Cairo University, Faculty of Computers
and Artificial Intelligence

FACULTY OF COMPUTERS AND AI, CAIRO UNIVERSITY

Supervised Learning
Year 2024-2025

Second Semester

Assignment #3

Supervised Learning
Assignment 3

Cairo University, Faculty of Computers and Artificial Intelligence

Assignment #3

Objective

The goal of this assignment is to implement a speech classification system using a Naïve Bayes classifier from scratch, compare its performance with built-in Logistic Regression, and enhance the results using bagging ensemble learning.

Instructions

1. Dataset Preparation

- **Obtain a speech dataset consisting of two classes (e.g., male vs. female, spoken digits, or any other binary classification task).**
- **Provide a brief description of the dataset, including the number of samples, features, and class distribution.**

2. Preprocessing and Feature Extraction

- **Perform data preprocessing steps such as:**
    - **Noise reduction**
    - **Silence removal (if applicable)**
    - **Normalization**
- **Extract relevant features from the speech signals, such as:**
    - **Mel-Frequency Cepstral Coefficients (MFCCs)**
    - **Spectral features (e.g., Spectral Centroid, Spectral Rolloff)**
    - **Zero-Crossing Rate**
    - **Other relevant features**

3. Implement Naïve Bayes Classifier from Scratch

- **Implement a Naïve Bayes classifier without using prebuilt machine learning libraries.**
- **Use the Gaussian Naïve Bayes approach to handle continuous features.**
- **Train the classifier on the extracted features.**
- **Evaluate the model using appropriate metrics (e.g., accuracy, precision, recall, F1-score).**

Cairo University, Faculty of Computers
and Artificial Intelligence

4. Apply Bagging with Naïve Bayes and Logistic Regression

- **Implement a bagging ensemble method using multiple instances of your Naïve Bayes classifier.**
- **Implement bagging ensemble using Logistic Regression.**
- **Train multiple models on different subsets of the dataset.**
- **Aggregate predictions using majority voting.**
- **Compare the performance of the ensemble methods with individual models.**