Assignment Part-II

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
Optimal value for lambda for ridge regression is 10 and
Optimal value for lasso regression is 0.0001.

If we double the lambda values for ridge regression the coefficients are lowered and
for lasso the coefficients are turning to 0 or negative values.

MS Subclass which identifies the type of dwelling involved in the sale, would be important predictor variables after the change is implemented in both the models.

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
Assignment was carried out with lambda values of –
10 for ridge regression and 0.0001 for lasso regression as it does not cause overfitting or underfitting.
In both the regression models the R2 scores are not 0 or 1 but nearing to 1.
For ridge regression, R2 scores (train:0.949 & test: 0.91)
For lasso regression, R2 scores (train:0.947 & test:0.91)

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
For ridge regression:
Electrical, GrLivArea, WoodDeckSF, OverallQual and RoofStyle

```
# Significant variables to predict price of house using ridge regression

ridge_coef.sort_values(by = 'coefficients',ascending = False).head(10)
```

	Features	coefficients
0	MSSubClass	0.123075
27	HalfBath	0.010828
6	YearBuilt	0.010011
24	LowQualFinSF	0.008961
18	BsmtUnfSF	0.007601
21	Electrical	0.007199
25	GrLivArea	0.006764
36	WoodDeckSF	0.006725
4	OverallQual	0.005946
7	RoofStyle	0.005114

For lasso regression:
KitchenAbvGr, Neighborhood_StoneBr, GarageArea, GarageQual and WoodDeckSF

```
# Significant variables to predict price of house using lasso regression

lasso_coef.sort_values(by = 'coefficients',ascending = False).head(10)
```

	Features	coefficients
0	MSSubClass	0.132490
46	Neighborhood_Gilbert	0.022268
18	BsmtUnfSF	0.020163
4	OverallQual	0.019631
55	Neighborhood_SWISU	0.016551
29	KitchenAbvGr	0.007531
58	Neighborhood_StoneBr	-0.000000
34	GarageArea	-0.000000
35	GarageQual	0.000000
36	WoodDeckSF	0.000000

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 can be made robust and generalisable if it is simple.

It's implications are:

1. Simpler models make higher error in training data but in case of out of box scenarios.
2. Simpler models perform better compared to complex models.
3. It doesn't change significantly with changes in training data.
4. It requires less data.