**CS288 Natural Language Processing:**

# Homework3 Report

Name: Diyun Lu SID: 3039752224

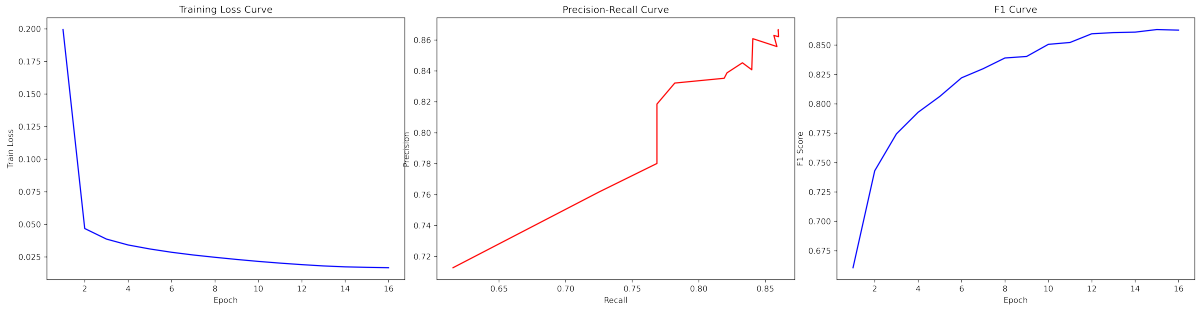
## 1 Part a: Tagging Model

At first, a calculation problem was in my code and I could not reach 95 accuracy anyway. So I experimented a lot by adjusting batch size, training more epochs, and adjusting each part's dropout rate. An overfit was observed at about epoch 20; even the big batch size cannot help. The slight difference in dropout rate almost has no effect on helping with the overfit while big batch size and more training epochs can increase the upper limit of the accuracy.

## 2 Part b: Parsing Model

### 2.1. Statistical Analysis

From the statistics of EVALB, we can see that the model generally reached the expected F1 score. From detailed graphs of the training loss, PR curve, and the F1 curve through the training epochs, we can tell that the training was going normally and did not appear to overfit. However, the PR curve showed that the precision oscillated back and forth greatly while the recall was increasing steadily. We may conclude that the model is having trouble identifying the false positives.



### 2.2. Comparison of Sentence Types

For the trained model, I examined the performance on NP, VP and PP separately. We can observe a high false positive on NP which is aligned with the conclusion of the last part. Semantically, NP actually includes many VP and PP, so it may be the reason that the model falsely identified them as NP which results in a high predicted positive NP.

| Types | NP | VP | PP |
|---|---|---|---|
| Predicted Positive # | 12889 | 4890 | 3888 |
| All Positive # | 12719 | 4913 | 3893 |
| F1 % | 87.42 | 87.41 | 87.40 |

**Table 1**: Comparison Table