MT Übung 4

Thema: RNNs

Varvara Stein

I have chosen the novels “Crime and Punishment” and “Brother Karamazov” by Fyodor Dostoevsky.
This data is relatively small, only 3,2 MB, but still enough for training, however, the performance could be poorer due to this reason.
For the first training I only used the novel “Crime and Punishment” (1,2 MB) and the default setting in romanesco. Once the model was trained for the first time, the perplexity scored 203.73 which is unfortunately rather a low result.
In order to improve the performance, I enlarged the size of the text data from 1,2MB to 3,2MB, i.e. added another novel. Assuming that the cleaner the text the better the performance is, I pre-processed the text data using the script “preprocessing.py”. The Pre-processing includes tokenization and lowercasing.
Moreover, some hyperparameters have been changed.

NUM_EPOCHS = 11

HIDDEN_SIZE = 1200

Other hyperparameters were set by default.

After the second training the perplexity was much lower than I expected and reached 63.40. Since no big changes have been made, the result is either not trustworthy enough, or greatly influenced by pre-processing.