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?

Optimal Value of Alpha

Model	Alpha
Ridge	10
Lasso	0.001

After doubling value of Alpha

- Respective code changes available in the same Jupiter note book shared.
- Discussing the outcome here.

	Ridge (10)	Ridge (20) - New	Lasso (0.001)	Lasso (0.002) – New
R2 score (train)	0.9165	0.9163	0.9156	0.9145
R2 score (test)	0.8727	0.8736	0.8763	0.8775
RMSE (train)	0.1131	0.1133	0.1138	0.1145
RMSE (test)	0.1525	0.152	0.1504	0.1496

- Old ridge model has better R2 Score in the train set then the new (doubled alpha) ridge model but on the test-data set new ridge model has better R2 value.
- Same observation found while comparing the new (doubled alpha) and old lasso model where R2 Score looks better for Lasso with Alpha=0.001 in train data-set but in the test-data set R2 score of the Lasso with Alpha=0.0002 looks better.
- Very similar observation found for the RMSE error comparison for both Lasso and Ridge between previous model and new model (doubled alpha) where the for the train-data set RMSE looks better for old model (both Lasso and Ridge) and for the test-data set the number looks better for new model (doubled Alpha).

- Comparing **new Ridge model against the new Lasso Model** with the highlighted values from the about matrix still Lasso is remains as the better model compare to the new ridge model (with better R2 Score/Accuracy and Low Error).

Top 10 Features in the new Lasso (Alpha = 0.002) Model

1. 1stFlrSF
2. 2ndFlrSF
3. OverallQual
4. OverallCond
5. LotArea
6. SaleCondition_Partial
7. BsmtFinSF1
8. SaleCondition_Normal
9. BsmtQual
10. MSZoning_RL

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?

	Ridge (10)	**Lasso (0.001)**
R2 score (train)	0.9165	0.9156
R2 score (test)	0.8727	0.8763
RMSE (train)	0.1131	0.1138
RMSE (test)	0.1525	0.1504

- Both Ridge and Lasso Regression has almost same train R2 Score (**0.92**) while Lasso has a slightly better test Score (**0.88** vs **0.87**)

- There are **6** common features if we compare the top 10 features from both models.

- Root Mean Square Error for Lasso is slightly lower than ridge hence Lasso is better.

- Some of the coefficient in Lasso model became 0 (eg: - Exterior2nd_HdBoard, MasVnrType_BrkFace etc.) hence it helps to do feature reduction.
    - 4 Features has been removed.
- Apart of this Lasso regularization is robust to outliers, ridge regularization is not.

- With these comparisons we can conclude that the Lasso model is better than Ridge model hence we go with the Lasso Model.

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?

- Code steps and explanations are provided in the Jupiter notebook.

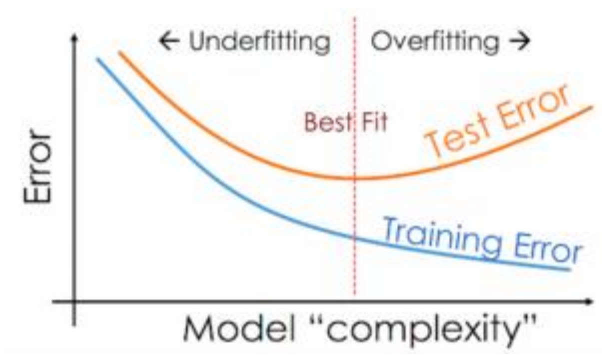
New top 5 most important variables

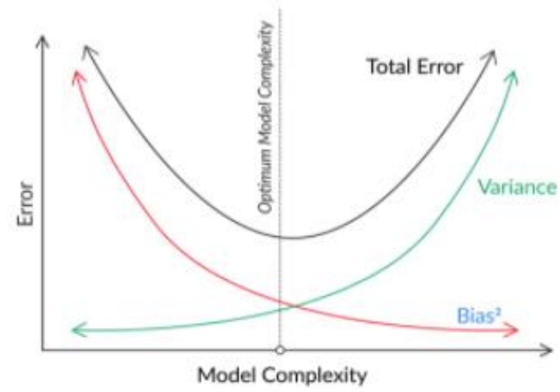
1. FullBath
2. GarageArea
3. Fireplaces
4. KitchenQual
5. MSZoning_RL

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?





- Performance of a robust model should not get affected much by any variation to the data.
- A generalizable model is defined based on its ability to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.
- When the model memorizes the noise in the train data and fits too closely to the training set, it would become “overfitted,” and it won’t be able to generalize well to new data.
- As in the above diagram we need to prevent overfitting to make the model more robust and generalizable. An overfitted model has a very high variance and any minor change in the data would impact the model’s predictability.
- Models which are too complex tend to be overfitted hence our focus should be on make the model simpler by taking measures to avoid overfitting to make sure that the model becomes robust and generalizable.
- Following are some ways to avoid overfitting.
    - Training with more data
    - Data augmentation
    - Addition of noise to the input data
    - Feature selection
        - Too much parameters would lead to high complexity.
    - Cross-validation
    - Regularization

- Regularization technique would be applying a penalty to the input parameters with the larger coefficients and this would limit the model's variance. L1(Lesso) regularization forces the weight parameters to become zero. L2(Ridge) regularization forces the weight parameters towards zero. This way regularization technique makes the model less complex and move it towards the bit fit model.

- What are the implications of the same for the accuracy of the model and why?

  - Models which are too complex (i.e overfitted) would have a very high accuracy but mostly would have very low accuracy in test data. As in the above diagrams while making the model generalizable (and robust) it has to decrease the variance and it would lead to some bias.
  - Addition of the bias to the model (to make it generalizable) would lead to decrease in the accuracy.
  - So, in order to find the best model, there should be a balance between model accuracy and complexity.
  - The regularization technique such as Lasso and Ridge discussed above are used to find the best model by balancing the model complexity and accuracy.