Advanced regression

Subjective Questions

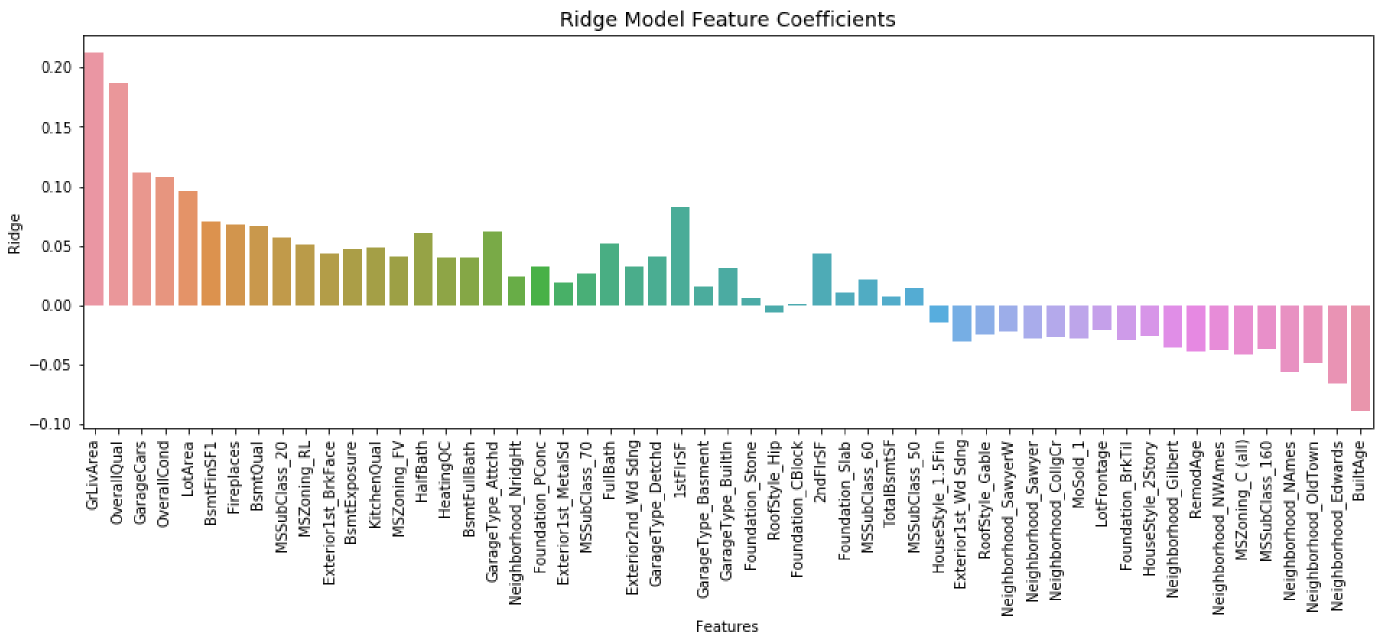
Jithin Prakash K

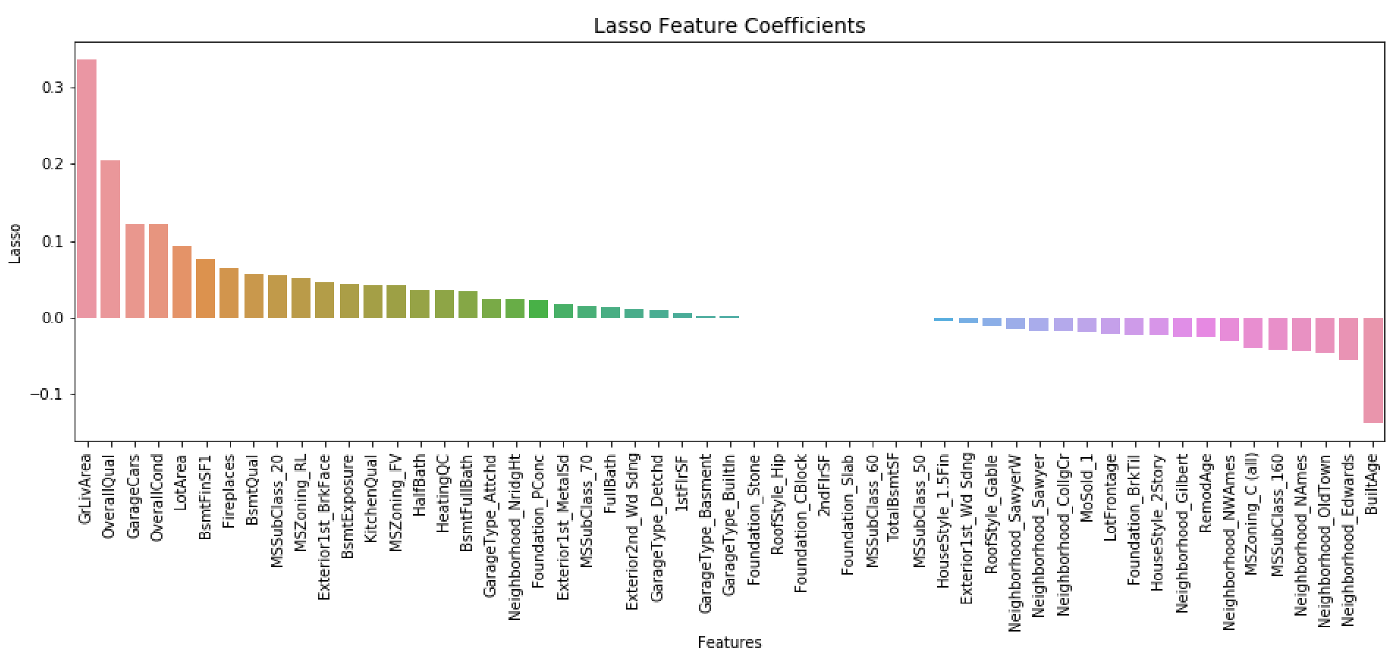
Question 1

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?

Answer:

Original Alpha value for Ridge Regression model was found to be: 80
Original Alpha value for Lasso Regression model was found to be: 0.005





When the alpha value is doubled for the Ridge regression, the change in the top 10 features is as following

	Ridge (Alpha = 80)	Ridge (Alpha = 160)
0	GrLivArea	GrLivArea
1	OverallQual	OverallQual
2	GarageCars	GarageCars
3	OverallCond	OverallCond
4	LotArea	1stFlrSF
5	BuiltAge	LotArea
6	1stFlrSF	BuiltAge
7	BsmtFinSF1	Fireplaces
8	Fireplaces	BsmtQual
9	BsmtQual	BsmtFinSF1

	Ridge (Alpha = 80)	Ridge (Alpha = 160)
Alpha	80.000000	160.000000
R-Squared (Train)	0.901160	0.897803
R-Squared (Test)	0.911499	0.911300
RSS (Train)	100.915982	104.343593
RSS (Test)	42.377363	42.472745
MSE (Train)	0.098840	0.102197
MSE (Test)	0.096752	0.096970

The features are almost the same, only that the order is changed when doubling the alpha value. Also, it is observed that 5 features have coefficients very low in range 0.0005 – 0.008 for double the value of lambda/alpha.

The train and test R-squared value is reduced when the alpha is increased, i.e., the higher alpha value the model becomes underfit.

	Lasso (Alpha = 0.005)	Lasso (Alpha = 0.010)
0	GrLivArea	GrLivArea
1	OverallQual	OverallQual
2	BuiltAge	GarageCars
3	GarageCars	BuiltAge
4	OverallCond	OverallCond
5	LotArea	LotArea
6	BsmtFinSF1	BsmtFinSF1
7	Fireplaces	Fireplaces
8	BsmtQual	BsmtQual
9	Neighborhood_Edwards	Exterior1st_BrkFace

	Lasso (Alpha = 0.005)	Lasso (Alpha = 0.010)
Alpha	0.005000	0.010000
R-Squared (Train)	0.900956	0.896422
R-Squared (Test)	0.912423	0.911257
RSS (Train)	101.123776	105.752874
RSS (Test)	41.935077	42.493456
MSE (Train)	0.099044	0.103578
MSE (Test)	0.095742	0.097017

The features are almost the same, except for the last feature out of top 10, also, the order is changed when doubling the alpha value. It is observed that for alpha = 0.005 there are 7 features with coefficients as zero and when the alpha is double, alpha = 0.010, there are 16 features with coefficients as zero.

The train and test R-squared value is reduced when the alpha is increased, i.e., the higher alpha value the model becomes underfit.

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:
Optimal value of lambda is chosen, and there are 55 parameters in the final model, and since there are 55 models, the lasso approach would be the best as we get some features eliminated as the lasso approach makes the coefficient zero for the less significant features. This simplifies the model complexity. The Lasso regression, helps in feature selection by making the coefficients of less significant features as zero. Hence Lasso regression model is chosen as the final model.

In addition, the lasso and ridge regression came up with similar accuracy metric values.

	Linear	Ridge	Lasso
R-Squared (Train)	0.9050	0.9012	0.9010
R-Squared (Test)	0.9062	0.9115	0.9124
RSS (Train)	97.0282	100.9160	101.1238
RSS (Test)	44.9361	42.3774	41.9351
MSE (Train)	0.0950	0.0988	0.0990
MSE (Test)	0.1026	0.0968	0.0957
Alpha	NaN	80.0000	0.0050

```
betas[betas.Lasso==0]
```

	Features	Linear	Ridge	Lasso
29	RoofStyle_Hip	-0.032111	-0.006488	0.0
30	Foundation_CBlock	0.194743	0.000562	-0.0
31	2ndFlrSF	-0.058884	0.043434	-0.0
32	Foundation_Slab	0.053699	0.011172	0.0
33	MSSubClass_60	0.059956	0.022090	-0.0
34	TotalBsmtSF	-0.025812	0.007217	-0.0
35	MSSubClass_50	0.044361	0.014321	-0.0

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?

Answer:

Five most important predictor variables from the lasso model initially were,

1. GrLivArea
2. OverallQual
3. BuiltAge
4. GarageCars
5. OverallCond

Since these predictor variables are no more available in the incoming data, we need to re-model the data using the new set of data, so a Grid Search is used with new Training data to obtain the optimum alpha value for the new set of incoming data, it was found that the alpha value remains same, even after dropping the important variables. Optimum alpha is observed to be = 0.005

	Lasso_New
R-Squared (Train)	0.8629
R-Squared (Test)	0.8826
RSS (Train)	139.9916
RSS (Test)	56.2224
MSE (Train)	0.1371
MSE (Test)	0.1284
Alpha	0.0050

It is also observed that the R-squared value has reduced to 86% for training set and 88% for testing set.

The first 5 important variables with the new set of incoming data are

	Lasso_New
1stFlrSF	0.296579
2ndFlrSF	0.211047
GarageType_Attchd	0.164347
GarageType_Detchd	0.136766
BsmtQual	0.131079
KitchenQual	0.116136
FullBath	0.104894
LotArea	0.103058
RemodAge	-0.102126
Neighborhood_Edwards	-0.100168

1. 1stFlrSF
2. 2ndFlrSF
3. Garage Type Attached
4. Garage Type Detached
5. Basement Quality

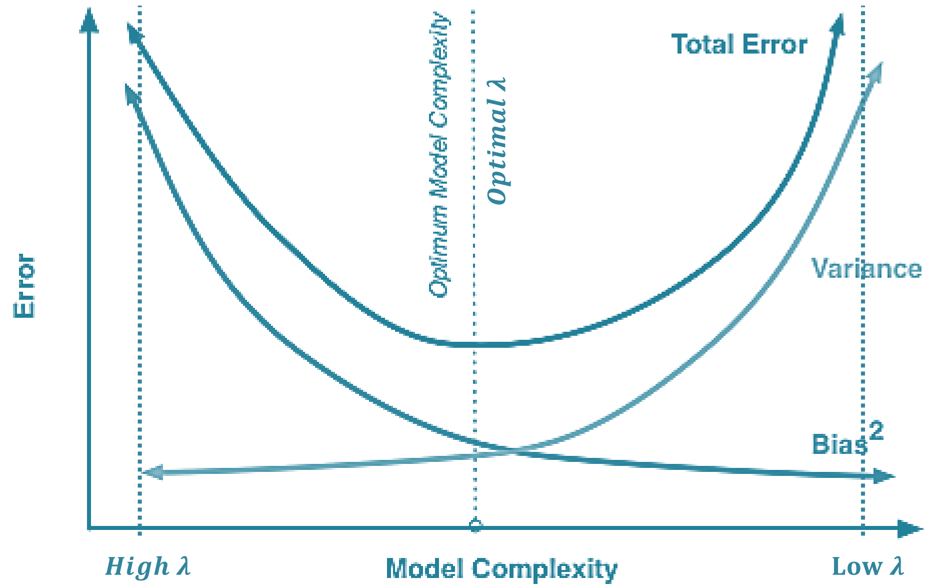
Question 4

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:
The model will be robust when the Bias (Bias^2) and Variance is low. When the model is too simple, the Bias will be too high and variance will be too low. This is a case of under fitted model. As the model complexity increases, the bias keeps reducing steadily and the variance keep increasing steadily. When model becomes too complex, the bias (Bias^2) will be lowest but the variance will be too high. This is an over-fitted model scenario.

The regularization technique is one of the methods that helps the model to be in the robust and generalized condition, but helping the model to be not overfitted by applying penalty by using a hyper-parameter using lambda/alpha.

Refer the graphical representation as below,



The model is robust and generalized when there is a $Bias^2$ – variance trade off, then the bias and variance are low or in other terms when the total error is low. This happens for optimal model complexity and optimal lambda value. Model is robust when the model is not over fitted or under fitted.

When the complexity of model is more, it means the Bias is low and variance is high, the value of lambda goes to 0. There is no penalization for the model. This means the model is over fitted. The model has learnt the pattern completely and it works very well for the training data. This model will give very high accuracy scores for the training data since the model has learnt all the data points in the training data, However, when the testing data is applied, the test accuracy will be low since the data is new and the pattern is now learnt by the model.

When the complexity of the model is low. It means the Bias is high and the variance is low. This means the model is not able to learn the pattern from that data, in this case both the train accuracy and the test accuracy will be low. When the lambda values in regularization is high, the higher lambda would eliminate the effect of many features thereby making the model not learn the pattern within the data. This is the case of under-fitting.

From underfitting to overfitting, the accuracy of the training set increases steadily, model learns patterns and then slowly learns all the data points when the model is overfitted. When the model is optimal, the training and testing accuracy will be optimal, and value will be in similar range. The lambda will penalize the insignificant features. The model will not be underfitted or overfitted. The bias and variance will be at the lower in total (Total error = $Bias^2$+Variance). Model is generalized and robust when the model complexity and lambda/alpha value is optimal.