Machine Learning Assignment - 1

Instructions.

- **This is an easy assignment which delves into the realm of multimodal data analysis, focusing on the analysis of audio, image and time-series data.**

- **Use [Google Colab / Jupyter Notebooks] only for this assignment. Not doing so will lead to (0 Marks) in complete assignment.**

- **Each question has its own instruction, follow precisely. If not (0 Marks) will be awarded.**

- **Submission Guidelines**

    - **Submit only a single [MLA1_*.ipynb] where * is your roll number.**
    - **For eg: if your roll number is [MT10002], your file name should be [MLA1_MT10002.ipynb]**

- **Not following naming convention will lead to (0 Marks) , whatever the reason you present.**

- **You are allowed to use only [*pandas, numpy, PIL, matplotlib, scipy*] for this assignment. Again not following the same will lead to (0 Marks) for entire assignment.**

- **If you are not able to explain in viva, irrespective of your answer correctnes you will get (0 Marks)**

- **If there is ● in front of question you have to implement it completely from scratch and if it is ● you can use functions from above mentioned packages.**

[30 Marks]

Dataset download link.

Dataset Description: A dataset comprising electroencephalography (EEG) signals, corresponding visual stimuli, and auditory data from human participants is provided for analysis. EEG data were collected at a sampling rate of 128 kHz using a five-electrode configuration (AF3, AF4, T7, T8, Pz) while participants viewed and cognitively processed visual stimuli for three-second intervals. The original EEG data, in .eeg format, has been converted to .csv for convenience and is accompanied by the respective image and audio files.

- **Beginner**

    1. ● **(2 Marks)** Make a random number generator which takes input as size and generate random numbers. Using this generate and plot random numbers of size:
        (a) 1×n
        (b) n×n

    2. ● **(1 Marks)** Using above implemented random number generator select 12 samples from the EEG data and print there metadata.

    3. ● **(3 Marks)** Implement a normalization function which can normalize given data in range [-x, x]. Normalize and plot 4 randomly selected EEG signals.

    4. **(4 Marks)** Use selected EEG signal from previous question and perform on:
        - Image Data:
            (a) ● Reshape image to (3,224,224).
            (b) ● Convert to Black&White.
            (c) ● Plot histogram of the image (Color & BW).
            (d) ● Calculate CenterOfMass for all images.
        - Audio Data:
            (a) ● Resample audio to 16000.
            (b) ● Convert to MelSpectrogram.
            (c) ● Plot time-domain & frequency-domain signal.
            (d) ● Calculate ZeroCrossingRate for all time-domain signals.

    5. ● **(3 Marks)** Calculate and plot PowerSpectralDensity (PSD) of randomly selected 4 EEG signals.

- **Intermediate**

  1. 🔴 **(4 Marks)** Calculate and plot Cross-Correlation and Auto-Correlation of randomly selected 4 EEG signals.
  2. 🟢 **(4 Marks)** Perform exploratory data analysis (EDA). (Visualization, statistical analysis, class imbalance, features visualization, outliers detection etc.)
     (a) Image data.
     (b) Audio data.
  3. 🔴 **(4 Marks)** Select features from your EDA and perform principal component analysis (PCA) on image and audio feature data separately. Plot the feature map before and after performing PCA.

- **Advanced**

  1. 🔴 **(5 Marks)** Randomly select 120 EEG samples and form a feature-bank for image, audio, and EEG data separately. Write a pipeline for selecting data, making train-test splits and training-testing classification using linear regression as classification algorithm.