Machine Learning Assignment 2
4th February 2025

Classification Using Random Forest and XGBoost

Introduction

This assignment will explore the application of Random Forest Classifier and XGBoost for classification tasks. The goal is to train machine learning models and use them to predict labels for a test dataset. Follow the instructions carefully to complete the assignment.

Objective

The main objectives of this assignment are:

- Train Random Forest and XGBoost models on given datasets.
- Use the trained models to predict labels for the test dataset.
- Submit the predicted labels in the required format for evaluation.

Instructions

1. Open the provided Jupyter Notebooks:
    - ML-A2-P1.ipynb
    - ML-A2-P2.ipynb
    - These notebooks contain skeleton code and the detailed comments
    - test dataset for Part 1 has been mailed to you all. Check your spam folder in case you haven't received it.
2. Model Implementation:
    - Use the training dataset to implement:
        - Random Forest Classifier
        - XGBoost Classifier
3. Prediction:
    - Use the trained models to predict labels for the test dataset.
    - Save the predictions in the required CSV format as mentioned in the notebook.

Submission Guidelines

- **Part 1:** Submit your predicted labels and probabilities as a CSV file. Follow the

naming format instructed in the notebook.
- Upload the required files as a single ZIP file named RollNo.zip to the submission portal. The ZIP file must contain:
    - The CSV file with predictions for Part 1
    - The Jupyter Notebook files (.ipynb) for Parts 1 and 2
    - Predicted probabilities (.npy) for Part 1
- **NOTE:** No CSV file with predicted labels is required for Part 2.
- All files should be in the same zip, don't create additional directories.
- **Deadline for Submission: 15th February 2025 (EOD)**
- **Important: Ensure your files follow the required naming convention to avoid penalties.**

Evaluation Criteria

Your submission will be evaluated based on:

- Accuracy of predictions on the test dataset.
- Correct implementation of the models in the provided Jupyter Notebook.
- Adherence to the submission format.

Important Note

- Plagiarism in any form will result in serious consequences. Ensure that all work submitted is your own.
- Mail to saransh03sharma@gmail.com for Part-1 and Bismay Parija parijabismay@gmail.com for Part-2 doubts.