



Homework1 Description

Outlier Finding

- A matrix $M \in R^{N \times d}$ is generated from a uniform distribution $U(0,1)$
- There is one special row is assigned as $0.5 + 0.1 * U(0.1)$
- Your goal is to learn a model to predict which row it is

[0.1, 0.2, 0.9, 0.6]

[0.3, 0.1, 0.4, 0.9]

[0.4, 0.5, 0.5, 0.5]

[0.4, 0.9, 0.1, 0.6]]

$Y = 2$

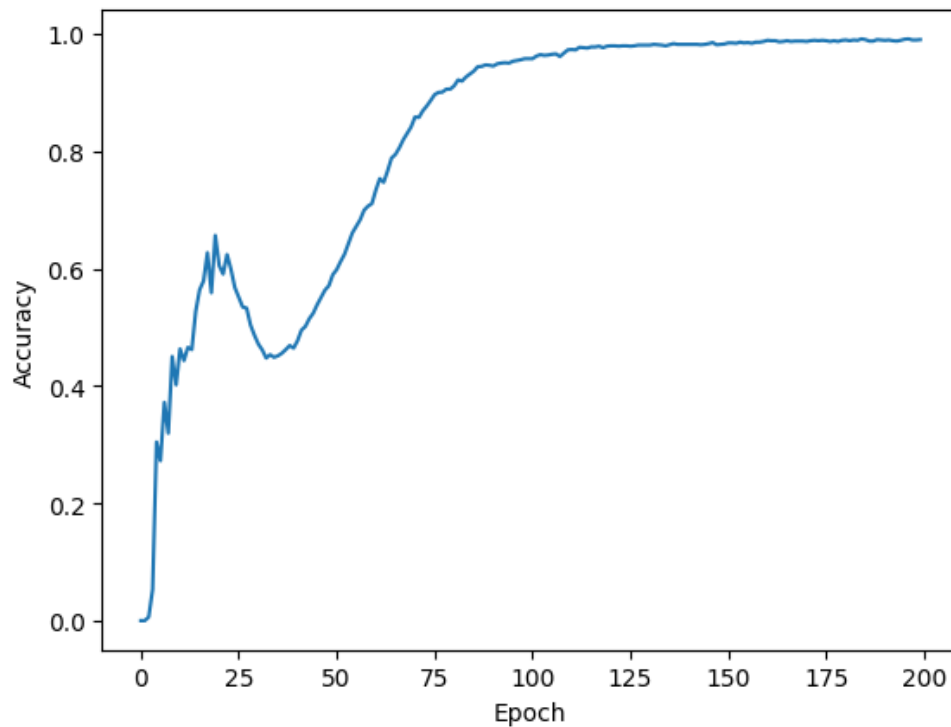
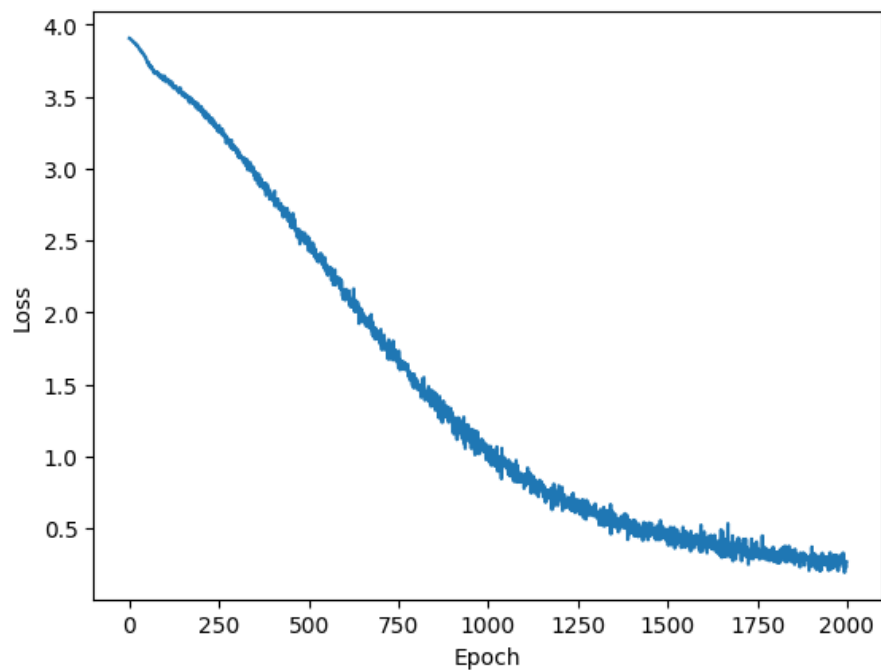
M



Outlier Finding




- Training: get training examples calling the `sample_batch` function
- Testing set: pre-generated 1600 data with `sample_batch` function
- Task: multi-class classification
- Evaluation metric: accuracy

Plot Your Learning Curve





























Practice: Multi-class classification

HOMEWORK

The information here is tentative and subject to change. Please read the requirement of each homework before deadline. There are different platforms for each homework, including Kaggle , NTU COOL , and JudgeBoi . Remember to submit to **EVERY** required platform.

- You should finish your homework on your own.
- 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 search or use additional data or pre-trained models.
- Your final grade will be subject to a 10% penalty if you violate any of the above rules.
- Prof. Lee & TAs preserve the rights to change the rules & grades.

#	HW	Slide	Code	Platforms	Video(En)	Video(Zh)	Date	TA
x	Colab Tutorial			N/A			N/A	許湛然
x	Pytorch Tutorial	 		N/A			N/A	張恆瑞、許湛然
HW1	Regression						3/05~3/26	張恆瑞、許湛然
HW2	Classification		 	 			3/12~4/02	孟妍、李威緒、陳宣叡、施貽仁
HW3	CNN						3/26~4/16	曾韋誠、黃健祐

<https://speech.ee.ntu.edu.tw/~hylee/ml/2021-spring.php>

<https://speech.ee.ntu.edu.tw/~hylee/ml/ml2021-course-data/hw/HW02/HW02.pdf>

Practice: Multi-class classification

Task Introduction

Task: Multiclass Classification

Frame-wise 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

<https://speech.ee.ntu.edu.tw/~hylee/ml/2021-spring.php>

<https://speech.ee.ntu.edu.tw/~hylee/ml/ml2021-course-data/hw/HW02/HW02.pdf>

Practice on Colab

Homework 2: Phoneme Classification

Objectives:

- Solve a classification problem with deep neural networks (DNNs).
- Understand recursive neural networks (RNNs).

If you have any questions, please contact the TAs via TA hours, NTU COOL, or email to mlta-2023-spring@googlegroups.com

Download Data

Download data from google drive, then unzip it.


You should have

- `libriphone/train_split.txt`: training metadata
- `libriphone/train_labels`: training labels
- `libriphone/test_split.txt`: testing metadata
- `libriphone/feat/train/*.pt`: training feature
- `libriphone/feat/test/*.pt`: testing feature

after running the following block.

Notes: if the google drive link is dead, you can download the data directly from [Kaggle](#) and upload it to the workspace.

Submit on Kaggle

 Community Prediction Competition

ML2021Spring-hw2

TIMIT framewise phoneme classification

1,522 teams · 2 years ago

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [Submissions](#) [Late Submission](#) [...](#)

Overview

Start

Mar 12, 2021

Close

Apr 2, 2021

Prizes & Awards

Kudos

Does not award Points or Medals

Participation

1,522 Competitors

1,522 Teams

18,499 Entries

Description

Task Description

Framewise phoneme classification of TIMIT dataset

- 1229932 training samples
- 451552 testing samples
- Evaluation metric: catagorization accuracy

Table of Contents

Description

Citation

Submit on Kaggle

Overview	Data	Code	Discussion	Leaderboard	Rules	Team	Submissions			Late Submission	...
12	▲ 6	b07705013_兵分2999路					0.76709	3	2y		
13	▲ 2	b06502011_					0.76705	43	2y		
14	▼ 5	b08902072_進擊敬能					0.76648	67	2y		
15	▼ 1	b07902010_cele					0.76646	19	2y		
16	▲ 5	b08902012_鎧之敬能					0.76631	48	2y		
17	▲ 5	b07508005_*					0.76617	13	2y		
18	▲ 6	b06703084_皓霆					0.76613	9	2y		
19	▼ 6	r09922115_台大陳傑憲					0.76604	45	2y		
20	▲ 3	r09921064_paka★地爆天星★					0.76597	63	2y		
21	▼ 2	B08902073_好想把code寫漂釀					0.76596	17	2y		
22	▲ 5	r09922081_StrongBaseline					0.76581	36	2y		
23	▲ 2	r09922112_ \ (●'v`●)/					0.76577	35	2y		
24	▲ 18	r09921096_誰給你一袋米呦					0.76511	11	2y		

<https://www.kaggle.com/c/ml2021spring-hw2>

Check you score on leaderboard