Deep Learning: Lab 7

Johnny Joyce (jvjj1u19@soton.ac.uk)

1 Sequence-to-Sequence Modelling

1.1 Complete and train a sequence-to-sequence model

The snippet of code used for `forward()` was as follows:

```
def forward(self, src):
        embedded = self.embedding(src)
        lstm_out, state = self.rnn(embedded)
        return state
```

When this was used to train the model, the loss curves were as shown in Figure 1.

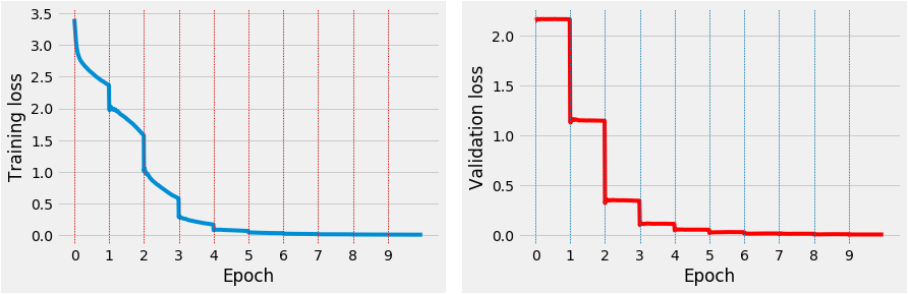

Figure 1: Plots of training loss and (left) validation loss (right).

1.2 Now use it!

- *Why is the order of the output reversed?*

By reversing the order of the output, some of the characters in the input will be closer in the string to the characters to which they correspond in the output. This makes it easier for the model to associate these characters with one another. Also, the *average* distance between corresponding characters does not change, meaning further away corresponding characters do not create an overall disadvantage. [1, Section 3.3]

- *What is the point of teacher forcing?*

Without teacher forcing, the network would have to use backpropagation on the previous hidden state, which would then depend on the hidden state before that, and so on — this would become intractible after a few iterations. With teacher forcing, the previous output is replaced by the previous output (which has no gradient), so the network does not need to backpropagate through each time step.

1.3 Sequence Lengths

The given morse code uses spaces to separate characters since each morse code character consists of multiple dots and dashes. By splitting the input when there is a space, each separate input given to the model will represent a single character. If longer chunks are used (e.g. morse code characters per chunk), the model will only output a single character for each chunk.

References

[1] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." In *Advances in neural information processing systems*, pp. 3104-3112. 2014.