Assignment 3 - Report 1

324827328, 213713522

May 2025

Hyper-Parameters:

- Char Embeddings Dim = 10
- LSTM Hidden State Dim = 50
- MLP Hidden Layer Dim = 50
- Epochs = 5
- Learning Rate = 0.01
- Amount Of Sequences = 6400
- Train/Dev Percentage = 0.8/0.2 (5120/1280 Samples)
- Batch Size = 64
- Accuracy Threshold = 0.5
- Each Subsequence max length = 10
- Max Sequence Length = 90 (Each Subsequence max length $\times$ 9 Sequences)

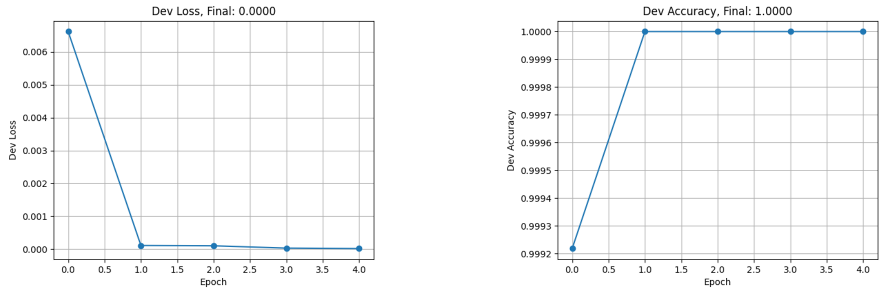

Figure 1: Dev Set Loss And Accuracy

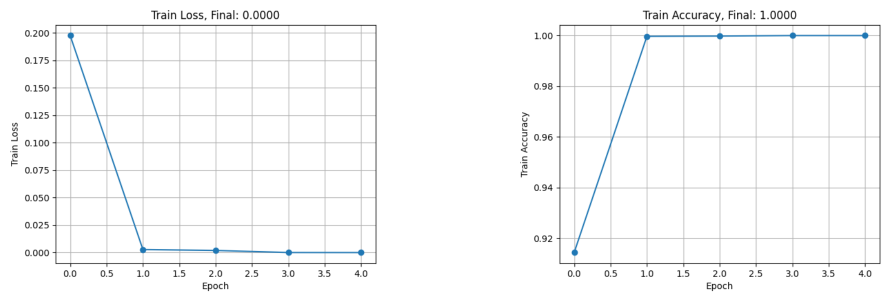

Figure 2: Train Set Loss And Accuracy

Summery

As we can see from figures 1 and 2, the experiment was a complete success. The model learned the data and got a perfect accuracy score of 1.0 both on the train and dev set after 5 epochs (Wall Clock time 15.0688 seconds). It was obvious this architecture will work on the generated languages, because the LSTM cell has a memory state passed on, making it possible for the model to learn and recognize when it sees ”b” before ”c”.

Padding

To make the LSTM cell work properly, we needed each sequence to be the same length, even though they are generated in different lengths. We solved this by padding the end of each short sequence with a <SHORT>token. Then, in order for the LSTM to not be affected by this token, we optimized it so it will not update its memory if its got to the padding token. This is important because we don’t want this token to contribute noise to the LSTM learned parameters.