ISL Project

Instructions:

1) Select a dataset for analysis.
2) Perform Data Preprocessing on the dataset.
3) Apply statistical learning techniques to analyze the dataset.
4) Present your findings, results, and insights in a final project report or presentation.
5) Based on your findings, iterate on your project to improve model performance or explore additional questions.

Essential topics:

Regression:

- Define an appropriate regression problem based on your dataset.
- Apply multiple linear regression to create a baseline model and evaluate its performance using metrics such as mean squared error (MSE) and R-squared.
- Apply Forward Stepwise Selection, Backward Stepwise Selection and compare the resulting models to determine which one works best for the dataset and modeling goals.
- Apply PCA to the dataset to reduce its dimensionality while preserving as much information as possible!
- Apply higher-order regression model using polynomial features and compare its performance with the baseline model.
- Visualize the relationships between predictor variables and the target variable using scatter plots and regression diagnostics.
- Explore and implement advanced regression techniques, such as ridge regression, lasso regression, and compare their performance.

Classification:

- Define an appropriate multi-class classification problem based on your dataset.
- Apply logistic regression as a baseline classification model and evaluate its performance using metrics such as accuracy, precision, recall, and F1 score.
- Implement dimensionality reduction techniques (e.g., PCA) to reduce the number of features while preserving information.

- Implement a tree-based classification model (e.g., decision trees, random forests) and compare its performance with the logistic regression baseline.
- Use visualizations to illustrate the separation between different classes in the feature space.
- Explore and implement one advanced classification technique such as k-nearest neighbors, naive Bayes, etc. Compare their performance with the baseline.