

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?

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?

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?

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?

Answers

- 1) For lasso optimal alpha is 0.001, for ridge alpha is 0.9. As computed in notebook, we can conclude that ridge is affected slightly more with the alpha change

Ridge Previous Values (alpha 0.9):

0.8512268074783247 0.8157240337525367

Ridge New Values (alpha: 1.8):

0.8481463335238923 0.820914014014055

Lasso Previous Values (alpha 0.001):

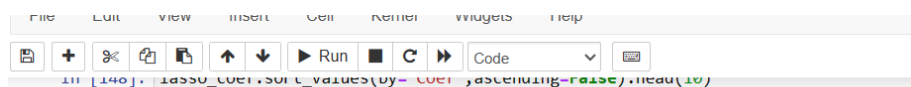
0.8479098486229748 0.818211575106701

Lasso New Values (alpha: 0.002):

0.843238116080688 0.8230719338643002

We can conclude that ridge is affected slightly more with the alpha change.

- 2) lasso regression would penalize more on no of variables in the model and help in feature elimination thus, using Lasso as final training model.
- 3) After creating a new lasso model by excluding previous top 5 predictor variables as shown in the notebook:



Out[148]:

	Feature	Coef
22	FullBath	1.674631
8	ExterCond	0.570775
11	BsmtFinType2	0.397174
4	OverallQual	0.368824
2	LotShape	0.276218
1	LotArea	0.271996
5	OverallCond	0.249381
29	GarageCars	0.213903
3	LandSlope	0.164061
20	BsmtFullBath	-0.000000

New top 5 predictors are

- FullBath: Full bathrooms above grade
- ExterCond: Evaluates the present condition of the material on the exterior
- BsmtFinType2: Rating of basement finished area
- OverallQual: Rates the overall material and finish of the house
- LotShape: General shape of property

New top 5 predictors: FullBath,ExterCond,BsmtFinType2,OverallQual,LotShape

- 4) We should build an overly complicated model. We should try to keep it as simple as possible. We should make sure that our model is not overfitting. If this is happening, we need to regularize using lasso or ridge regression. Also we should select features carefully and make sure that the model is not learning the noise. Outliers need to be treated as they have impact on overfitting.

Implications for accuracy of the model:

- 1) If a model is overfitting, regularization will help us address this
- 2) The model must perform give stable results across different datasets
- 3) The model should consistently perform over various time periods as the datasets tend to change over period of time