

## Assignment Questions

### 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 1(a):

Ridge and Lasso regression are methods to estimate the coefficients, where independent variables are highly correlated. The former method, doesn't zero down the coefficients but would be nearly zero, whereas the latter method zero downs the coefficients and thus helps us in regularizing the coefficients with successful elimination or choice.

Alpha - Ridge	Alpha - Lasso
0.2	0.005

#### Answer 1(b): Case when the alpha is double

Alpha - Ridge	Alpha - Lasso
0.2*2	0.005*2
0.939315456456721 0.9154237896976329	0.8628856212963936 0.8543005415187783

#### Observations:

1.(b).1 I've observed that the R2 Score of Lasso became less compared to Ridge score.

1.(b).2 The idea of doubling is to ensure that coefficients values are lowered in the former method whereas more zeroing of coefficients happens more in case of latter method. Furthermore, it also to be understood how its affecting the features and coefficients.

## 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 2 : In my assignment work and data study, I've found Ridge score to be more compared to Lasso score and would select Ridge method for feature selection and regression expression.

There is a possibility of reversal here, probably cases not thought by me or found by me. In any case, the selection would further increment the zeal of identifying a better approach for the model and fine tuning of the same. There by suggesting the company to set ground rules for their property expansion or sale.

## 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 3 :

Before Dropping Predictor Variables - ['LotFrontage', 'LotArea', 'MasVnrArea', 'BsmtFinSF1', 'TotalBsmtSF']

After Dropping Predictor Variables -

BsmtFinSF2	BsmtUnfSF	1stFlrSF	2ndFlrSF	LowQualFinSF	GrLivArea
------------	-----------	----------	----------	--------------	-----------

Note : Relevant code is available in the bottom section of the Jupyter Notebook.

#### 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 4 :

To make model robust and generalisable, we've to consider the following points

1. Train-Test Validation : We need to effectively select the train-test data-subsets and relevant feature selections to gauge its performance with test or unknown data for the models generalisation capabilities.
2. Regularization : Preventing overfitting of data while training by subduing or imposing penalties or restrictions using L1 or L2 methods thereby penalizing large parameter values, overly complex models etc.
3. Ensemble Learning of Multiple models : Ensure our data is ensembled through different known models to multiply the diverse perspectives and thereby reducing overfitting involved, if any. Methods like bagging, boosting, stacking also enhance the robustness and generalization capabilities of the model.
4. Feature Engineering : Careful selection and pre-processing of the features provides significant insights while irrelevant information and relevant or irrelevant data noise is casted away. In addition to this, domain knowledge is always shines like fine wine.
5. Tuning of Hyper Parameters : Using techniques like Grid-Search, Randomized Search, carefully select optimized and tuned parameters, thereby, increasing the optimization and robustness of the considered or developed model.
6. Model Eval Metrics : Right selection of eval metrics prioritizing generalization compared to performance over training accuracy. Metrics like F1-score, recall, ROC-Curve can provide more insights considering different aspects.

## Implications of Robustness and Generalization

1. We may have to trade off with training accuracy. At times we may observe that the models outperforms unknown or unseen data set but may not achieve high accuracy on the trained data set, if it's avoiding overfitting.
2. Focussing solely on training accuracy and missing the point of generalisability results in, inadequate performance in practical applications.