
Machine Learning HW2

ML TAs

mlta-2022-spring@googlegroups.com

Outline

- Task Introduction
- Dataset & Data Format
- Submission & Grading

Task Introduction

Task Introduction

1. Data Preprocessing: Extract MFCC features from raw waveform (already done by TAs!)
2. Classification: Perform framewise phoneme classification using pre-extracted MFCC features

Task Introduction

Task: Multiclass Classification

Framewise phoneme prediction from speech.

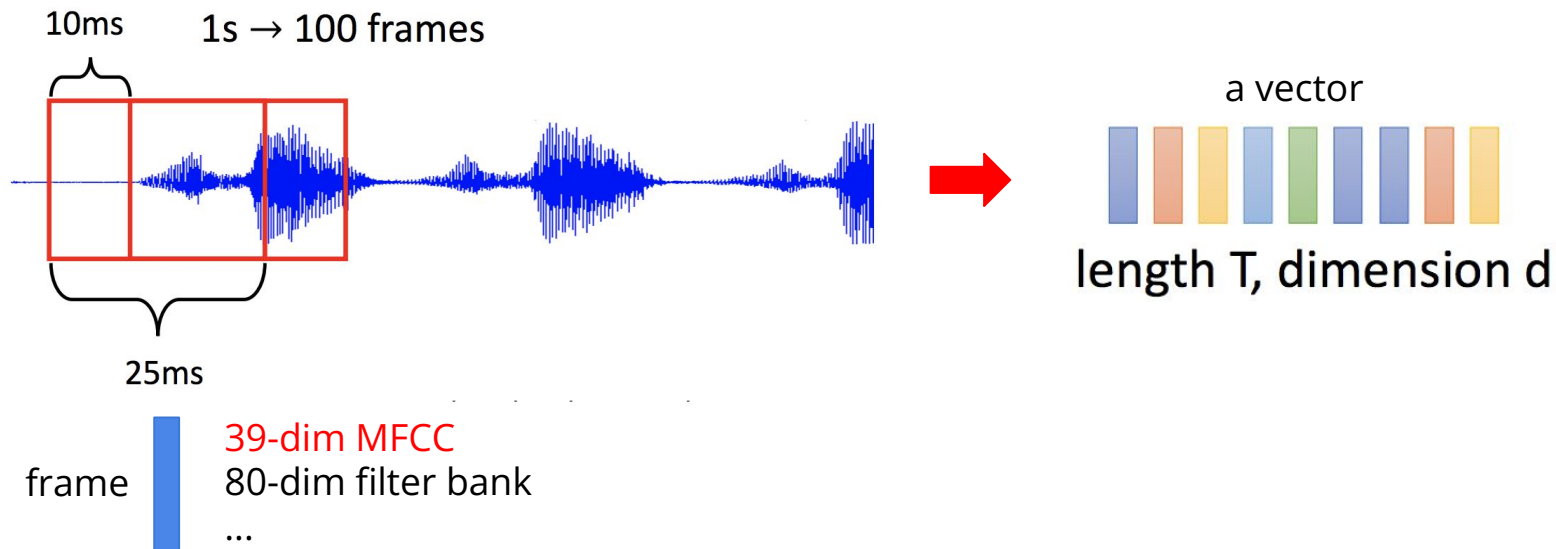


What is a phoneme?

A unit of speech sound in a language that can serve to distinguish one word from the other.

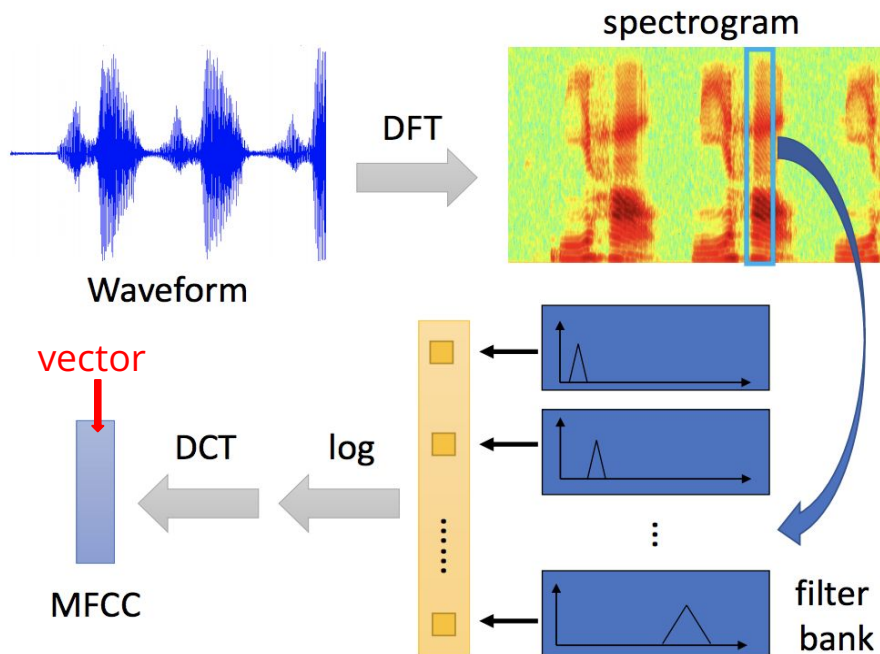
- bat / pat , bad / bed
- Machine Learning → M AH SH IH N L ER N IH NG

Data Preprocessing



Data Preprocessing

Acoustic Features - MFCCs (Mel Frequency Cepstral Coefficients)



For more details,
please refer to Prof. Lin-Shan Lee's
[\[Introduction to Digital Speech Processing\]
Chap.7](#)

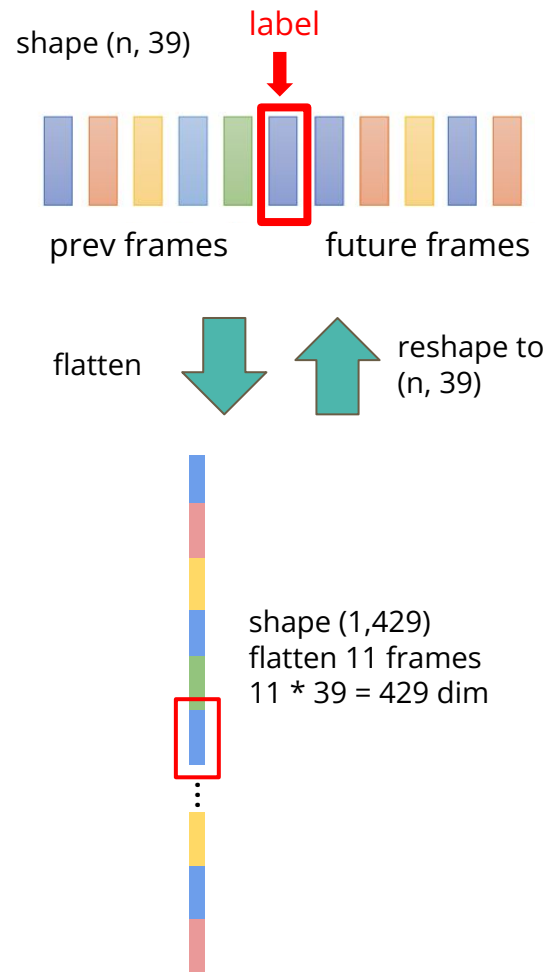
Image ref.
Prof. Hung-Yi Lee
[\[2020Spring DLHLP\] Speech Recognition](#)

More Information About the Data

Since each frame only contains 25 ms of speech, a single frame is unlikely to represent a complete phoneme

- Usually, a phoneme will span several frames
- Concatenate the neighboring phonemes for training

Finding testing labels or doing human labeling are strictly prohibited!



Dataset & Data Format

Dataset

LibriSpeech (subset of train-clean-100)

- Training: 4268 preprocessed audio features with labels (total 2644158 frames)
- Testing: 1078 preprocessed audio features (total 646268 frames)
- Label: 41 classes, each class represents a phoneme

Data Format

- Data Format (The TAs have already extracted the features)

libriphone/

- train_split.txt (train metadata)
- train_labels.txt (train labels)
- test_split.txt (test metadata)
- feat/
 - train/
 - test/

train_split.txt

```
1 2007-149877-0023
2 60-121082-0044
3 5688-41232-0018
4 1898-145702-0018
5 307-127539-0028
6 6147-34607-0009
7 7226-86965-0049
8 6078-54007-0024
9 5688-15787-0042
10 5808-48608-0055
```

train_labels.txt

```
1 2007-149877-0023 0 0 0 0 0 0 0 0 0
2 60-121082-0044 0 0 0 0 0 0 0 0 0
3 5688-41232-0018 0 0 0 0 0 0 0 0 0
4 1898-145702-0018 0 0 0 0 0 0 0 0 0
5 307-127539-0028 0 0 0 0 0 0 0 0 0
6 6147-34607-0009 0 0 0 0 0 0 0 0 0
7 7226-86965-0049 0 0 0 0 0 0 0 0 0
8 6078-54007-0024 0 0 0 0 0 0 0 0 0
9 5688-15787-0042 0 0 0 0 0 0 0 0 0
10 5808-48608-0055 0 0 0 0 0 0 0 0 0
```

features: 39-dim MFCC w/ CMVN
{filename}.pt for each utterance(audio)

```
└─ feat
   └─ test
      └─ train
         19-198-0008.pt
         19-227-0070.pt
         26-495-0000.pt
         26-495-0007.pt
         26-495-0017.pt
```

Using additional data is prohibited. Your final grade will be multiplied by 0.9!

Data Format

- Each .pt file is extracted from one original wav file
- Use torch.load() to read in .pt files as torch tensors
- Each tensor has a shape of (T, 39)

39 dims

T frames

```
tensor([[ -0.9555, -0.9062,  0.9451, ..., -1.4516, -1.5912, -1.3270],  
        [ -0.9434, -0.9633,  0.7211, ...,  0.1566, -0.0150, -0.1353],  
        [ -0.8907, -0.9749,  0.6556, ...,  1.1867,  0.4603, -0.0459],  
        ...,  
        [-1.0778, -0.7979,  0.8335, ...,  0.6452, -0.3527, -0.7415],  
        [-1.1911, -1.0670,  0.6462, ...,  0.3025, -0.6755, -0.9707],  
        [-1.1044, -1.0259,  0.7016, ..., -0.1956, -0.4646, -0.5964]])
```

Submission & Grading

Submission & Grading

- Leaderboard (4%): Kaggle
- Code submission (2%): NTU COOL
- Report submission (4%): Gradescope

Kaggle Public Baselines

- (1%) Simple baseline: 0.45797 (sample code)
- (1%) Medium baseline: 0.69747 (concat n frames, add layers)
- (1%) Strong baseline: 0.75028 (concat n, batchnorm, dropout, add layers)
- (1%) Boss baseline: 0.82324 (sequence-labeling(using RNN))

For boss baseline, you can refer to pervious course recording of RNN: [video 1](#) and [video 2](#).

Kaggle Submission

Kaggle Link: <https://www.kaggle.com/c/ml2022spring-hw2>

- Displayed name: **<student ID>_<anything>**
 - e.g. b06901020_puipui
- You do NOT have to change your account name, just modify “team name” under the “team” tab
- Submission format: **.csv** file
- Evaluation metric: accuracy
- Submission deadline:
 - **2022/3/18 23:59 (UTC+8)**

```
1 Id,Class
2 0,0
3 1,0
4 2,0
5 3,0
6 4,0
```


Kaggle Submission

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#)

Manage Team

Team Name

[Save Team Name](#)

This name will appear on your team's leaderboard position.

Kaggle Submission

- You may submit up to **5** results each day (UTC).
- Up to **2** submissions will be considered for the private leaderboard.

prediction_large.csv 2 years ago by ntuee_jizz model_large3_684_compressed.pth, size = 201KB, params: 93139 (rabbit ensemble)	0.65059	0.66341	<input checked="" type="checkbox"/>
prediction_large.csv 2 years ago by ntuee_jizz model_large3_676_compressed.pth, size = 201KB, params: 93139 (rabbit ensemble)	0.65282	0.65422	<input type="checkbox"/>
prediction_large.csv 2 years ago by ntuee_jizz model_large2_669_compressed.pth, size = 222KB, params: 103623	0.65394	0.65254	<input checked="" type="checkbox"/>

remember to select **2**
results for your final
scores before the
competition ends!

Code Submission (2%)

- Compress your code and report, then submit it to NTU COOL.

<student ID>_hw2.zip

e.g. b06901999_hw2.zip

- We can only see your last submission.
- Do not submit your model or dataset.
- If your code is not reasonable, your final grade will be multiplied by 0.9!
- Submission deadline:
 - **2022/3/18 23:59 (UTC+8)**

Report Questions

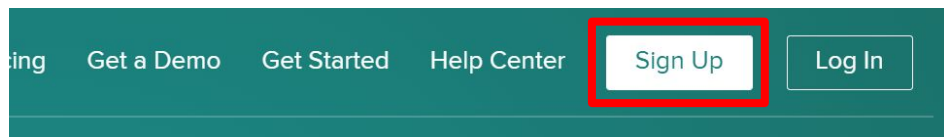
1. (2%) Implement 2 models with approximately the same number of parameters, (A) one narrower and deeper (e.g. hidden_layers=6, hidden_dim=1024) and (B) the other wider and shallower (e.g. hidden_layers=2, hidden_dim=1700). Report training/validation accuracies for both models.
2. (2%) Add dropout layers, and report training/validation accuracies with dropout rates equal to (A) 0.25/(B) 0.5/(C) 0.75 respectively.

Report Submission

- Submit with gradescope, no need to upload any files.
- We can only see your last submission.
- Submission deadline:
 - **2022/3/18 24:00 (UTC+8)**

Gradescope Registration

1.

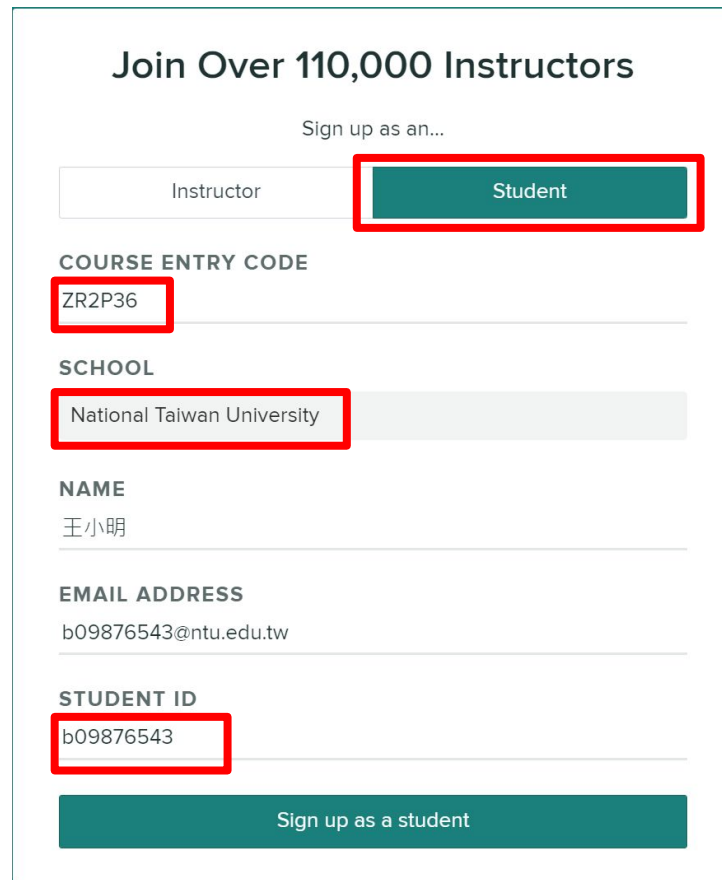


3.

Welcome to Gradescope! You are now enrolled as a student in 921 U2

To get started, you will first need to [set your password](#) (link will expire (CST)).

2. Entry Code: ZR2P36

A registration form titled 'Join Over 110,000 Instructors'. It includes a 'Sign up as an...' section with 'Instructor' and 'Student' buttons, where 'Student' is selected and highlighted with a red box. Below is the 'COURSE ENTRY CODE' field with 'ZR2P36' entered and highlighted with a red box. The 'SCHOOL' field contains 'National Taiwan University' and is also highlighted with a red box. The 'NAME' field contains '王小明'. The 'EMAIL ADDRESS' field contains 'b09876543@ntu.edu.tw'. The 'STUDENT ID' field contains 'b09876543' and is highlighted with a red box. At the bottom is a large teal button labeled 'Sign up as a student'.

Gradescope Submission

NAME	STATUS	RELEASED	DUE (CST) ▼
ML2022Spring HW2	● No Submission	<div><div></div></div> FEB 23	3 weeks, 2 days left MAR 19 AT 12:00AM

ML2022Spring HW2

Q1

2 Points

Implement 2 models with approximately the same number of parameters, (A) one narrower and deeper (e.g. hidden_layers=6, hidden_dim=1024) and (B) the other wider and shallower (e.g. hidden_layers=2, hidden_dim=1700). Briefly discuss the difference in training processes and training/validation accuracies.

Regulations

(*) [Academic Ethics Guidelines for Researchers by the Ministry of Science and Technology](#)

- You should NOT plagiarize, if you use any other resource, you should cite it in the reference. (*)
- You should NOT modify your prediction files manually.
- Do NOT share codes or prediction files with any living creatures.
- Do NOT use any approaches to submit your results more than 5 times a day.
- Do NOT use additional data or pre-trained models.
- Your **assignment will not be graded** and your **final grade x 0.9** if you violate any of the above rules.
- Prof. Lee & TAs preserve the rights to change the rules & grades.

If you have any questions, you can ask us via...

- NTU COOL (recommended)
 - <https://cool.ntu.edu.tw/courses/11666>
- Email
 - mlta-2022-spring@googlegroups.com
 - The title should begin with “[hwX]” (X is the homework number)
- TA hour
 - To be determined