Deep Learning - Lab 7 Exercise

Han Zhang hz5g21@soton.ac.uk

Exercise 1

Exercise 1.1:Complete and train a sequence-to-sequence model.
The small snippet of code is listed as follows.

```
def forward(self, src):
    # TODO
    embedded = self.embedding(src)
    output,(hidden, cell) =
        self.rnn(embedded)
    return hidden, cell
```

The loss curves of training and validation are shown in Figures 1 and 2.

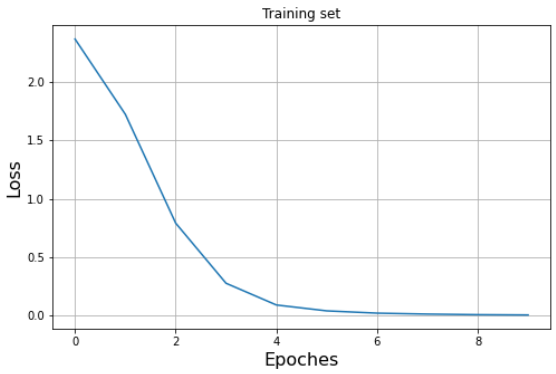

Figure 1: Loss function of training

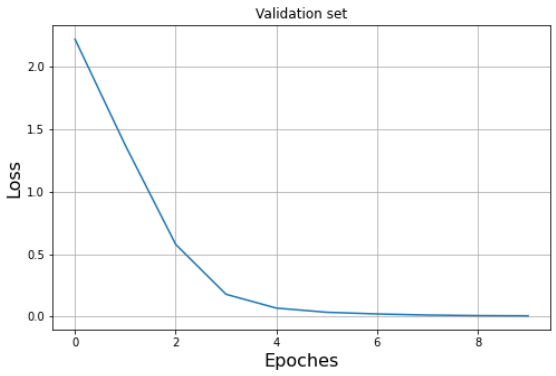

Figure 2: Loss function of validation

Exercise 1.2:now use it!

1. **why is the order of the output reversed?**

   Reversing the order of words in all source sentences will improve the LSTM's performance markedly since it introduced many short-term dependencies between the source and the target sentence which made the optimization problem easier. This causes the output reversed.

2. **what is the point of teacher forcing?**

   Teacher forcing is a strategy for training recurrent neural networks that use ground truth as input for t+1, instead of model output from a prior time step t as an input. This is done to speed up convergence and improve model skills.

Exercise 1.3:Sequence Lengths.
Predicting a longer chunk is problematic in this model. Modify the function to work with 2 sets in one chunk. The results demonstrate that sentences have missing letters at the beginning and end. This indicates that a longer chunk has some problems to work with at the beginning and end. It is because of the maximum sequence length of only 6. A model with a shorter training sequence performs poorly on longer chunks.

```
.- -. ... .-- . .-. / - .... . / ..-. ---
    .-.. .-.. --- .-- .. -. --.
original = answer the following
longer chunk = nswer the followin
```

This phenomenon may result from the mechanism in LSTM. Based on the training set, an LSTM model is trained to have a fixed term of memory. If it is subsequently given a piece of code that is far longer than the sequence in the training dataset, it will invariably fail to preserve that lengthy memory.