Lab 7 Exercise – Transforming Sequences

Justin Ugwudike (jknu1g19@soton.ac.uk)

April 5, 2023

**1.1 Question 1.1**

The complete code and the loss while training the model is shown in Figure 1 and Figure 2.

**Text

Description automatically generated**

Figure 1 – Completed encoder forward function.

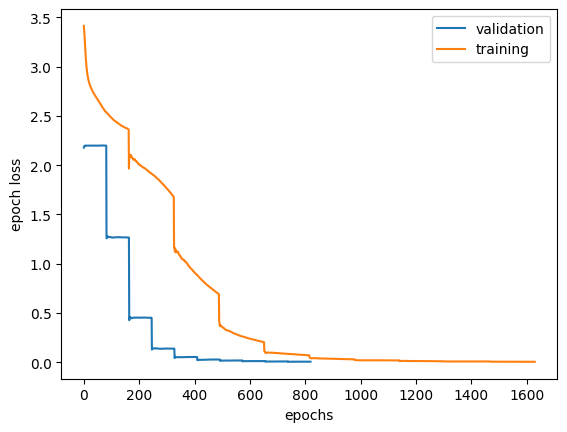


Figure 2 – Epoch loss curves on the training and validation data.

**1.2 Question 1.2**

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

Because in the original transformer paper they reversed the order input to help train the model to remember short term dependencies. Therefore, the output needs to be reversed as well to match the original input sequence.

**What is the point of teacher forcing?**

Teacher forcing allows the model to learn correct sequences of the output more efficient and in turn, provides faster convergence when training. Teacher forcing is done when the ground truth of a problem is inputted to the model at each time step.

**1.3 Question 1.3**

The model breaks when larger chunks are used. Splitting the text at every ‘/’ character or diving all the morse character codes into groups of three produces invalid outputs.

**Graphical user interface, text

Description automatically generated**

Figure 3 – Decode method output when the input is chunked into 3 words per chunk.

This is expected because during the training of the model the training data is enumerated through where a prediction is made, and the gradient of the loss is calculated for each datapoint and label pair.