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 for alpha for Ridge is 10 and Lasso is 0.001. Doubling the value of alpha for both ridge and lasso will make coefficients lower. As per the analysis, the most important predictor variables will still remain the same i.e. SaleCondition_Partial, SaleCondition_Normal, SaleCondition_Others, GarageFinish_Unf

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: Lasso regression would be chosen as less important feature coefficients become zero.

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: GarageType_BuiltIn, GarageType_Detchd, GarageType_NoGarage,GarageType_Others

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: The test accuracy should not be lesser than the training score, this is one way that we can make sure that the model is robust and generalizable. The outlier management is another area which needs to be kept in mind. The model should not be changed by outliers in the training data. Also the outlier analysis needs to be done properly.

The implications if the model is not robust are huge. The predictive analysis cannot be trusted if the model is not generalizable. This means that the analysis goes for a toss.