AIG100 – Project 2: Regression and Classification Methods

Overview

This project is to explore applications of regression and classification methods in solving real-world problems using machine learning. You will select datasets and apply appropriate regression and classification methods to predict outcomes and classify data points, respectively. Be reminded that this may look the same as the labs you have completed but this is more comprehensive, and expectations are higher. Moreover, please use a different dataset from your labs. If you wish to use the same dataset, then choose a different model to solve your problem. The aim here is not to redo what you have already implemented! You are free to use any tools/platforms for this project.

Objectives

- To understand and apply regression methods to predict numerical outcomes.
- To utilize classification methods to categorize data into predefined labels.
- To evaluate and compare the performance of different regression and classification models.
- To interpret and communicate the results of the analysis effectively.

Tasks

1. Dataset Selection and Objective Definition
    a. Choose a dataset for each task: one for regression and another for classification.
        - As discussed in class, each set should have an appropriate column to use as a target variable (i.e., the variable should be both of correct datatype, and of interest to predict), and useful features.
    b. Define clear objectives for what you aim to predict or classify.
2. Data Preprocessing
    a. Clean the data, handling missing values and outliers.
    b. Perform any necessary transformations to prepare the data for modeling.
3. Model Implementation
    a. For regression: Apply at least two different regression techniques (e.g., linear regression, decision trees).
    b. For classification: Implement at least two classification methods (e.g., logistic regression, SVM).
4. Model Evaluation
    a. Evaluate the models using appropriate metrics.
    b. Compare the performance of the models and discuss any findings.
5. Results Interpretation
    a. Interpret the results of your models.
    b. Create visualizations to present your findings (where possible/applicable).
6. Report Writing
    a. Document your process, findings, and conclusions in a detailed report.
    b. Besides the necessary sections, include a section on the potential implications of your results in real-world scenarios.

    c. Provide references (e.g. codes, blogs, GenAI tools).
7. Reflection
    a. Reflect on the challenges faced during the project and how they were addressed.
    b. Discuss any insights gained from applying regression and classification techniques.

(The points above explain the requirements for the project, but they do not constitute a list of topics to include in your report! For example, it will be strange to include a 'Report Writing' section… It should be obvious which elements should be report sections. Moreover, please report on each dataset separately: one dataset's entire selection/preprocessing/implementation/evaluation/interpretation/etc., before reporting all of those elements for the second dataset.)

Submission

Implement everything in a Jupyter Notebook file. Make sure that code is written in ode cells and textual explanations are written in Markdown cells. You need to run your notebook so that visualizations are saved in the submitted notebook. Submit your completed Jupyter Notebook file through the submission link in Blackboard.