Part 1 – Report

Name

Simon Raviv

Hyper-parameters

- epochs: 25
- embedding-dim: 30
- mlp-hidden-dim: 16
- lstm-hidden-dim: 32
- batch-size: 16
- lr: 0.003
- weight-decay: 0
- dropout: 0

I have implemented batch training using PYtorch LSTMCell.

Dataset Size

I have experimented with 2 dataset sizes:

- Total of 1000 samples with train/dev ratio of 900/100
- Total of 500 samples with train/dev ratio of 400/100

In addition, with several types of length for the subsequences:

- Random uniform length between 1 to 10 for each sub sequence
- Random uniform length between 1 to 20 for each sub sequence

There was difference between the datasets and the type of data. Using shorter sequences, it was easier to train.

In addition, using smaller/bigger dataset sets, with shorter sub sequence, it looked like it was easy to get high accuracy on both train and dev, but this is due to overfitting, caused by the short sub sequence length.

I chose the bigger dataset with bigger sub sequence, since it will generalize better for other sizes.

Network Accuracy

The network reaches 98-99 [%] on both train and dev sets. It is also possible to get to 100 %, which I saw in different experiments.

Once the network was implemented correctly, it was quite easy to get high accuracy without too much tuning or needing any kind of regularization techniques or learning rate scheduling.

Fitting Duration

- Time: On a 4 physical cores (8 hyperthreaded) 2.3 GHz CPU, the overall fit duration is ~ 4.5 minutes, where each epoch takes about 14 seconds.
- Epochs: It took 25 epochs in the setting I chose.

On a GPU machine, it's even much faster of course.