Question 1:

The optimal values of alpha for

- Ridge Regression: 0.3
- Lasso Regression: 0.0001

With these values, the most important predictor variables are

S.No	Ridge Regression		Lasso Regression	
	Feature	Coefficient	Feature	Coefficient
1	TotalBsmtSF	0.177357	TotalBsmtSF	0.195026
2	OverallQual_9	0.163498	OverallQual_9	0.127218
3	OverallQual_10	0.157866	OverallQual_10	0.112882
4	OverallQual_8	0.113734	GarageArea	0.100489
5	GarageArea	0.103064	TotRmsAbvGrd	0.097226
6	TotRmsAbvGrd	0.101761	2ndFlrSF	0.093867
7	OverallQual_7	0.09362	OverallQual_8	0.077145
8	2ndFlrSF	0.089511	YearBuilt_(2000, 2010]	0.073054
9	LotArea	0.081781	OverallQual_7	0.056326
10	YearBuilt_(2000, 2010]	0.078381	FullBath	0.05318

There are 9 common variables between top 10 variables of Lasso and Ridge regression.
Top 3 in Lasso are also top 3 in Ridge. Remaining variables have different ranks with different coefficients.
'LotArea' in Ridge is not in top 10 of Lasso and 'FullBath' in Lasso is not in top 10 of Ridge regression.

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

- Ridge Regression: 0.6
- Lasso Regression: 0.0002.

S.No	Ridge Regression		Lasso Regression	
	Feature	Coefficient	Feature	Coefficient
1	TotalBsmtSF	0.174391	TotalBsmtSF	0.203274
2	OverallQual_9	0.148596	OverallQual_9	0.104202
3	OverallQual_10	0.139398	GarageArea	0.099097
4	TotRmsAbvGrd	0.102624	2ndFlrSF	0.095236
5	GarageArea	0.100974	TotRmsAbvGrd	0.093702
6	OverallQual_8	0.100813	OverallQual_10	0.07792
7	2ndFlrSF	0.088602	YearBuilt_(2000, 2010]	0.06618
8	OverallQual_7	0.081182	OverallQual_8	0.056933
9	YearBuilt_(2000, 2010]	0.078199	FullBath	0.052747

10	LotArea	0.071225	OverallCond_9	0.043458

With alpha values doubled, below changes are observed:

- Ridge regression:
    - Coefficients of the variables decreased.
    - No change in the top 10 variables but their rankings are modified.
- Lasso regression:
    - Coefficients of the variables decreased.
    - 'OverallQual_7' is not listed after doubling the lambda. Instead, 'OverallCond_9' came into top 10 variables.
    - Also, a change in the rankings of the top-10 variables is observed.

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 tends 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 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)	0.918828	0.918558	0.915599
1	R2 Score (Test)	0.803692	0.804615	0.802487
2	RSS (Train)	1.455724	1.460564	1.513623
3	RSS (Test)	1.629296	1.621635	1.639303
4	MSE (Train)	0.001424	0.001429	0.001481
5	MSE (Test)	0.003720	0.003702	0.003743
6	RMSE (Train)	0.037741	0.037804	0.038484
7	RMSE (Test)	0.060991	0.060847	0.061178

Question 3:

Before deleting top 5 variables, the top predictors are

Lasso Regression	
Feature	**Coefficient**
TotalBsmtSF	0.195026
OverallQual_9	0.127218
OverallQual_10	0.112882
GarageArea	0.100489
TotRmsAbvGrd	0.097226
2ndFlrSF	0.093867
OverallQual_8	0.077145
YearBuilt_(2000, 2010]	0.073054
OverallQual_7	0.056326
FullBath	0.05318

The r-square value for training and test data are 0.915 and 0.802.

After deleting top 5 variables, the r-square value dropped significantly, and those values training and test data are 0.866 and 0.774.

The top variables after deletion are

Lasso Regression	
Feature	**Coefficient**
LotArea	0.146406
FullBath	0.125854
BsmtFinSF1	0.102067
2ndFlrSF	0.094559
YearBuilt_(2000, 2010]	0.087245
Neighborhood_NoRidge	0.06917
Neighborhood_StoneBr	0.065418
Neighborhood_NridgHt	0.061821
OpenPorchSF	0.059846
RoofMatl_others	0.056563

Another observation is that the ranking of the sixth to tenth placed variables didn't happen to be first to fifth place respectively after removing the top five variables.
Three variables, 'FullBath', '2ndFlrSF' and 'YearBuilt_(2000, 2010]' which were in 6-10 place 1-5 place.

Question 4:

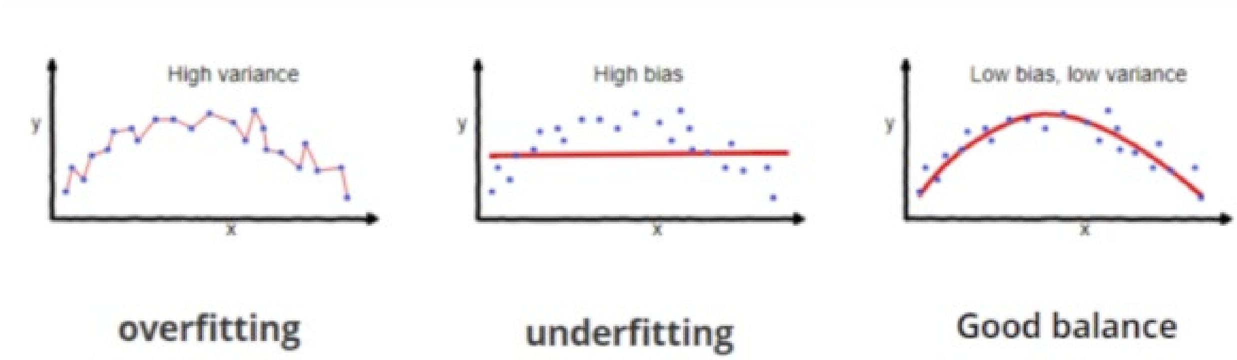
A model is robust and generalizable if

- 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:

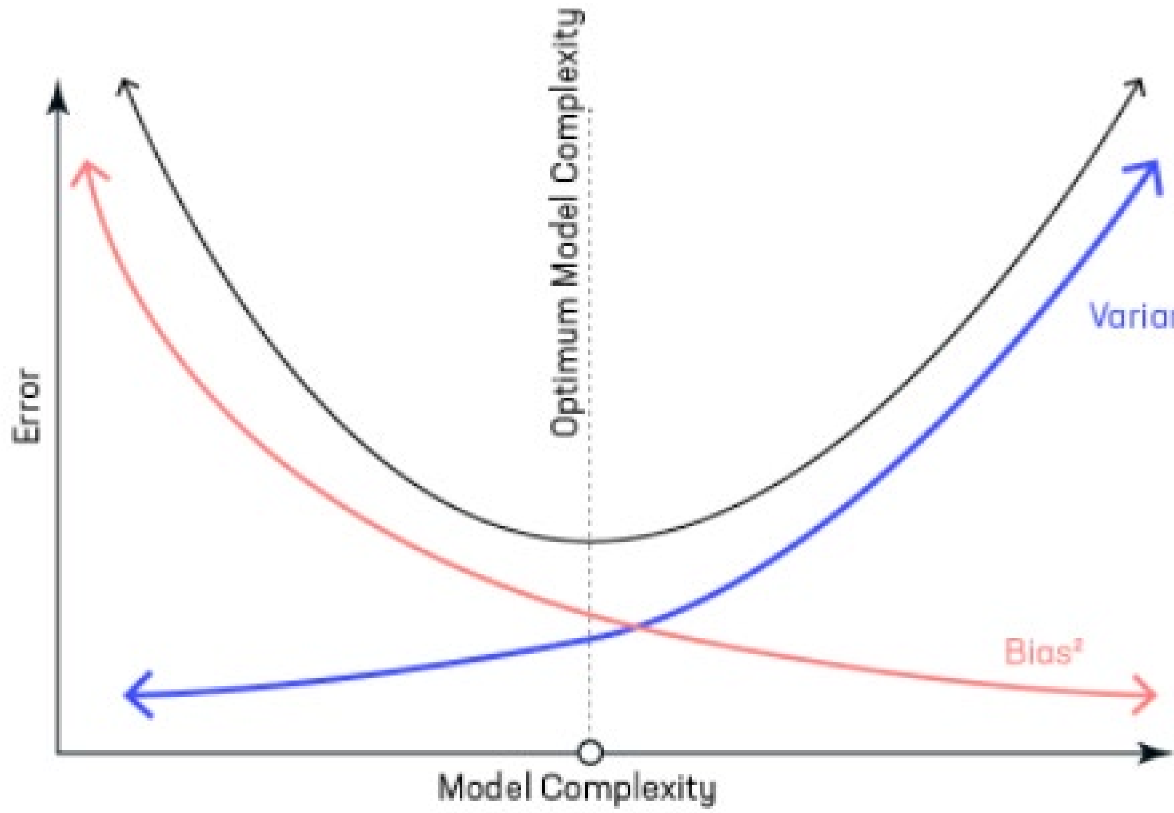
- When a model is too complex, it may lead to overfitting. It means the model may produce good training results but would fail to perform well on the test data. One possible solution for overfitting is to increase the amount and diversity of the training data. Another solution is regularization, which 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.