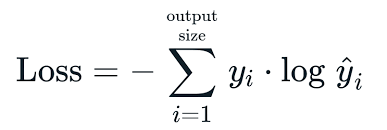
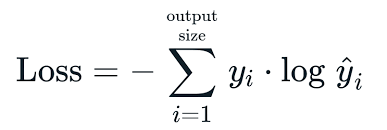
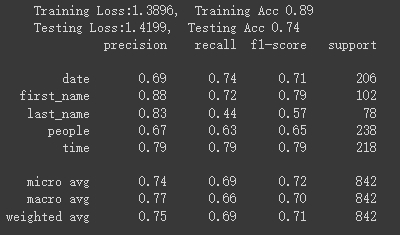
Applied Deep Learning HW1 report

B08310054 歐崇愷

1. Data Processing
   * I use NLTK to tokenize 'text' field and didn’t remove stop word and didn’t set the max input length of text
   * I use GLOVE as a fixed, pretrained word embedding. Embed each word into 300 dim vector
2. Describe your intent classification model
   * I use LSTM model
     + *ht*, *ct* = LSTM(*wt*, *ht-1*, *ct-1*), where *wt* is the word embedding of the t-th token.
     + *ht* is a 2-dim vector, representing the hidden state of LSTM at the t-th timestamp
     + choose softmax as activation function
   * performance of your model
     + **0.92222 acc**
   * **cross entropy loss **
   * I use torch.optim.Adam as the optimizer, with
     + Learning rate = 0.003
     + Batch size = 128
3. Describe your slot tagging model
   * I use LSTM model
     + *ht*, *ct* = LSTM(*wt*, *ht-1*, *ct-1*), where *wt* is the word embedding of the t-th token.
     + *ht* is a 2-dim vector, representing the hidden state of LSTM at the t-th timestamp
     + choose softmax as activation function
   * performance of your model
     + **0.76300 acc**
   * **cross entropy loss **
   * I use torch.optim.Adam as the optimizer, with
     + Learning rate = 0.0003
     + Batch size = 1
4. Sequence Tagging Evaluation
   * 在自己計算的joint acc 中得到0.74的命中率，然而在token accurancy中僅得到約0.71的命中率，由此可見在較長句子中token的預測率較為不理想，才會使token accuracy小於joint accuracy.
5. Compare with different configurations
   * In intent classification case，我初始化參數為learning rate 0.05, hidden dim 512, numbers of layer 2，發現訓練不起來，卡在acc = 0.6，嘗試降低hidden dim至256表現略微提升，調整numbers of layer 至1或3反而更差，最後調整learning rate 至0.005後acc上漲至0.92
   * In slot tagging case，我初始化參數為learning rate 0.005, hidden dim 512, numbers of layer 2，發現訓練不起來，acc = 0.01，嘗試降低hidden dim至64表現略微提升，最後調整learning rate 至0.0003後acc上漲至0.73