HW1P2 IDL

- Author Name: Nora Ngo
- AndrewID: nngongoc
- Kaggle Username: norango

Description:

This homework involved building a Multi-Layer Perceptron (MLP) network for phoneme detection and classification. Using the provided dataset, the goal was to classify phonemes into 42 distinct classes using a trained model.

Requirements:

Given in the requirements.txt file

Dataset:

The dataset used is the MFCC dataset provided by the professor and can be found in the given Kaggle competition: https://www.kaggle.com/competitions/11785-spring-25-hw-1-p-2

Model Architecture

- The model follows a **pyramid MLP architecture** with approximately **19.88M parameters**.
- It comprises 32 layers. There are 9 linear layers, 9 batch normalization layers, 8 Mish activation layers, 7 dropout layers
- **Activation function**: Mish
- **Optimizer**: AdamW
- **Scheduler**: CosineAnnealingWarmRestarts
- **Weight initialization**: Xavier_Normal
- **Dropout**: 0.25
- **Batch size**: 8162
- **Epochs**: 30
- **Augmentations**: Both (Frequency Masking and Time Masking)
- **Masking parameters**: Frequency Mask = 5, Time Mask = 10
- **Training time**: Approximately 5-6 hours

Experimentation:

For given HW1P2, 57 ablations were run until an accuracy of **86.811%** accuracy was achieved. To summarize the given studies:

- **(1 - 30 Ablations):**
  - Initial experiments focused on using GELU activation and tuning the number of perceptrons from 1024-4096 for input layers and 512-64 for output layers.
  - Network sizes ranged between 4M to 10M parameters.
  - Achieved **~75% validation accuracy**.
- **(31 - 40 Ablations):**
  - Fix the code for the dataset classification
  - Validation accuracy improved to **~83%**.
  - Network sizes ranged between 4M to 10M parameters.
- **(41 - 50 Ablations):**
  - Switch to Mish activation and tune the number of perceptrons to 8162.
  - Network sizes ranged between 4M to 10M parameters.
  - Validation accuracy stagnated at **~84%**.
- **(51 - 57 Ablations):**
  - Final models focused on a **pyramid architecture with Mish activation**, simplifying earlier complex designs.
  - Achieved **86.8% validation accuracy**, the best observed.
  - Network sizes ranged around **19 parameters.**
  - Training loss and validation loss showed stable convergence patterns, as seen in WandB logs.
  - The learning rate schedule with CosineAnnealingWarmRestarts contributed to smoother convergence.

Key Takeaways from the Updated Studies:

- **Best-performing model**: Pyramid MLP, Mish activation, AdamW optimizer, CosineAnnealingWarmRestarts scheduler.
- **Training strategy**: Running for **30+ epochs** significantly improved results.
- **Avoid overly complex gating mechanisms**: Simple architectures performed better.
- **A batch size of 8162** worked optimally without convergence issues.

Training

Instructions on how to train the model.

- For training, just run the attached .ipynb file in Google Colab and ensure a "saved_models" folder is made for the given runtime.
- Also, ensure that the given Wandb part of the ipynb is renamed for the new training model for checkpoints.

Evaluation

How to test the trained model.

- Directly run the code after the "Experimentation" section of the ipynb file, ensuring that "saved_models" has the wandb .pth file of the model for testing.

Inference

Steps to use the trained model for predictions.

- Predictions can be checked directly through the generated "submission.csv" file after running the .ipynb file.

Notes

- The WandB link for the updated ablation studies: https://wandb.ai/ngon78790-carnegie-mellon-university/hw1p2/workspace?nw=nwusernngon78790

Acknowledgment

- I would like to acknowledge my study group members: Aditya Sannabhadti for helping me reach the initial 83% accuracy by suggesting model architecture, parameter configuration changes, and dataset class, and Phoebe Zhou for helping me change the model architecture.