



Question1) 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 lambda for Ridge Regression = **10**
- Optimal value of lambda for Lasso = **0.001**

Now after doubling the value of alpha i.e. 20 for Ridge and 0.002 for lasso. We can see following changes are occurred given below:

Changes in Ridge Regression metrics:

- R2 score of train set decreased from 0.94 to 0.93
- R2 score of test set remained same at 0.93

Changes in Lasso metrics:

- R2 score of train set decreased from 0.92 to 0.91
- R2 score of test set decreased from 0.93 to 0.91

So, the most important predictor variables after we double the alpha values are:-

- GrLivArea
- OverallQual_8
- OverallQual_9
- Functional_Typ
- Neighborhood_Crawfor
- Exterior1st_BrkFace
- TotalBsmtSF
- CentralAir_Y



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

- We have to choose optimal value of lambda i.e. we can't choose higher value of lambda because it causes overfitting and also can't choose lowest value of alpha/lambda due to underfitting
- Ridge and lasso regression both are useful for handling overfitting issues but if we want to do feature selection Lasso is best
- Lasso regression makes coefficient reducing to zero while ridge makes value of coefficient nearer to zero (not exactly zero) if we don't want too much reduction in coefficient Ridge is best else lasso is good.
- Also we have to choose model depend on problem domain

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

- Here, we will drop the top 5 features in Lasso model and build the model again.

Top 5 Lasso predictors

were: OverallQual_9, GrLivArea, OverallQual_8, Neighborhood_Crawfor and Exterior1st_BrkFace

After dropping our top 5 lasso predictors, we get the following new top 5 predictors:-

- 2ndFlrSF
- Functional_Typ
- 1stFlrSF
- MSSubClass_70
- Neighborhood_Somerst



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 is robust when any data variation has little effect on its performance.
- A model that is generalizable is able to properly adapt to new data that has never been seen before and comes from the same distribution as the one that was used to create the model.
- We must ensure that a model does not overfit in order to ensure that it is robust and generalizable. This is due to the fact that an overfitting model has a very high variance and that even the smallest data change has a big impact on the model prediction. This kind of model will pick up all the patterns in training data, but it won't pick up the patterns in test data that haven't been seen.
- To put it another way, for the model to be robust and generalizable, it should not be too complicated.