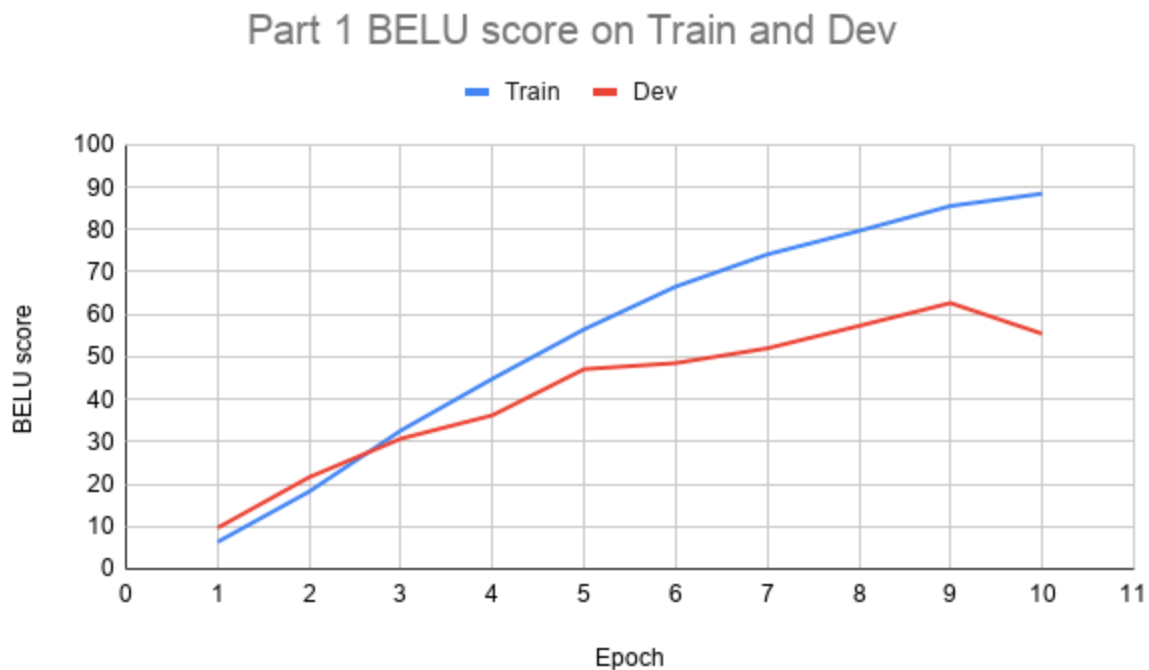Ofer Sabo
201511110

Part 1

I used the Optuna package to find the hyperparameters which produce the best BELU score. As a result, these values are not "round".

Hyperparameters:
"decoder_type": "gru",
"drop_out_rate": 0.19828927415429567,
"hidden_size": 150,
"embedding_dim": 850,
Learning rate: 0.0007125761361357518.
Optimizer: adam.

Description of the model architecture:
Encoder - single direction and single-layer LSTM, with size hidden.
Decoder - single direction and single-layer GRU, with size hidden.
Source embedding: size "embedding_dim".
Target embedding: size "embedding_dim".
Max decoding steps is 100.
Also I clip the gradient into a fixed value of 1.

Ofer Sabo
201511110

Here is the loss value graph:
- Orange is the training loss curve.
- Blue is the Dev loss curve.