

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 our Ridge Regression, the alpha value of 2 is quite optimal.

With this alpha value we get R2 values of 0.9355272112524963 in the training and 0.9008237523486304 in the test.

Regression	Alpha	R2 Train	R2 Test
Lasso	0.01	0.8869690265467509	0.8747706716572385
Lasso	0.02	0.8730268618276438	0.8619966740098093
Ridge	2	0.9355272112524963	0.9008237523486304
Ridge	4	0.9335335112615245	0.9020913837926449

The important predictor variables after the change has been implemented for ridge regression are:

- Exterior1st_BrkFace
- GrLivArea
- MSZoning_FV
- MSZoning_RH
- MSZoning_RL
- MSZoning_RM
- Neighborhood_Crawfor
- Neighborhood_StoneBr
- SaleCondition_Normal
- SaleCondition_Partial

The important predictor variables after the change has been implemented for lasso regression are:

- BsmtFinSF1
- Fireplaces
- GarageArea
- GrLivArea
- LotArea
- LotFrontage
- OverallCond
- OverallQual
- TotalBsmtSF

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: As per studies, we have come to understand that a model should be simple and should produce consistent prediction results.

If we are to go by the R² values, we could go with Ridge as it has higher R² values compared to Lasso.

Here we have a large number of features to select and lasso makes selection of features and drives the coefficients to zero unlike ridge. Lasso generates a simple model with few features and is more interpretable than Ridge. Thus, due to having only few features that are relevant for prediction, we should choose lasso over ridge.

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: The five most important predictor variables are

- GarageArea
- GrLivArea
- OverallQual
- OverallCond
- TotalBsmtSF

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: In order to make a robust and generalisable model, some of the things that we need to take care of in the model building process are

- Make sure the data we have is diverse and has real world scenarios
- Cross Validation
- Bias Variance trade-off, which has regularization, hyperparameter tuning as well
- Make the model as simple as possible, but not make it too simple as it can sometimes affect its accuracy