Research Question: Is Ridge Regression or LASSO a better model to predict the price of houses in Ames, Iowa?

Project Title: Predicting House Prices Using Ridge and Lasso Regression

Project Goal:

The primary objective of this project is to develop a predictive model for housing prices in Ames, Iowa, utilizing ridge and lasso regression techniques. By applying these regularization methods, we aim to enhance the model's accuracy and interpretability, effectively addressing issues such as multicollinearity and overfitting. Our goal is to determine which property features significantly influence house prices and to assess the performance of ridge and lasso regression in this context.

Machine Learning Methods:

We will employ two regularization techniques:

1. **Ridge Regression**
2. **Lasso Regression**

Data Source:

"House Prices: Advanced Regression Techniques" dataset from Kaggle. This dataset comprises 79 explanatory variables detailing various aspects of residential properties in Ames, Iowa, collected between 2006 and 2010. The features include property size, age, condition, among others. We will choose the 20 variables that will think will be the most informative predictors to reduce the scope of the project.

https://www.kaggle.com/competitions/home-data-for-ml-course/data

Analysis:

- Which property features are most significant in predicting house prices in Ames, Iowa?
- How do ridge and lasso regression techniques compare in terms of predictive performance and feature selection?
- What are the optimal regularization parameters for each method to achieve the best model performance?

Expected Outcomes:

We anticipate that lasso regression will effectively perform feature selection by shrinking less important coefficients to zero, thereby simplifying the model without compromising accuracy. Ridge regression is expected to handle multicollinearity effectively, providing stable coefficient estimates. We will expect lasso to work better on our data set due to the number of redundant features (e.g. flatness of property and slope of property are both included), as well as seemingly insignificant features (e.g. roof material, pool quality, fireplace, etc.). Lasso will shrink these less important coefficients to zero and simplify the final model, leading to better predictive power.