

Q1. 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:

The optimal value for Ridge was 5.0 and for Lasso was 0.001.

The model will start overfitting and the accuracy will keep on decreasing.

For ridge regression, the alpha will be 10.0 now, most important variables are:

**['2ndFlrSF', '1stFlrSF', 'LotArea', 'Neighborhood_CollgCr',
'Neighborhood_IDOTRR']**

For lasso regression, the alpha will be 0.002 now, most important variables are:

**['Neighborhood_IDOTRR', 'BsmtQual', '1stFlrSF', '2ndFlrSF',
'Neighborhood_CollgCr']**

Q2. 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:

Even though the R2_score were nearly same for both Ridge and Lasso but it was greater for Lasso so I will chose Lasso for my final model and also Lasso penalizes more for each error and can help in feature elimination to reduce the complexity.

Q3. 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:

Now the five most important predictors will be:

['GrLiveArea', 'Neighborhood_CollgCr', 'LowQualFinSF', 'BsmtFullBath', 'BsmtFinSF2']

Q4. 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:

By regularisation we can keep the overfitting in control and hence increasing a little bit of bias but that reduces the complexity and thus the model is more robust and generalisable.

Even though there will be some dip in the accuracy overall because of reducing the complexity as the model will not capture unnecessary patterns i.e. noise with itself and also highly accurate model will not perform well on unseen data.