

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: : In ridge regression model:- When we plot the curve between negative mean absolute error and alpha we see that as the value of alpha increases from 0, the error term decreases and the train error is showing an increasing trend when value of alpha increases. When the value of alpha is 2 the test error is minimum, so decided to go with the value of alpha equal to 2 for our ridge regression model.

For lasso regression, decided to keep very small value that is 0.01, when we increase the value of alpha the model try to penalize more and try to make most of the coefficient value zero. Initially it came as 0.4 in negative mean absolute error and alpha. When we double the value of alpha for our ridge regression, now we will take the value of alpha equal to 10 the model will apply more penalty on the curve and try to make the model more generalized that is making model more simpler and doesn't fit every data of the data set. The most important variable after the changes has been implemented for **Ridge Regression** are as follows:- 1. MSZoning_FV 2. MSZoning_RL 3. Neighborhood_Crawfor 4. MSZoning_RH 5. MSZoning_RM 6. SaleCondition_Partial 7. Neighborhood_StoneBr 8. GrLivArea 9. SaleCondition_Normal 10. Exterior1st_BrkFace

The most important variable after the changes has been implemented for **Lasso Regression** are as follows:- 1. GrLivArea 2. OverallQual 3. OverallCond 4. TotalBsmtSF 5. BsmtFinSF1 6. GarageArea 7. Fireplaces 8. LotArea 9. LotArea 10. LotFrontage

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: The Ridge model performance is better than Lasso based on their r^2 scores. The train and the test scores are matching well. Will proceed with Ridge Regression

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: Those 5 most important predictor variables that will be excluded are :- 1. GrLivArea 2. OverallQual 3. OverallCond 4. TotalBsmtSF 5. GarageArea

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: The model should be as simple as possible, though its accuracy will decrease but it will be more robust and generalizable. It can also be understood using the Bias-Variance trade-off. The simpler the model the more the bias but less variance and more generalizable. Its implication in terms of accuracy that a robust and generalizable model will perform equally well on both training and test data i.e. the accuracy does not change much for training and test data.