

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?

Ans-

Optimal value of alpha for ridge is 1.0 and for Lasso is 0.0001. Features remains almost same after doubling the value.

- LotFrontage : If the house Linear feet of street connected to property area increase then the Price increase.
- BsmtFullBath : : If the BsmtFullBath area is more the SalePrice is higher
- Overall Condition: If the Overall Condition is Excellent the SalePrice is higher
- MSZoning_RH : If the house i near residential area then the SalePrice is higher
- Overall quality: If the Overall Condition is Excellent the SalePrice is higher
- Exterior1st_CBlock : IF the house Exterior1st is CBlock then price is less.
- Garage Area: If the Garage area is high the SalePrice is higher
- CentralAir: If the CentralAir is Yes the SalePrice is higher

Ridge Coefficient after changing value-

	Features	Coefficient	Mod
0	LotFrontage	10.261566	10.261566
3	OverallCond	0.527282	0.527282
14	BsmtUnfSF	0.406426	0.406426
12	BsmtFinType2	0.355562	0.355562
2	OverallQual	0.338037	0.338037
11	BsmtFinSF1	0.322512	0.322512
9	BsmtExposure	0.316265	0.316265
33	GarageFinish	0.303286	0.303286
73	LotConfig_CulDSac	-0.272453	0.272453
6	ExterCond	0.264079	0.264079

Lasso Coef-

:

	Feature	Coef	mod
0	LotFrontage	10.174821	10.174821
14	BsmtFullBath	0.822796	0.822796
3	OverallCond	0.568885	0.568885
9	CentralAir	0.492011	0.492011
2	OverallQual	0.484268	0.484268
73	Exterior1st_CBlock	-0.479688	0.479688
33	MSZoning_RH	0.379703	0.379703
35	MSZoning_RM	0.297522	0.297522
36	Street_Pave	0.246404	0.246404
20	GarageQual	0.240831	0.240831

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?

Ans-

We will make use of Lasso Regression model because it helps in reducing the number of features and giving almost accurate result . Its more efficient model than Ridge regression 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?

Ans- After removing top 5 features we will get new predictors as mention below.

LotArea- Lot size in square feet

FullBath- Full bathrooms above grade

1stFlrSF-First Floor square feet

ExterCond-

MSZoning_RH- Evaluates the present condition of the material on the exterior

	Feature	Coef	mod
0	LotArea	10.205140	10.205140
10	FullBath	1.029837	1.029837
6	1stFlrSF	0.606857	0.606857
1	ExterCond	0.561906	0.561906
28	MSZoning_RH	0.512930	0.512930

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?

Ans-We can validate of our model on test data and add optimal alpha to generalise the model so that there should not be major difference in term of accuracy on unseen data as well.

So minimise the value of error and bias both and keep model simple.