Assignment 1: Named Entity Recognition   2020-2021 Spring Semester

Submission Due: 11:59pm on February 13 (Sat)

Submission: Send iPython notebook (or Python codes) and report (only Question 1.c results) in PDF to ai6127.assignments@gmail.com with title "AS1-[YourName]"

1. Question One **[50 marks]**
    a. Download a named entity recognition dataset from https://github.com/leondz/emerging_entities_17, modify the format of the dataset as input to the hands-on implementation of "Named entity recognition by using CRF" of Lecture 3 Slide 38 and run the hands-on implementation, reporting F-score **[15 marks]**
        i. Training data: wnut17train.conll (Twitter)
        ii. Development data: emerging.dev.conll (YouTube)
        iii. Test data: emerging.test.conll (YouTube)
    b. Modify the format of the dataset as input to the Softmax classifier of Tutorial 3 and run the Softmax classifier, comparing with the CRF model's performance in terms of F-score **[15 marks]**
    c. Optimize the hyper-parameters of the Softmax classifier in terms of F-score, by alternating at least two values of each of the following hyper-parameters **[20 marks]**:
        i. Window size
        ii. Embedding size
        iii. Hidden layer size
        iv. Number of hidden layers
        v. Freeze word embeddings or not
        vi. Learning rate
        vii. Number of epochs

    You may select small numbers of epochs for the optimization experiments if each experiment takes long time. Display the experiment results in a tabular format which compares alternative values of the hyper-parameters.