

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 of alpha for

Ridge Regression = 100

Lasso Regression = 0.01

On doubling the value of alpha,

In case of Ridge Regression (alpha = 200), the accuracy reduced by appx 1 % for both test and train data. Also, a slight increase in MSE is also observed. The coefficients are much more smaller and near zero.

	Alpha = 200	Alpha = 100
R2 Train	0.9287291690842262	0.9354498473230527
R2 test	0.9054186708470979	0.9103160773829793
MSE train	0.07127083091577381	0.06455015267694732
MSE test	0.11598485920146992	0.10997918120345535

In case of Lasso Regression (alpha = 0.02) as well the accuracy reduced by appx 1 % for both test and train data. Also, a slight increase in MSE is also observed. More number of coefficients are now zero than when alpha was 0.01.

	Alpha = 0.02	Alpha = 0.01
R2 Train	0.9104611101619372	0.9258744327494841
R2 test	0.9003858383100153	0.9073876618757044
MSE train	0.08953888983806285	0.07412556725051596
MSE test	0.12215660978296612	0.11357029018169343

Even though the coefficient has been reduced further to zero, in both cases the most significant predictor remains **GrLiveArea**

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:

With the obtained optimal alpha value for Ridge and Lasso regression,

	Ridge	Lasso
R2 Train	0.9354498473230527	0.9258744327494841
R2 test	0.9103160773829793	0.9073876618757044
MSE train	0.06455015267694732	0.07412556725051596
MSE test	0.10997918120345535	0.11357029018169343

we could observe the following:

- R2 (accuracy) is greater for Ridge than Lasso
- MSE is less for Ridge than Lasso.
- Ridge regression makes use of all the predictors, but in lasso some of the coefficients are reduced to zero. Hence Ridge model is much more complex than Lasso.

Even though the accuracy and MSE are better for Ridge since these values are almost near based on the complexity, we could choose Lasso as it is much simpler model than Ridge.

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:

The first 5 predictors are:

- GrLivArea – 0.340265
- OverallQual - 0.207533
- BsmtFinSF1 - 0.114922
- Neighborhood_NridgHt - 0.104913
- HouseAge (YrSold-YearBuilt) - 0.089145

On dropping these 5 predictors from the data set and rebuilding the Lasso regression model we can see that the new 5 most important predictors are as follows:

- 2ndFlrSF - 0.300617
- 1stFlrSF - 0.295888
- TotalBsmtSF - 0.183027
- BsmtQual_TA - 0.158782
- BsmtQual_Gd - 0.121876

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:

A model can be said robust and generalisable only if

- It performs well on an unseen data. Performance on training data will not give a concrete idea about how the model will perform when exposed to unseen data.
- Overfitting of the model on the train data needs to be avoided. i.e the model should not learn the exact pattern from the train data.
- The model should be simpler enough to understand the pattern and the variance should be low, a slight increase in bias can provide significant reduction in variance.

The Accuracy of the model can give us an idea whether the model is robust and generalisable. The difference in the accuracy b/w the train data and the unseen test data should be less. This tells us that the model has not overfit itself on the training data, rather it keeps its accuracy on unseen test data as well.