ADVANCED REGRESSION ASSIGNMENT PART II
SUBJECTIVE QUESTIONS

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:

Optimal Value of alpha for Ridge Regression: **6.0**

Optimal Value of alpha for Lasso Regression: **0.0001**

Doubling the value of Alpha:

- There was no change in the model parameters / coefficients, its just that the R-Squared, and coefficients of Betas have changed after doubling the model
- **Ridge Regression (alpha=12):**
    - **R-squared Value on training data: ~92.3%**
    - **R-squared Value on test data: ~87.8%**
    - **Important Predictor Variables (top 5) with Beta values:**
        - `('SaleCondition_Partial', 0.1003)`
        - `('SaleCondition_Others',0.0806)`
        - `('SaleCondition_Normal', 0.0799)`
        - `('GarageFinish_Unf', 0.0719)`
        - `('GarageFinish_RFn', 0.0683)`
- **Lasso Regression (alpha= 0.0002):**
    - **R-squared Value on training data: ~93.3%**
    - **R-squared Value on test data: ~88.3%.**
    - **Important Predictor Variables (top 5) with Beta values:**
        - `('SaleCondition_Partial', 0.2438)`
        - `('SaleCondition_Others', 0.2237)`
        - `('SaleCondition_Normal', 0.201)`
        - `('GarageFinish_RFn', 0.1963)`
        - `('GarageFinish_Unf', 0.1876)`

ADVANCED REGRESSION ASSIGNMENT PART II
SUBJECTIVE QUESTIONS

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 the results ridge lasso looks to have similar metrics.

Ridge: Alpha = 6, Train R-Squared=93.3%, Test R-Squared=88.0%

Lasso: Alpha=0.0001, Train R-Squared=94.0%, Test R-Squared=88.4%

As per the results both models look fine, as both resulted in good R-Squared scores. Lasso has slightly higher values of R-Squared for both train, test set. As Lasso does automatic feature selection, making coefficients of lesser important features to zero, Lasso can be chosen.

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: After removing the top 5 variables and rerunning the model.

- **Ridge Regression (alpha=5), top 5 predictors with Beta values:**
  - ('GarageFinish_RFn', 0.1588)
  - ('GarageFinish_NoGarage', 0.1287)
  - ('GarageType_NoGarage', 0.1232)
  - ('GarageType_Detchd', 0.1004)
  - ('GarageType_CarPort', 0.0997)
- **Lasso Regression (alpha=0.0001), top 5 predictors with Beta values:**
  - ('GarageFinish_RFn', 0.3465)
  - ('GarageFinish_NoGarage', 0.3373)
  - ('GarageType_NoGarage', 0.3222)
  - ('GarageType_Detchd', 0.2893)
  - ('GarageType_CarPort', 0.2052)

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?

ADVANCED REGRESSION ASSIGNMENT PART II
SUBJECTIVE QUESTIONS

Ans: The model should be simple in general, to that it can be generalisable, and robust.

- **Generalisable model**, can be able to adopt properly for unseen dataset**.**
- **Robust Model**, ensure performance / results are not affecting much, thought there is variation in the data.
- To make sure model is **robust, generalisable**, we need to ensure:
    - Model doesn't cause **overfitting.** As an overfitting model has high variance. Overfitting is an undesirable behaviour as it gives accurate predictions for training data but not for new data.
    - **Bias, Variance Trade off**:
        - Bias (Correctness of model) refers to how much error does the model make in test data. Variance(Consistency of model) refers to how sensible is the model to input data.
        - Simple model might have higher Bias, but lower variance. So, this leads to underfitting. Complex models might have lower Bias, and higher variance, causing overfitting. So, it is important to have balance between both to avoid overfitting, underfitting of data.
- **Accuracy of Model**:
    - Complex models generally will have high accuracy. For making model to be more generalisable, robust, need to reduce the variance, which incurs to some additional bias. This helps in having an optimal model, this mean accuracy will be reduced.
    - The balance between accuracy, and complexity can be achieved with help of **regularization techniques** such as Ridge Regression, Lasso Regression.