

**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 to double the value of alpha for both ridge and lasso regression? What will be the most important predictor variables after the change is implemented?**

**Answer:**

The optimal value for ridge regression is **477.85** and the optimal value for lasso regression is **0.0019**.

The importance of variables changes if we double the value of alpha in case of ridge:

features_alpha	ranking_alpha	features_2alpha	ranking_2alpha
0	OverallQual	1	OverallQual
1	GrLivArea	2	GrLivArea
2	GarageCars	3	GarageCars
3	KitchenQual	4	KitchenQual
4	BsmtQual	5	BsmtQual
5	1stFlrSF	6	1stFlrSF
6	YearBuilt	7	YearBuilt
7	FireplaceQu	8	Fireplaces
8	OverallCond	9	ExterQual
9	MSZoning_RL	10	TotRmsAbvGrd

As we can see that the top 7 features are exactly same for both version for ridge regression just the ranking of the features are shifted here and there.

**Note: The following ranking of feature is obtained using RFE.**

The importance of variables changes if we double the value of alpha in case of lasso:

features_alpha	ranking_alpha	features_2alpha	ranking_2alpha
0	OverallQual	1	OverallQual 1
1	GrLivArea	2	GrLivArea 2
2	YearBuilt	3	YearBuilt 3
3	OverallCond	4	GarageCars 4
4	TotalBsmtSF	5	OverallCond 5
5	RoofMatl_CompShg	6	TotalBsmtSF 6
6	RoofMatl_Tar&Grv	7	MSZoning_RL 7
7	RoofMatl_WdShngl	8	RoofMatl_CompShg 8
8	RoofMatl_Metal	9	RoofMatl_Tar&Grv 9
9	RoofMatl_Membran	10	RoofMatl_WdShngl 10

As we can see that the top 3 features are exactly same for both version for lasso regression, but the other features and their ranking are quite different for both versions.

**Note: The following ranking of feature is obtained using RFE.**

**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?**

**Answer:**

While determining the optimal value of lambda we chose lasso regression model because the predictive  $r^2$  score is **90%** (which is slightly better than ridge **88%**) and it uses **161 features** only which is significantly lesser than the ridge regression model which makes the lasso model lighter model than ridge regression.

**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 predictor variables. Which are the five most predictor variables now?**

**Answer:**

	features_alpha	ranking_alpha
0	OverallQual	1
1	GrLivArea	2
2	YearBuilt	3
3	OverallCond	4
4	TotalBsmtSF	5
5	RoofMatl_CompShg	6
6	RoofMatl_Tar&Grv	7
7	RoofMatl_WdShngl	8
8	RoofMatl_Metal	9
9	RoofMatl_Membran	10

These are the top 10 features and the most important variables of lasso model as per the RFE. Since the first five variables are not available in the new incoming data, we will use the next five variables from the above list which consist of the list starting from RoofMatl\_CompShg to RoofMatl\_Membran.

**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?**

**Answer:**

We can make our model robust and generalizable by reducing the variance by injecting some bias into the model artificially means shrinking the parameter length vector towards 0. This makes the problem of overfitting, and the predictive power of model will improve significantly.