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

The optimal values of alpha –

Rigde = 0.2
Lasso = 50

	Ridge			Lasso		
	alpha = default	alpha = 0.2 (opt)	alpha=0.4 (2 x opt)	alpha = default	alpha = 50 (opt)	alpha=100 (2 x opt)
R^2 (train)	0.9217	0.9345	0.9301	0.9425	0.9209	0.9046
R^2 (test)	0.8175	0.822	0.8187	0.7016	0.8479	0.8434

If we choose to double the value of alpha (2 x optimal value), we expect model to get simpler. In other words, bias would increase and variance would further decrease ie, model may start underfitting both train and test data. And indeed, we see a decrease in the R^2 value of both ridge and lasso regression after doubling the value of alpha. Hence, doubling the value of from its optimal value makes model underfit the data.

Five Most Important predictor variables after doubling the value of alpha –

Ridge –

- 'BsmtFinSF1_square',
- 'RoofMatl_WdShngl',
- 'Condition2_PosN',
- '1stFlrSF',
- 'TotalBsmtSF_square'

Lasso –

- 'BsmtFinSF1_square',
- 'RoofMatl_WdShngl',
- 'Condition2_PosN',
- 'OverallQual_square',
- 'GrLivArea'

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

I will choose lasso with optimal value of alpha (50) for two reasons –

1) R^2 on test is higher than ridge regression

R^2 ridge = 0.822
R^2 lasso = 0.8479

2) Difference between train and test R^2 of lasso is less than ridge
ridge (R^2 train - R^2 test) = 0.1125
lasso (R^2 train - R^2 test) = 0.073

Question-3

After building the model, you realised 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

- 'BsmtFinSF1_square',
- 'BsmtUnfSF'
- 'TotalBsmtSF'
- 'BsmtUnfSF_square'
- MSSubClass_square'

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?

Answer

I will add regularization to the model for making model robust and generalizable. Regularization makes model simpler reducing the variance at the cost of bias. Because of regularization accuracy on train set will decrease and on test set will increase.