Lab Assignment 2 - Regression

Objective: Constructing linear and polynomial regression models and analysing the performance of these models.

1. Load Boston housing data from sklearn (sklearn.datasets.load_boston). Dataset description https://www.kaggle.com/prasadperera/the-boston-housing-dataset
2. Do the exploratory analysis of the dataset (i.e. univariate and multivariate analysis) and plot the graphs
3. Remove any undesirable feature from the dataset and scale the remaining features.
4. Split the dataset into train and test.
5. Write a function that fits a polynomial LinearRegression model on the training subset for degrees 1, 3 and 6
6. Compare the performance of various models based on the coefficient of determinant (R^2) and RMSE.
7. Based on the above scores from step 6 (degree levels 1 through 6), what degree level corresponds to a model that is underfitting? What degree level corresponds to a model that is overfitting? What choice of degree level would provide a model with good generalization performance on this dataset? This function should return one tuple with the degree values in this order: "(Underfitting, Overfitting, Good_Generalization)"
   Note: There may be multiple correct solutions to this question.
8. Training models on high degree polynomial features can result in overly complex models that overfit, so we often use regularized versions of the model to constrain model complexity, as we saw with Ridge and Lasso linear regression.

   # For this question, train two models: a non-regularized LinearRegression model (default parameters) and a regularized Lasso Regression model (with parameters `alpha=0.01`, `max_iter=10000`) on polynomial features with high degree. Return the R2 score for both the LinearRegression and Lasso model's test sets.
9. Look for real world applications where you can apply regression.