



1.1 Complete and train a sequence-to-sequence model

The plots show the loss curves for first few epochs(0 to 1 and 1 to 2 epoch) and the last epoch(1 to 8 epoch). I used the following forward method to complete the code:

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

1.2 now use it!

Decoded Morse Code translation (output):

“answer the following
why is the order of the output reversed
what is the point of teacher forcing”

Output observation:

In the paper proposed by Sutskever et al, they concluded that the extent of the improvement obtained by reversing the words in the source sentences was remarkable and that it is important to find a problem encoding that has the greatest number of short term dependencies, as they make the learning problem much simpler.

The reversed input improved the LSTM’s ability to model long-range dependencies, which in turn increased the overall translation quality.

In seq2seq models, especially those using LSTM (Long Short-Term Memory), reversing the input can help alleviate issues related to vanishing gradient which prevents weights from changing their values. This is significant in networks with long sequences where early inputs have a minimal impact because their gradients vanish by the time back-propagation reaches them.

Teachforcing:

The model used is seq2seq and a training method named “teacherforcing” is used. The seq2seq is a neural network architecture and can handle variable length inputs and outputs. It has an encoder and a decoder used to process the input by converting it to vector and finally outputting it as a readable sequence.

In teacherforcing method the decoder is actually fed the target output as input for the next timestep instead of its own prediction. This allows the model to use the correct previous output to to predict the current output.

This allows the model to learn faster as it learns using accurate feedbacks. Using its own predictions can cause noise especially during the early stages where the model is not sufficiently trained. Excessive teacher forcing can induce exposure bias, and can cause the model to become overly reliant on the teacher-forced inputs during training. This may result in poorer generalization to unseen data during inference. The model will lose its ability to recover from its own incorrect predictions. So the teacher forcing ratio is set to 0.5.

1.3 How well does the model work with longer chunks?

The model worked the same for longer chunks. When I entered longer morse code, it decoded them accurately.