**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?

Answers

Optimal Value of alpha in lasso is 0.01 and ridge is 20 . in case of lasso

Predictor variable coefficient becomes 0 which allow us for future

selection.



After changing alpha to 0.02.



Ridge model with optimal value of alpha=20. There is decrease in coefficient of predictor variable however predictor variable coefficient will not become 0.

[('Constant', -0.601),

('LotArea', 0.067),

('OverallQual', 0.387),

('BsmtQual', 0.096),

('BsmtFinSF1', 0.088),

('TotalBsmtSF', 0.097),

('GrLivArea', 0.315),

('BedroomAbvGr', -0.038),

('KitchenQual', 0.038),

('GarageArea', 0.131),

('Condition1\_Norm', 0.206),

('Condition2\_PosN', -0.209),

('Condition2\_RRAe', -0.004),

('BldgType\_Duplex', -0.093),

('BldgType\_TwnhsE', -0.093),

('HouseStyle\_2.5Fin', -0.057),

('HouseStyle\_2.5Unf', -0.121),

('HouseStyle\_SFoyer', 0.021),

('RoofStyle\_Gambrel', 0.024),

('RoofMatl\_CompShg', 0.07),

('RoofMatl\_Membran', 0.033),

('RoofMatl\_Metal', 0.006),

('RoofMatl\_Roll', 0.002),

('RoofMatl\_Tar&Grv', 0.014),

('RoofMatl\_WdShake', 0.014),

('RoofMatl\_WdShngl', 0.202),

('Exterior1st\_AsphShn', -0.003),

('Exterior1st\_BrkComm', -0.023),

('Exterior1st\_CBlock', -0.006),

('Exterior1st\_CemntBd', 0.013),

('Exterior1st\_ImStucc', -0.007),

('Exterior1st\_Plywood', -0.002),

('Exterior1st\_Stone', -0.003),

('Exterior2nd\_AsphShn', -0.003),

('Exterior2nd\_CBlock', -0.006),

('Exterior2nd\_CmentBd', 0.043),

('Exterior2nd\_Stone', 0.031),

('Exterior2nd\_Wd Sdng', -0.082),

('Foundation\_PConc', 0.176),

('Foundation\_Slab', 0.068),

('Heating\_GasA', 0.042),

('Heating\_GasW', -0.048),

('Heating\_Grav', -0.006),

('Heating\_OthW', -0.059),

('Heating\_Wall', 0.041),

('Electrical\_FuseP', -0.056),

('Functional\_Min1', 0.038),

('Functional\_Min2', 0.061),

('Functional\_Sev', -0.043),

('Functional\_Typ', 0.163),

('GarageType\_BuiltIn', 0.091)]

After doubling the alpha i.e 40

[('Constant', -0.488),

('LotArea', 0.067),

('OverallQual', 0.384),

('BsmtQual', 0.091),

('BsmtFinSF1', 0.088),

('TotalBsmtSF', 0.099),

('GrLivArea', 0.305),

('BedroomAbvGr', -0.032),

('KitchenQual', 0.033),

('GarageArea', 0.136),

('Condition1\_Norm', 0.182),

('Condition2\_PosN', -0.105),

('Condition2\_RRAe', -0.003),

('BldgType\_Duplex', -0.069),

('BldgType\_TwnhsE', -0.069),

('HouseStyle\_2.5Fin', -0.03),

('HouseStyle\_2.5Unf', -0.069),

('HouseStyle\_SFoyer', 0.01),

('RoofStyle\_Gambrel', 0.013),

('RoofMatl\_CompShg', 0.036),

('RoofMatl\_Membran', 0.015),

('RoofMatl\_Metal', 0.003),

('RoofMatl\_Roll', -0.001),

('RoofMatl\_Tar&Grv', 0.003),

('RoofMatl\_WdShake', 0.005),

('RoofMatl\_WdShngl', 0.109),

('Exterior1st\_AsphShn', -0.002),

('Exterior1st\_BrkComm', -0.014),

('Exterior1st\_CBlock', -0.003),

('Exterior1st\_CemntBd', 0.016),

('Exterior1st\_ImStucc', -0.004),

('Exterior1st\_Plywood', -0.006),

('Exterior1st\_Stone', -0.001),

('Exterior2nd\_AsphShn', -0.002),

('Exterior2nd\_CBlock', -0.003),

('Exterior2nd\_CmentBd', 0.033),

('Exterior2nd\_Stone', 0.017),

('Exterior2nd\_Wd Sdng', -0.073),

('Foundation\_PConc', 0.168),

('Foundation\_Slab', 0.04),

('Heating\_GasA', 0.033),

('Heating\_GasW', -0.036),

('Heating\_Grav', -0.005),

('Heating\_OthW', -0.03),

('Heating\_Wall', 0.022),

('Electrical\_FuseP', -0.03),

('Functional\_Min1', 0.008),

('Functional\_Min2', 0.026),

('Functional\_Sev', -0.023),

('Functional\_Typ', 0.117),

('GarageType\_BuiltIn', 0.081)]

**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?

I will choose lasso model because lasso model allow me to filter predictor variable which are least corelated with target variable. Ridge model unnecessary add all feature in the model building process and model will become more complex as compared to lasso model.

**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:

These are the below predictor variable

[('OverallQual', 0.417),

('GrLivArea', 0.294),

('Condition1\_Norm', 0.159),

('Foundation\_PConc', 0.144),

('GarageArea', 0.134),

These are the predictor variable after removing five most predictor variable specified as above.

'TotalBsmtSF', 0.333),

('GarageCars', 0.297),

('GarageType\_Attchd', 0.277),

('2ndFlrSF', 0.253),

('TotRmsAbvGrd', 0.221),

('GarageType\_BuiltIn', 0.183

**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:-

Model should predict very well on training data as well as test data. Model should not have overfit and underfit. This can be done by using regularization technique.

We have used ridge and lasso regularization to gernerialize model. Model should have low bias and high variance. Model should know pattern for prediction.