Assignment 8

Advanced Applied Machine Learning

1. [100 pts]

   In the module, the attention mechanism implementation is based on a single sequence input and iterated over every token individually. This implementation matches the block diagram in the module.

   Convert the implementation to batched using BATCH_SIZE to take advantage of PyTorch and a GPGPU device. Code updates and hints towards batch-by-batch processing have been provided throughout the implementation, including the dataset class NMTDataset. General steps towards batching in an RNN attention neural network,

   - Each sequence is created as a fixed-length tensor and padded to fill the tokens to the length after the `EOS` token.
   - Training has to input the sequences batch by batch
   - It is permissible to go over the tokens one by one, but in batches
   - RNN layers have `batch_first` to control the batch order, either batch-sequence-features or sequence-batch-features for True and False, respectively. The initial hidden layer order is always the same, sequence-batch-features

   Show that batched implementation is faster and generates the same error convergence.