

## CS288 Natural Language Processing:

# Homework1 Report

Name: Diyun Lu      SID: 3039752224

## 1 Experiments and Results

I actually had the best perplexity before implementing any other configurations to the LSTM model. At the very beginning, the perplexity can reach around 61. It appears a overfit with a rise of the perplexity after 10 epochs.

Since the dataset is relatively small, it makes sense and regularization may help. I tried several ways separately including adding a l2 regularization penalty, schedule learning rate, and I also tried another optimizer.

### 1.1. L2 Regularization

With the decay coefficient  $10^{-4}$  and other settings unchanged, the final perplexity stayed at around 64.

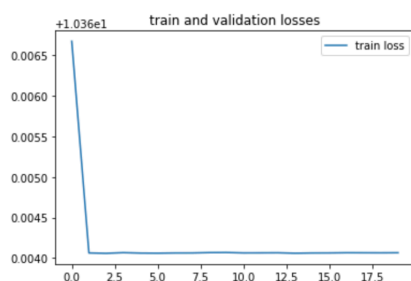
### 1.2. Optimization

I also tried to schedule the learning rate. Since the original model's perplexity stopped to decrease at about epoch 10, I set the learning rate to decay every 5 epochs at a rate of 0.5. Hoping the slow decrease of learning rate can affect the result. But it the perplexity stayed the same at 64.

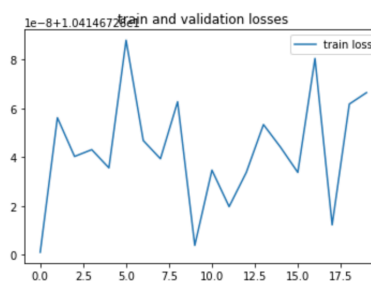
According to the thesis, SGD with momentum outperforms than other optimizers in the context of word-level language modeling in many cases. Thus I replaced the optimizer with SGD with momentum of 0.9. But it seemed not working well and I did not reserve that part of codes.

## 2 Analysis

From the loss of training we can see that it is actually low enough. So it appears that the perplexity so good for the original model should be a overfit. Adding regularization ways somehow helps correct it but my strength for regularization may not be enough.



(a) L2-train loss



(b) Scheduler-train loss