1. 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?

a) Optimal value of alpha/lambda for Ridge and Lasso Regularization?

The optimal lambda values for Lasso and Ridge are:

- Ridge: 2
- Lasso: 0.0001

b) Changes to the model when lambda values are doubled?

Model metrics with optimal lambda value are

Key	R2-Training	R2-Test	Adj-R2-Test	RSS-Train	RSS-Test	RMSE-Train	RMSE-Test
Ridge	0.9043	0.8740	0.861754	2.2364	1.2979	0.0473	0.0549
Lasso	0.9039	0.8778	0.868282	2.2470	1.2592	0.0474	0.0541

Model metrics with lambda value double the optimal value are

Key	R2-Training	R2-Test	Adj-R2-Test	RSS-Train	RSS-Test	RMSE-Train	RMSE-Test
Ridge	0.9012	0.8738	0.861535	2.3105	1.3002	0.0481	0.0550
Lasso	0.8987	0.8778	0.870559	2.3678	1.2590	0.0487	0.0541

- For Ridge, the R^2 values decrease and the RMSE and RSS values go up for the doubled optimal lambda value.
- For Lasso as well, the R^2 values decrease and the RSS and RMSE values go up for the doubled optimal lambda value. Only the RMSE for the test data set remains unchanged.

This indicates that the optimal lambda value is a much better choice for the model regularization.

c) What will be the most important predictor variables after the change is implemented?

The important features with double the optimal lambda values are

```
Top Features for Ridge with doubled alpha:
              Feature  Coefficient
0         OverallQual       0.1611
1           GrLivArea       0.1399
2         BuildingAge      -0.0902
3          GarageCars       0.0815
4          MSZoning_FV      0.0621
5          TotalBaths       0.0608
6         TotalBsmtSF       0.0600
7         OverallCond       0.0588
8  Neighborhood_Crawfor     0.0514
9  Neighborhood_StoneBr     0.0486

Top Features for Lasso with doubled alpha:
              Feature  Coefficient
0         OverallQual       0.2299
1           GrLivArea       0.1503
2         BuildingAge      -0.0972
3          GarageCars       0.0830
4         TotalBsmtSF       0.0637
5         OverallCond       0.0627
6          TotalBaths       0.0488
7  Neighborhood_Crawfor     0.0455
8             LotArea       0.0390
9          MSZoning_FV      0.0381
```

- For Ridge, the top 3 features slightly change in terms of their coefficients but their order remains. Total basement square footage rises to the top 10 features. And, there is a difference in the significance(coefficient strength) and order for the remaining top features.
- For Lasso, the top 2 features remain the same, however the coefficient values are different as compared to the model coefficients for the optimal lasso lambda value. Lot area rises to the top 10 features. And, there is a difference in the significance(coefficient strength) and order for the remaining top features.

2. 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?

Key	R2-Training	R2-Test	Adj-R2-Test	RSS-Train	RSS-Test	RMSE-Train	RMSE-Test
Ridge	0.9043	0.8740	0.861754	2.2364	1.2979	0.0473	0.0549
Lasso	0.9039	0.8778	0.868282	2.2470	1.2592	0.0474	0.0541

The optimal lambda values are

- Ridge - 2
- Lasso - 0.0001

Based on the metrics for Ridge and Lasso based optimal models, the test dataset fares better with the Lasso model than the Ridge model. The R^2 and the adjusted R^2 are higher for the Lasso model and the RMSE is lower for the Lasso model. Also, since Lasso removes insignificant features by adjusting their coefficient values to 0 (vs Ridge which only brings the coefficient values of insignificant features to near 0). My choice for this use case would be Lasso regularized regression model.

The top predictors/features for Lasso Optimal regression model are as below.

S.No.	Feature	Coefficient	Interpretation
1.	OverallQual	0.2100	For every unit increase in Overall quality there is a (100* 0.2100) i.e. 21% increase in Sales Price of the house.
2.	GrLivArea	0.1441	For every unit increase in Above grade living area there is a (100* 0.1441) i.e. 14.41% increase in Sales Price of the house.
3.	MSZoning_FV	0.1050	For houses in the Floating Village Residential zone there is a (100* 0.1050) i.e. 10.5% increase in Sales Price of the house.
4.	BuildingAge	-0.1016	For every unit increase in Building Age/Year Built there is a (100* 0.1016) i.e. 10.16% decrease in Sales Price of the house.
5.	Exterior1st_BrkComm	-0.0907	For houses with exterior with brick common there is a (100* 0.0907) i.e. 9.07% decrease in Sales Price of the house.

3. 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?

After removing the top 5 predictors and rebuilding the lasso regression model using the optimal alpha value, the new top 5 predictors/features are

```
Top Features for Lasso with original top 5 features removed:
                Feature  Coefficient
0            TotalBaths       0.1766
1            GarageCars       0.1691
2            Fireplaces       0.1013
3  Neighborhood_StoneBr       0.0899
4         MSSubClass_30      -0.0879
```

S.No.	Feature	Coefficient	Interpretation
1.	TotalBaths	0.2100	For every unit increase in Total Bathrooms there is a (100* 0.1766) i.e. 17.66% increase in Sales Price of the house.
2.	GarageCars	0.1691	For every unit increase in Cars accomodated by the garage there is a (100* 0.1691) i.e. 16.91% increase in Sales Price of the house.
3.	Fireplaces	0.1013	For every unit increase in the fireplaces available there is a (100* 0.1013) i.e. 10.13% increase in Sales Price of the house.
4.	Neighborhood_StoneBr	0.0899	For houses in the Stonebrook neighborhood there is a (100* 0.0899) i.e. 8.99% increase in Sales Price of the house.
5.	MSSubClass_30	-0.0879	For houses of type '1-Story 1945 & older' there is a (100* 0.0879) i.e. 8.79% decrease in Sales Price of the house.

4. Question 4

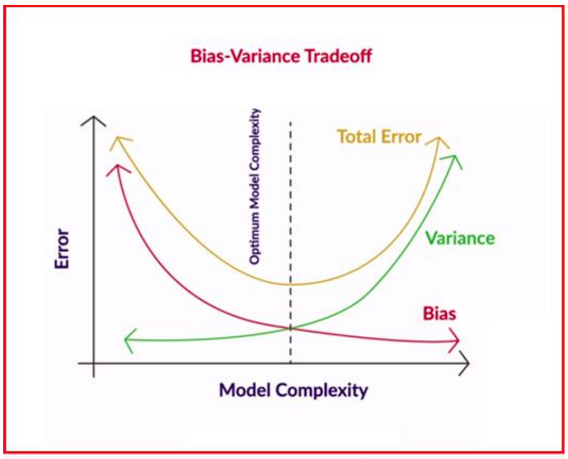
How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Robustness of a model is the capability of the model to handle unseen real data and still perform optimally. Real world data for which the model is deployed and intended for use, could pose several challenges like noisy/erroneous/outlier data and changes in its distribution due to natural/artificial situations. A robust model should be able to handle such cases and maintain its performance. For this to work, the model should be generalized enough that it is able to perform with acceptable accuracy and yet handle the unknown situations to a certain extent.

For a model to be sufficiently robust and generalized, it should not be too simple (under-fit) or too complex (over-fit). This process involves a certain trade off to achieve between the robustness of the model and the accuracy of the model. This is needed because if the model should be interpretable, simple, robust and generalizable, it would make some sacrifice on the model accuracy. This is also referred to as the 'Bias-Variance Trade Off'.

Bias indicates how accurate the model is going to be on the unseen future test data. If the model is under-fit i.e. extremely simple, they might not be able to perform well on the future test data. On the other hand, when the model has over-fit (memorized the training data too much), any small change in training data will have huge impact on the model output. This is called high variance. It is not possible to reduce both bias and variance. If we attempt to build a complex model, the bias becomes low but the variance will increase. On the contrary, if we attempt to build a simple model, the bias becomes high while the variance becomes low.

So, we strive for a trade-off between bias and the variance, so that the model is not too simple or too complex and can perform optimally once it is deployed for the unseen future data.



Bias Variance Tradeoff

There are several strategies available to achieve this trade-off and select an optimally performing robust and generalizable model. Regularization is one such strategy through which L1 (Lasso) or L2(Ridge) regularization can be applied. Ridge and Lasso regularization help in simplifying the model by preventing over-fitting (penalize over-fitting) and also take care of the effect of multicollinearity that might exist among the model features. Regularization especially works well as compared to OLS such that it can reduce the variance without increasing bias a lot. Lasso regularization also helps with feature selections by bringing the coefficient values of insignificant features to 0 (whereas ridge just brings these coefficient values near 0).

Formula for Ridge - L2 regularization:

$\sum [y_i — (\beta_0 + \beta_1 * x_{1i} + \beta_2 * x_{2i} + \beta_3 * x_{3i} \ldots)]^2 + \sum \lambda|\beta_i|^2$
(where i is between 1 and n)

Formula for Lasso - L1 regularization:

$\sum [y_i — (\beta_0 + \beta_1 * x_{1i} + \beta_2 * x_{2i} + \beta_3 * x_{3i} \ldots)]^2 + \sum \lambda|\beta_i|$
(where i is between 1 and n)