Surprise Housing Assignment Part-2

1. **What is the optimal value of 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?**
   - The optimal value of Ridge Regression is 20 and the optimal value of Lasso Regression is 0.001.
   - If we double the value of alpha for both ridge and lasso, Ridge co-efficient values are lowered, whereas the less valued coefficients in lasso becomes zero
   - The most important variables after the change is implemented are:
     SaleCondition_Partial
     SaleCondition_Normal
     SaleCondition_Family
     SaleCondition_Alloca
     SaleCondition_AdjLand

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?**
   - Both Lasso and Ridge Regression Models are decent enough with almost same evaluation metric values but there are more fluctuations in considering variables in Ridge Regularization since it deals with zeroing coefficients. Thus, I will go with **Lasso Regression.**

3. **After building the model, you realized 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 variables. Which are the five most important predictor variables now?**
   - SaleType_WD
   - SaleType_Oth
   - SaleType_New
   - SaleType_ConLw
   - SaleType_ConLI

4. **How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?**
   - In order to have a robust model we need to treat the outliers carefully and clean the data with the most relevant data, also while hyper parameters like alpha in case of Regularization or test-train split are chosen so as to avoid any underfitting/overfitting. Proper metrics for whose the total metric score (testing and training scores) is minimal is chosen considering the bias-variance tradeoff. This way we can build a robust and generalized model with a good accuracy