

## ML HW4 Report - CNN

### Basic Part

#### 1. Data Preprocess

在 basic part 中，我們需要 column Edema 的資料，因此將 df\_data(讀出的 csv 資料) 中的 column Edema 存入 Y\_train, Y\_val，用以做 basic prediction。

#### 2. Model Structure

1. 一開始進到第一層 layer 先使用 convolution layer(filter=32)，一開始因為資料較新，選擇的 kernel\_size 較大，strides 也較大，後續再逐漸縮小。這一層選擇的 kernel\_size 為 5x5，strides=3x3，使用的 activation function 為 relu。
2. 接著第一層的 Pooling，我們使用的是 MaxPooling(3x3)。
3. 為了避免連接太多，第一層最後做 Dropout(0.25)然後進入第二層。
4. 第二層用的方式大致上與第一層相同，不同之處在於 convolution layer 的 kernel\_size 改為 3x3，strides 也改為 1x1。後續同樣用 Maxpooling 與 Dropout(0.25)。
5. 進到第三層，與前兩層也大致相同，不同處是 kernel\_size=1x1、strides=1x1。
6. 在三層 convolution layer 後，先加入 Flatten()，準備做 Fully Connected layer。
7. 第一個 fully connected layer，output 數為 128，使用的 activation function=relu。完成後加入 Dropout(0.5)。
8. 最後一層同樣為 fully connected layer，output 數為 1，使用的 activation function=sigmoid。

#### 3. Train step

在建構完 model 後先進行 compile (Optimizer=Adam(lr=default=0.001))，就可以進入 train step。訓練用的資料為 X\_train 與 Y\_train，而 validation\_data 則是從中 split 出的 (split\_rate=0.3)，在 basic part 使用的 epochs=10。

#### 4. Validation step

使用 X\_val 進行預測，並與 Y\_val(Ground Truth)進行比較。且因為最後結果是以 f1\_score 測試，這裡特別使用 f1\_score(average=binary)進行 validation，得出的 **f1\_score=0.6607773851590106**

#### 5. Predict & Output step

最後就是用 X\_test 的資料進行預測，預測後將  $\geq 0.5$  的值 set as 1， $< 0.5$  的值 set as 0，然後輸出為 csv。

## Advanced Part

### 1. Data Preprocess

在 Advanced part 中，我們需要除了 Age 與 Gender 以外的資料，因此將 df\_data(讀出的 csv 資料)中的所需 column 存入 Y\_train\_ad, Y\_val\_ad，用以做 prediction prediction。

### 2. Model Structure

這部分與 basic part 大致都相同，唯一的不一樣是最後一層，因為要預測的數量從 1->7，因此 output 數也改設為 7（即 Dense(7)）。

### 3. Train step

在建構完 model 後先進行 compile，就可以進入 train step。訓練用的資料為 X\_train 與 Y\_train\_ad，而 validation\_data 則是從中 split 出的（split\_rate=0.3），在 Advanced part 使用的 epochs=20。

### 4. Validation step

使用 X\_val\_ad 進行預測，並與 Y\_val\_ad(Ground Truth)進行比較。且因為最後結果是以 weighted f1\_score 測試，這裡特別使用 f1\_score(average=weighted)進行 validation，得出的 f1\_score=0.5087549687590339

### 5. Predict & Output step

最後就是用 X\_test 的資料進行預測，預測後將  $\geq 0.5$  的值 set as 1， $< 0.5$  的值 set as 0，然後輸出為 csv。