Question 1:

The optimal values of alpha for

- Ridge Regression: 0.0001
- Lasso Regression: 0.0001

With these values, the most important predictor variables are

S.No	Ridge Regression		Lasso Regression	
	Feature	Coefficient	Feature	Coefficient
1	OverallQual_10	0.262889	OverallQual_10	0.250867
2	LotArea	0.249444	LotArea	0.193845
3	OverallQual_9	0.193882	OverallQual_9	0.189699
4	BsmtFinSF1	0.096038	BsmtFinSF1	0.098465
5	OverallQual_8	0.093552	OverallQual_8	0.092219
6	Neighborhood_NoRidge	0.090144	Neighborhood_NoRidge	0.087443
7	2ndFlrSF	0.080531	2ndFlrSF	0.080587
8	YearBuilt_(2000, 2010]	0.077985	YearBuilt_(2000, 2010]	0.072682
9	PoolArea	0.074868	Fireplaces	0.058555
10	Neighborhood_StoneBr	0.062024	BsmtFinSF2	0.056027

Top 8 variables are same with different values but the 9^{th} and 10^{th} places in Ridge regression are Pool Area and Stone_br neighbourhood, where as in Lasso, the 9^{th} and 10^{th} places are Fireplaces and BsmtFinSF2.

When we doubled the value of alpha, i.e., with alpha values as below, the important predictor variables are

- Ridge Regression: 0.0002
- Lasso Regression: 0.0002.

S.No	Ridge Regression		Lasso Regression	
	Feature	Coefficient	Feature	Coefficient
1	OverallQual_10	0.262885	OverallQual_10	0.238826
2	LotArea	0.249429	LotArea	0.185501
3	OverallQual_9	0.193880	OverallQual_9	0.138238
4	BsmtFinSF1	0.096039	BsmtFinSF1	0.100885
5	OverallQual_8	0.093551	OverallQual_8	0.090874
6	Neighborhood_NoRidge	0.090144	Neighborhood_NoRidge	0.084743
7	2ndFlrSF	0.080531	2ndFlrSF	0.080649
8	YearBuilt_(2000, 2010]	0.077985	YearBuilt_(2000, 2010]	0.067403
9	PoolArea	0.074866	PoolArea	0.062909
10	Neighborhood_StoneBr	0.062024	Neighborhood_StoneBr	0.050526

With alpha values doubles, there is no change in top 10 parameters in both ridge and lasso regressions, but only the values of coefficients became smaller.

Theory:
When the lambda value is doubled, it means it is moving towards higher values, so, the shrinkage penalty increases, pushing the coefficients further towards 0, which may lead to either

- Model underfitting or
- Coefficients of more predictions are moved to zero, thereby reducing the complexity of the model.

Here in our case, the model complexity is decreased as the coefficients are pushed towards zero.

Question 2:

Ridge regression:

- It enforces the β coefficients to be lower, but it does not enforce them to be zero. That is, it will not get rid of irrelevant features but rather minimize their impact on the trained model.

$$L = \sum(\hat{Y}i - Yi)^2 + \lambda\sum \beta^2$$

- As the value of lambda is increased, the value of the coefficient tend towards zero. This leads to both low variance and low bias, but as stated earlier, this model may not be good for feature reduction.

Lasso regression:

- It is like Ridge regression, but the only difference is that the regularization term is in absolute value.
- This difference makes a huge impact on the trade-off as it makes Lasso method overcome the disadvantage of Ridge regression by not only punishing high values of the coefficients β but actually setting them to zero if they are not relevant. Therefore, we might end up with fewer features, thereby reducing the complexity of the model.

$$L = \sum(\hat{Y}i - Yi)^2 + \lambda\sum |\beta|$$

- But, if the model is already a better model and all the independent variables are closely related to dependent variables, then the Lasso regression may reduce the r2 value by unnecessarily penalizing the coefficients to move closer to zero or by setting them to zero.

I have chosen Lasso regression as my model is a complex model with close to 40 parameters. Lasso can further drive the coefficients slightly towards zero thereby reducing the complexity. Even the results also prove this as every metric in test data, i.e., r-square, RSS, MSE and RMSE are better for Lasso regression than Ridge regression.

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.859441	0.859441	0.857970
1	R2 Score (Test)	0.716086	0.716090	0.729651
2	RSS (Train)	1.813255	1.813255	1.832236
3	RSS (Test)	1.565109	1.565092	1.490333
4	MSE (Train)	0.001774	0.001774	0.001793
5	MSE (Test)	0.003573	0.003573	0.003403
6	RMSE (Train)	0.042122	0.042122	0.042341
7	RMSE (Test)	0.059777	0.059777	0.058332

Question 3:

Before deleting top 5 variables, the top predictors are

Lasso Regression	
Feature	**Coefficient**
OverallQual_10	0.238826
LotArea	0.185501
OverallQual_9	0.138238
BsmtFinSF1	0.100885
OverallQual_8	0.090874
Neighborhood_NoRidge	0.084743
2ndFlrSF	0.080649
YearBuilt_(2000, 2010]	0.067403
PoolArea	0.062909
Neighborhood_StoneBr	0.050526

The r-square value for training and test data are 0.857 and 0.729.

After deleting top 5 variables, the r-square value dropped significantly, and those values training and test data are 0.758 and 0.666.

The top variables are

Lasso Regression	
Feature	**Coefficient**
Neighborhood_NoRidge	0.144265
YearBuilt_(2000, 2010]	0.124805
Fireplaces	0.121867
Neighborhood_StoneBr	0.121382
PoolArea	0.118539
YearBuilt_(1980, 2000]	0.088403
Neighborhood_NridgHt	0.084488
2ndFlrSF	0.075537
OverallCond_9	0.064988
KitchenQual_others	0.055346

Another observation is that the sixth to tenth placed variables didn't take-up the first to fifth place after removing the top five variables. The ranking among the remaining variables is again changed as the values of coefficients is changed.

Question 4:

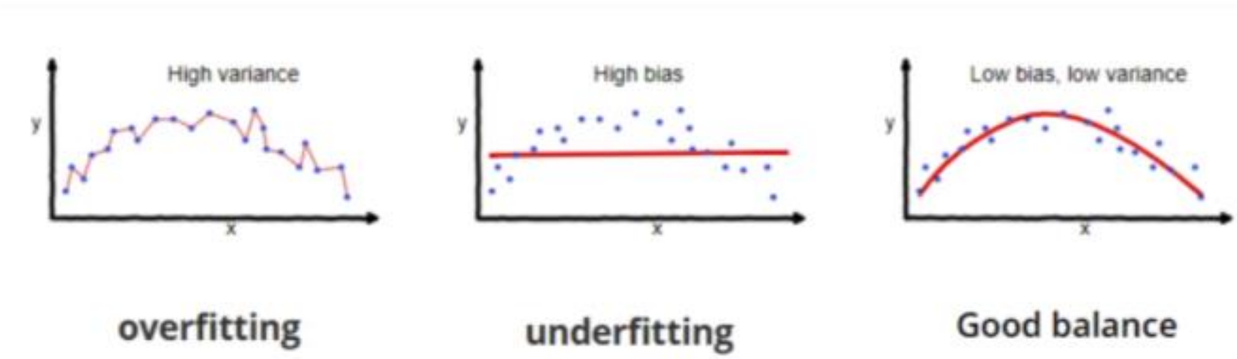
A model is robust and generalizable if

- R-square value in both Test and train data is same or at least with a least difference.

But the r2 value is impacted in Test data because of many reasons:

1. Non-constant variance
2. Autocorrelation
3. Multicollinearity
4. Overfitting
5. Extrapolation

Also, a model is said to be robust when it is in good balance, i.e., it is rather overfitting or under fitting.



Reasons for Overfitting:

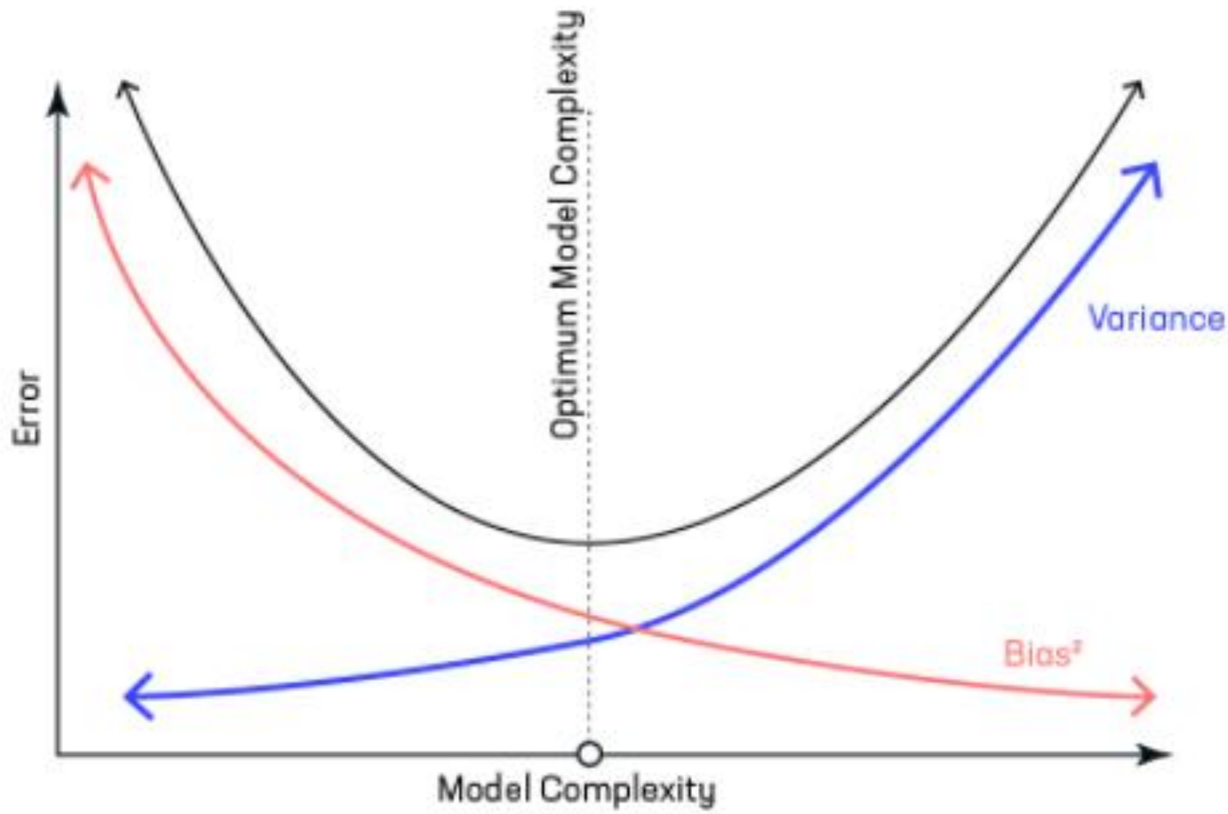
- When a model is too complex, it may lead to overfitting. It means the model may produce good training results but would fail to perform well on the test data. One possible solution for overfitting is to increase the amount and diversity of the training data. Another solution is regularization, which we will cover in the next session.
- In overfitting, training data can be properly predicted and hence usually a high r-square value in training data but not in test-data.

Reasons for Underfitting:

- When a model is not able to predict the values for training data itself, then it is underfitting. It means the chosen variables are not able to properly determine the values in training data.

Finally, a model is robust when it has low variance and low bias, but this is difficult as these two parameters bias and variance are inversely proportional.

So, we need to find an optimal spot where both variance and bias are low



In our case, the model seems to be over fitting as there is a considerable difference between the r-square values of Test and Train data.