Machine Learning

ASSIGNMENT 2: Naive Bayes Classifier

Date: October 3, 2024

Deadline: October 16, 2024

Total Credit: 10

- Markings will be based on the correctness and soundness of the outputs.
- Marks will be deducted in case of plagiarism.
- Make proper indentation and appropriate comments.
- You are required to submit 2 separate files, 1 for code (in .py format) and 1 for documentation (in .pdf format). The code file should only contain the assignment code, and the documentation file should contain the responses to the assignment questions. Do not add your whole code to the documentation file. Code snippets for reasoning can be added to the documentation file but are not mandatory.
- For submission, name your files using roll number(s) of all the group members. Ex: roll1_roll2_roll3.py/pdf. There can be at most three members in the group.
- **Note: Do not send your files to us by email.**

Problem Statement: You are to design and implement a binary classification module to determine whether the data point is ‘good’ or ‘bad’. Corresponding to column name ‘Class’ in the dataset.

Dataset Details: The Ionosphere dataset contains features obtained from radar signals focused on the ionosphere layer of the Earth's atmosphere. The task is to determine whether the signal shows the presence of some object, or just empty air. For further details visit Link .

Dataset: Link .

Implementation Details:

1. Design and implement a Naive Bayes module from scratch. Use a train vs test split of 80:20.
2. Report accuracy, Precision, Recall and F1 score.
3. Draw the confusion matrix to get an idea about the types of misclassified examples.

4. Now, take the readily available implementation from the Sklearn library and compare with your own implementations in terms of accuracy; Precision, Recall, and F1Score; as well as the confusion matrix.

Files to submit:

- Code
- Detailed document describing the implementation details, results and the time taken for the execution.
- Analysis of comparison between the two implementations.