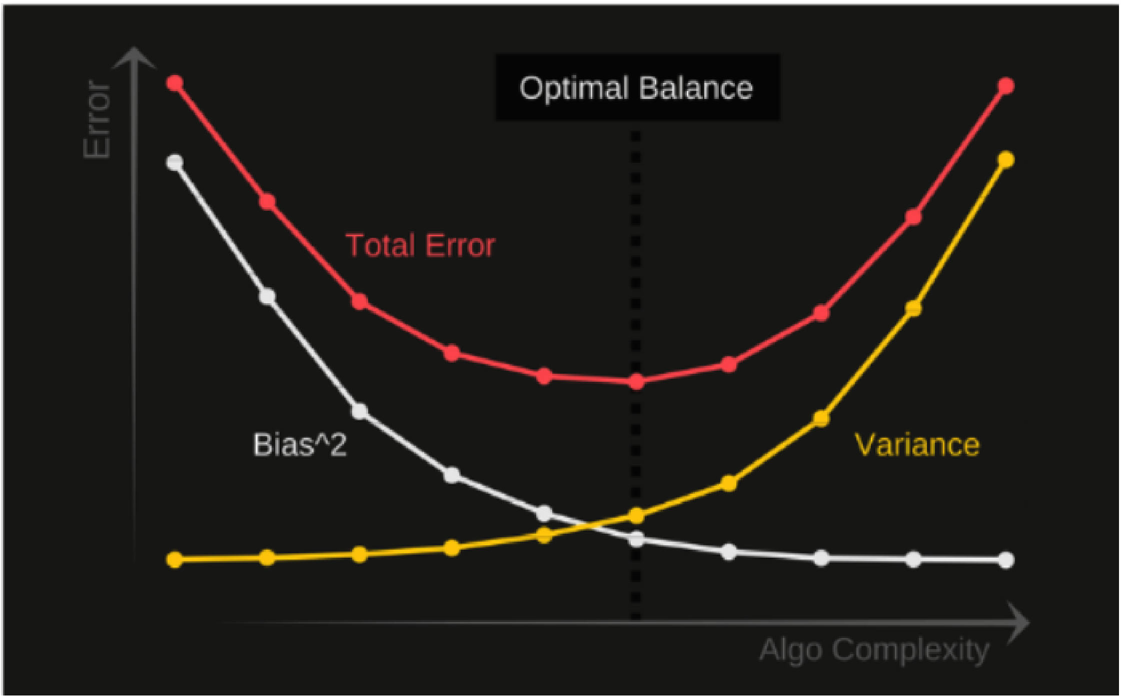
R2-Score (Train set) for lambda 0.0001 is = 0.862

R2-Score (Test set) for lambda 0.0001 is = 0.859

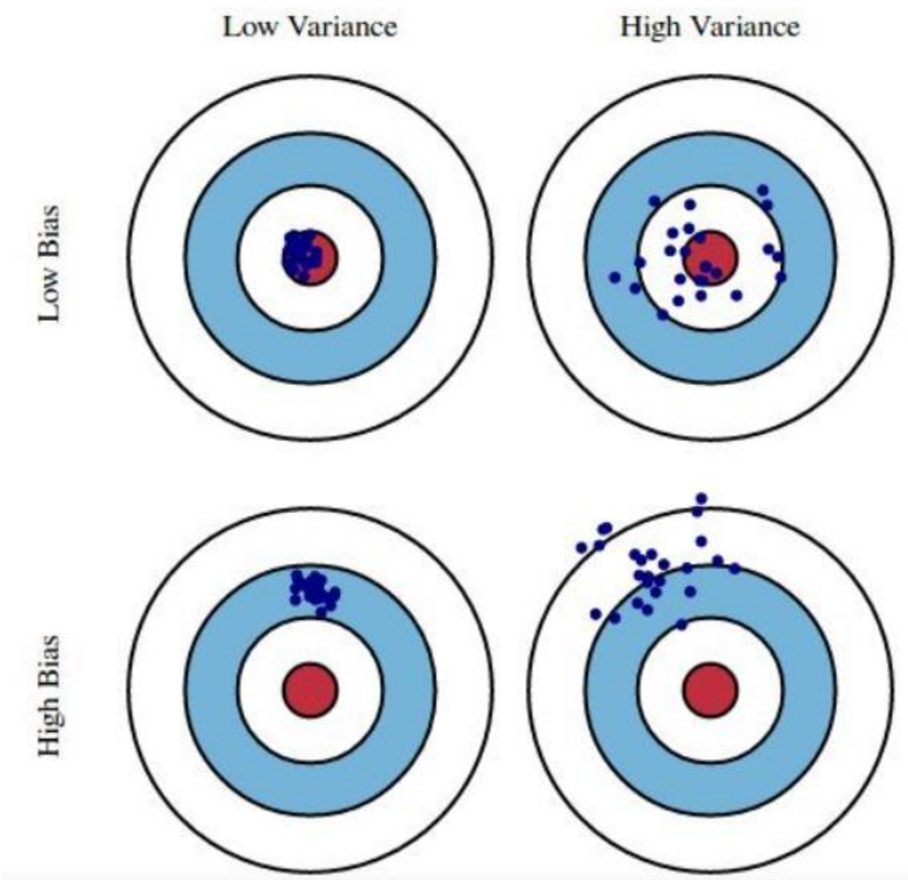
Question 4

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?

When the model is simple, the variance is low, but the bias is high. If the model works well on the train set as well as the test set, it is said to be robust (i.e., it is neither underfitted nor overfitted). Simple models can be easily generalisable though the bias is reasonably compromised. A model should be as simple as necessary but not simpler than that. There is always a trade-off between bias and variance



Optimal Balance
Error
Total Error
Bias^2
Variance
Algo Complexity



Low Variance
High Variance
Low Bias
High Bias

We should use regularisation to make the model simpler by choosing optimal model that will neither overfit nor underfit. It penalises the model parameters so that it won’t overfit. And, Lasso, a type of regularisation does feature selection by making the non-significant features 0. We can use residual analysis to check whether our model is actually performing well or there are some patterns we missed. We can use regularisation along with proper EDA to make the model simpler because simpler models are more robust and can perform well on unseen data giving less errors on test data making it more generalisable. And when the model gives the less errors on unseen data, it impacts the accuracy positively. The accuracy of robust and generalisable model is always more than any model that is not generalisable.