

Solutions

Ans4: A typical problem for GRU is when we are analysing long sentences, the information in (t) cannot carry necessary information to (t+N) if N is too large.

Ans5:

We use LSTM model with softmax activation function.

It's difficult to say definitively that one approach is "better" than the other, as the choice of model will depend on the specific task and data at hand. In some cases, LSTMs may be more effective, while in others, BERT or other transformer-based models may be more suitable. It's important to experiment with different models and evaluate their performance on the task to determine which one is the best fit.

LSTMs are well-suited to these tasks because they are able to capture long-term dependencies in the input data. This means that they are able to take into account the context of a word in a sentence or document, which is important for accurately understanding the meaning of the text.

BERT can be a good candidate for this scenario in which it can be better to process long sentences though we may suffer from a high time complexity but still might be better than GRU. The common activation functions for BERT are ReLu or Tanh to minimize the impact of vanishing gradient.

For this situation LSTM works well.

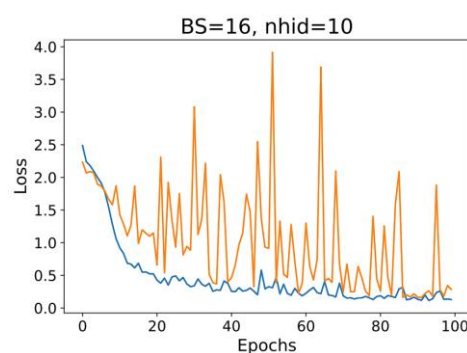
Assignment 3

Ans1: In adversarial machine translation, it is common to group words pairs based on the length of the source and target sequences. This is done to ensure that all of the sequences within a mini batch have the same length, which allows the model to process them efficiently.

By grouping words pairs in this way, you can avoid the issue of dealing with variable-length sequences, which can be challenging to process. This can be especially useful when working with neural machine translation models, which typically require fixed-length input and output sequences.

It's worth noting that there are other ways to handle variable-length sequences in machine translation, such as using padding or bucketing techniques. However, grouping words pairs based on length is often a simple and effective approach that can work well in many cases.

Ans2:



```
Test loss: 0.44588130712509155
Test accuracy: 0.8655410408973694
```