

# 2022 fall ADL HW1 Report

R11922A15 張仲喆

## Q1: Tokenizer

### 1. Describe detail.

在這次的作業，我所使用的tokenizer是內建的中文tokenizer「**bert-base-chinese**」，tokenizer會利用「，，。：：；；！！??」等符號切割句子，並利用token library找尋對應的字詞進行替換。

### 2. Answer Span.

我是直接使用huggingface中的對應方法，依照我對程式內容的理解，是利用迴圈去對應token、尋找答案開始的位置。而再inference時，預測出來的位置即為token，經過轉換便可以很容易的對應回中文字。

## Q2: Modeling with BERTs and their variants

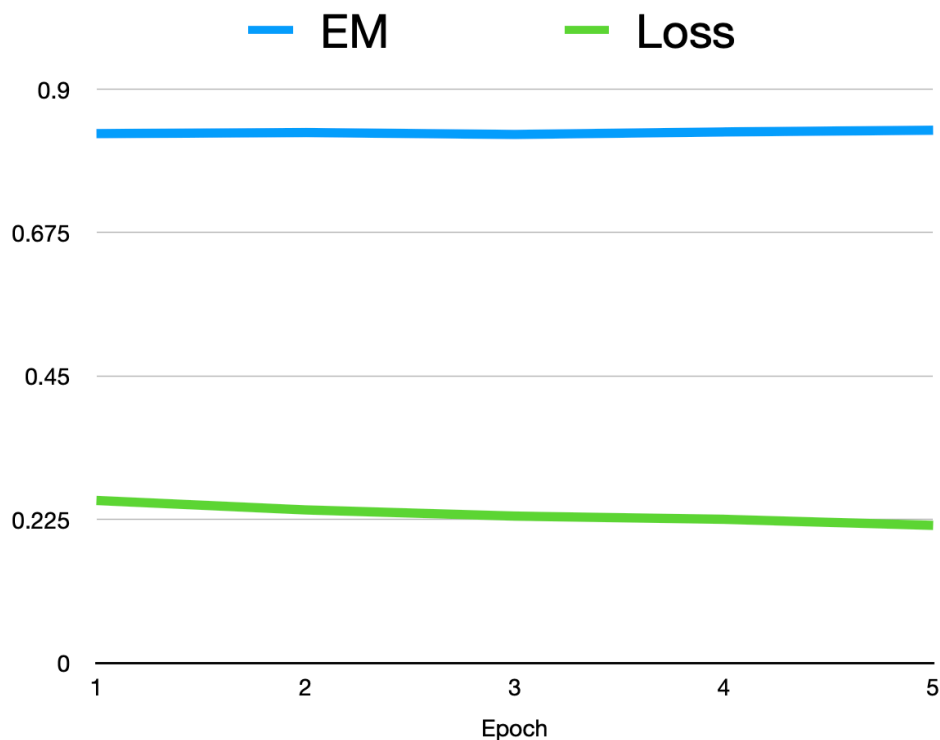
### 1. Model (MC)

- a. BEET-wwm-ext
- b. Acc.: 0.962
- c. Loss function: CrossEntropyLoss
- d. Optimization algorithm: AdamW
  - Learning rate: 2e-5
  - Batch size: 8 \* 8 (batch \* accumulation)
  - max\_length: 1024

### 2. Another Model

- a. Bert-wwm-large-ext
- b. Acc.: 0.968
- c. Loss function: CrossEntropyLoss
- d. Optimization algorithm: AdamW
  - Learning rate: 3e-5
  - Weight\_decay: 1e-6
  - Batch size: 2 \* 16 (batch \* accumulation)
  - max\_length: 512

### Q3: Curve



### Q4: Pretrained vs. Not Pretrained

在這個問題中，我測試了基本的BERT-Base model 在是否pretrain的狀態下的performance差異，兩者的訓練設定是相同的，只差異在是否使用 pertained weight。

- BEET-Base
- Loss function: CrossEntropyLoss
- Optimization algorithm: AdamW
- Learning rate:  $2e-5$
- Batch size:  $8 * 8$  (batch \* accumulation)
- max\_length: 1024
- Performance: (Acc.)
  - Non-pretrained: 0.51
  - Pretrained: 0.962