CS-770 Machine Learning
Assignment 1: Linear regression

Due date: 09/26/2024

_____

• Submit a separate PDF report and code file.
• Prepare a comprehensive report documenting all aspects of the assignment.
• Ensure that your PDF report is well-structured, with clear headings, figures, and tables to present your findings effectively.
• Submit separate code files written in Python (.py format) for data preprocessing, model training, and evaluation.
• Ensure that the code is well-commented, organized, and follows best practices in programming.

_____

Q1 Perform the regression analysis using Boston Housing dataset (50 points).

Task 1: Data Exploration and Preprocessing (25 points).
• Exploratory Data Analysis (EDA): Perform a comprehensive EDA on the Boston Housing dataset. Visualize the distribution of the target variable and important features using histograms, box plots, and scatter plots.
• Feature Scaling: Discuss the importance of feature scaling in regression analysis.

Task 2: Regression Analysis (25 points).
• Model Implementation: Using the provided code as a base, implement linear, ridge, lasso, and OLS regression models.
• Model Evaluation: Evaluate each model's performance using the R^2 score and Mean Squared Error (MSE). Discuss the implications of these metrics in the context of model selection.
• Report results for simple and multiple linear regression.

Q2. Perform the regression analysis using Fish dataset (50 points).

Task 1: Linear regression (25 points).
• Fit a predictive linear regression model to estimate weight of the fish from its length, height, and width? (the data source fish.csv is provided).
• Report the coefficients values by using the standard Least Square Estimates.
• What is the standard error of the estimated coefficients, R-squared term.

Task 2: Ridge and Lasso Regression (25 points).
• Fit the Ridge and Lasso Regression Models
• Report the coefficients for Ridge and Lasso Regression models and the error of the estimated coefficients, R-squared term.