Report

Part 1 - A Basic Encoder-Decoder Model

The model's parameters settings:

Shared:

- NLL Loss (Cross Entropy Loss).

Encoder:

- The encoder is composed of one unidirectional LSTM layer with hidden dim of 128.
- The embedding dim is 128.
- Adam-based optimization.
- Learning rate of 0.0007.

Decoder:

- The decoder is composed of one unidirectional LSTM layer with hidden dim of 256.
- The embedding dim is 128.
- Dropout layer is applied on the output of the LSTM layer, with dropout rate set to 0.3
- Adam-based optimization.
- Learning rate of 0.0007.

Training time (including evaluation at the end of each epoch) : **3.607 minutes**

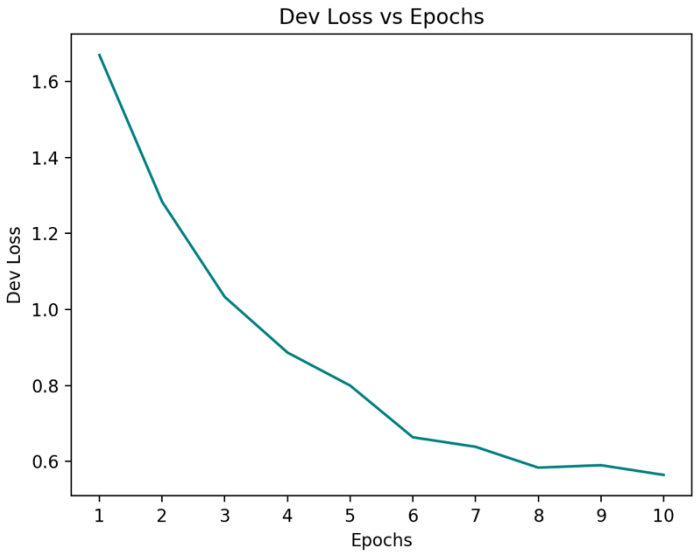
Note:

The loss on both data sets, training set and dev set, was calculated with NLL loss by summing the negative log likelihoods for each token in the output sequence and then dividing this sum by the number of predicted tokens (->target_length-1)

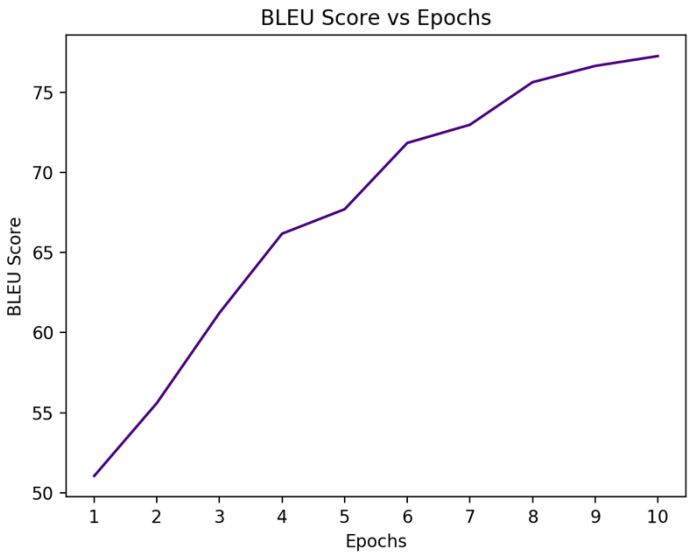
Loss on the training set : 0.164

Learning Curves - Dev set:

Loss - 0.564



BLEU score - 77.261



Result on the Test set:

BLEU score - 81.949