Report

Github repo - https://github.com/saikrishnareddysanda/SU_PA_2/tree/main

Q1

Tasks

- Chosen the following pretrained models that are trained on VoxCeleb1 dataset
    - Wavlm-base-sv
    - unispeech-sat-base-sv
    - Spkrec-ecapa-voxceleb

- Model loading
    - Wavlm-base-sv - Model is loaded using Huggingface transformers library. (https://huggingface.co/microsoft/wavlm-base-plus-sv)

    - Unispeech-sat-base-sv - Model is loaded using Huggingface transformers library. (https://huggingface.co/microsoft/unispeech-sat-base-plus-sv)

    - Spkrec-ecapa-voxceleb - Model is loaded using Speechbrain library. (https://huggingface.co/speechbrain/spkrec-ecapa-voxceleb)

- Dataset Loading
    - A subset of Voxceleb1 dataset is uploaded to my drive and loaded in colab for evaluation.

- EER Calculation
    - For the test subset pairs that are present in the dataset, Similarity scores for each of the models are calculated.
    - Results
        - Wavlm-base-sv - 0.166
        - Unispeech-sat-base-sv - 0.166
        - Spkrec-ecapa-voxceleb - 0.50

- Comparison with results from WavLM paper. Note: Whole dataset is used for this evaluation whereas in previous only a subset is used.
    - Wavlm-base-sv - 1.75
    - Unispeech-sat-base-sv -
    - Spkrec-ecapa-voxceleb - 2.127

- Evaluation on Kathbath dataset
    - The data is directly downloaded and extracted.
    - In this evaluation, telugu language is chosen.
    - To convert files from .m4a to .wav and create the data in a right format, the following code is used. (https://github.com/AI4Bharat/IndicSUPERB/blob/master/utilities/structure.py)
    - Pairs of the same speaker and different speakers are created randomly.
    - Results
        - Wavlm-base-sv - 0.1
        - Unispeech-sat-base-sv - 0.083
        - Spkrec-ecapa-voxceleb - 0.066