Question1

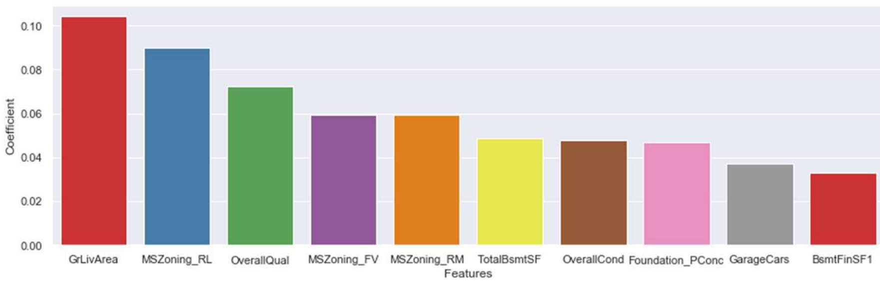
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

Below are the top 5 features and their coefficients.

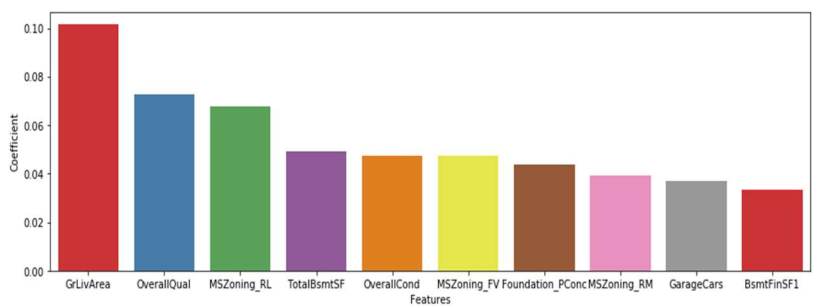
a. Ridge Regression

The optimal value of alpha for Ridge Regression is 10.



	Features	rfe_support	rfe_ranking	Coefficient
5	GrLivArea	True	1	0.1042
12	MSZoning_RL	True	1	0.0901
1	OverallQual	True	1	0.0723
10	MSZoning_FV	True	1	0.0594
13	MSZoning_RM	True	1	0.0594
4	TotalBsmtSF	True	1	0.0489
2	OverallCond	True	1	0.0478
14	Foundation_PConc	True	1	0.0468
7	GarageCars	True	1	0.0370
3	BsmtFinSF1	True	1	0.0332

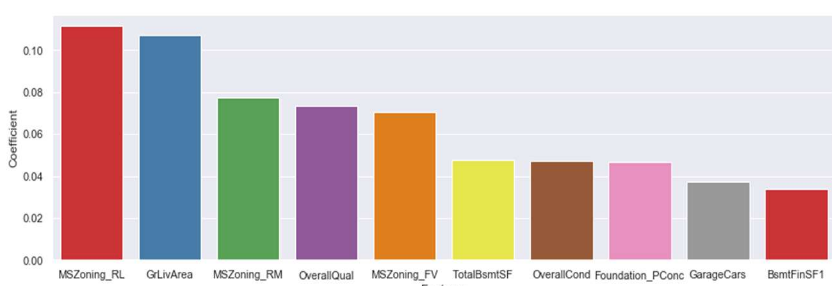
Ridge Regression with alpha=20



	Features	rfe_support	rfe_ranking	Coefficient
5	GrLivArea	True	1	0.1017
1	OverallQual	True	1	0.0727
12	MSZoning_RL	True	1	0.0675
4	TotalBsmtSF	True	1	0.0492
2	OverallCond	True	1	0.0476
10	MSZoning_FV	True	1	0.0476
14	Foundation_PConc	True	1	0.0440
13	MSZoning_RM	True	1	0.0391
7	GarageCars	True	1	0.0368
3	BsmtFinSF1	True	1	0.0332

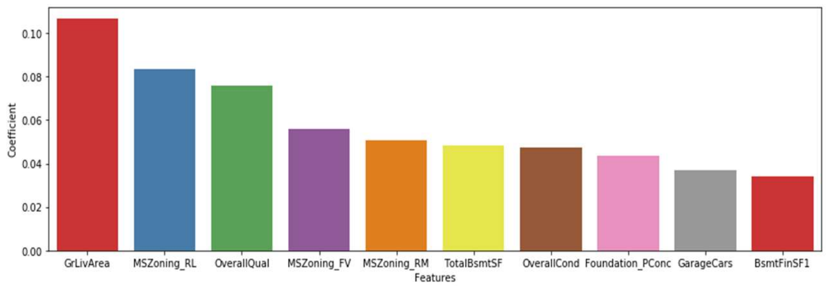
b. Lasso Regression

The optimal value of Alpha for Lasso Regression is 0.0004



	Features	rfe_support	rfe_ranking	Coefficient
12	MSZoning_RL	True	1	0.111460
5	GrLivArea	True	1	0.107201
13	MSZoning_RM	True	1	0.077262
1	OverallQual	True	1	0.073446
10	MSZoning_FV	True	1	0.070426
4	TotalBsmtSF	True	1	0.047874
2	OverallCond	True	1	0.047355
14	Foundation_PConc	True	1	0.046559
7	GarageCars	True	1	0.037357
3	BsmtFinSF1	True	1	0.033709

Lasso Regression with alpha=0.0008



	Features	rfe_support	rfe_ranking	Coefficient
5	GrLivArea	True	1	0.106688
12	MSZoning_RL	True	1	0.083239
1	OverallQual	True	1	0.075624
10	MSZoning_FV	True	1	0.055649
13	MSZoning_RM	True	1	0.050470
4	TotalBsmtSF	True	1	0.048391
2	OverallCond	True	1	0.047293
14	Foundation_PConc	True	1	0.043690
7	GarageCars	True	1	0.036974
3	BsmtFinSF1	True	1	0.034101

From the result above, we can see that as the penalty value, α increases:

- **Lasso regression** shrinks coefficients all the way to zero, thus removing them from the model.
- **Ridge regression** shrinks coefficients toward zero, but they rarely reach zero

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:

As per the below metrics, Lasso has better R2 score and less Mean error square values compared to Ridge regression. Additionally, lasso helps in Variable Selection. So, Lasso regression will be used for deployments.

Metrics	**Lasso Regression**	**Ridge Regression**
R2score – train	0.9184	0.9181
R2 score – test	0.9032	0.9011
Mean squared error train	0.0114	0.01151
Mean squared error test	0.01346	0.01375

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:

Below are the five top predictors after removing the top 5 features from lasso in incoming data :

	Features	rfe_support	rfe_ranking	Coefficient
5	2ndFlrSF	True	1	0.096213
4	1stFlrSF	True	1	0.085884
3	TotalBsmtSF	True	1	0.056528
1	OverallCond	True	1	0.056401
14	Foundation_PConc	True	1	0.042647

Question 4

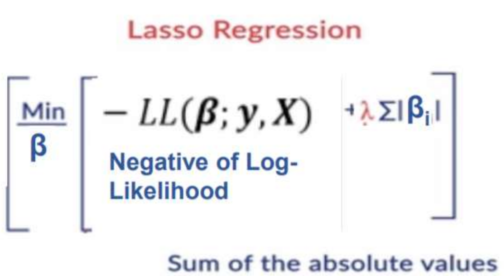
How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer:

- Regularization can be used to avoid overfitting by creating and optimally complex model.
- Regularization adds a penalty on the different parameters of the model to reduce the freedom of the model.
- Hence the model will be less likely to fit he noise of the training data and will improve the generalization abilities of the model.
- In regularized logistic regression, the objective function has two parts, the cost function and regularization term.

$$\hat{\boldsymbol{\beta}} = \min_{\beta} \; -LL(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X}) + \lambda\, R(\boldsymbol{\beta})$$

  - In case of **lasso regression**, a regularization term of "sum of the absolute value of the coefficients" is added.



- In **ridge regression**, an additional term of "sum of the squares of the coefficients" is added to the cost function along with the error term

Lasso Regression

$$\min_{\beta} \left[-LL(\boldsymbol{\beta}; \boldsymbol{y}, \boldsymbol{X}) + \lambda \Sigma |\beta_i| \right]$$

Negative of Log-Likelihood

Sum of the absolute values

So, by using regularization model can become robust and generalizable.

Implications of Regularization:

- We have understood that the main reason of overfitting is due to the random errors and fluctuations which are learned by the model while training. The model learns it so well that training data model's accuracy becomes very high and resulting in overfitting.
- We have observed that sudden change in the feature weights resulting in non-linearity of the model. The easy way is to penalize the feature weights when they get updated. As explained above two regularization methods make use of this technique.
- Here, λ is a hyper parameter whose value is decided by us. If λ is high it adds high penalty to error term making the learned hyper plain almost linear, if λ is close to 0 it has almost no effect on the error term causing no regularization. So, these regularization methods help us to get optimum coefficients such that Train and Test data accuracy will be high and almost nearby.
- Also, in Lasso regularization helps in feature selection technique too as it zero out the respective weights of undesired features, but it is computationally expensive.