Deep Learning- RNN

**Describe in the readme file how to train each setting, and how to test it**

**with the saved weights.**

We implemented the "small" model as described in "Recurrent Neural Network Regularization", by Zaremba et al by Google COLAB. The txt files were added to the model by URL addresses of their location in Github website. In order to run the model for training and evaluation, press Ctrl+ F9. We submitted our training and evaluation results on the Moodle website in PDF format.

For each one of the models we wrote its own test parameters and we ran each one of them separately.

* We implemented an architecture for each one of the following settings:

- LSTM-based network without dropout.

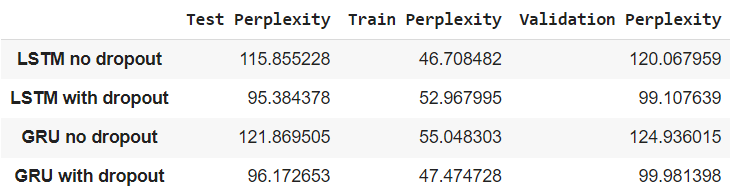
- LSTM-based network with dropout.

- GRU-based network without dropout.

- GRU-based network with dropout.

* In the next step, we adjusted the test parameters in order to make sure that without dropout the validation perplexity is below 125, and with it is below 100.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RNN** | **Dropout** | **Epochs** | **lr** | **winit** | **sequence length** | **Factor**  **epoch** | **lr factor decrease rate** | **Max gradients norm** |
| **LSTM-based network without dropout.** | 0 | 15 | 0.95 | 0.05 | 35 | 5 | 2 | 5 |
| **LSTM-based network with dropout.** | 0.5 | 80 | 0.22 | 0.05 | 35 | 60 | 1.2 | 5 |
| **GRU-based network without dropout.** | 0 | 20 | 0.38 | 0.05 | 35 | 5 | 2 | 5 |
| **GRU-based network with dropout.** | 0.5 | 100 | 0.29995 | 0.05 | 37 | 65 | 1.21 | 6 |

* In the final step, we presented the perplexity curves of each model by a plot and summary via table:

According to the table, both of the RNN networks which are based on dropout demand more epochs in order to converge under the required perplexity value. We believe that we needed more epochs for convergence due to the disregarding of nodes in the layers which caused a loss of information during the process. On the other hand, we noticed that adding too many epochs to the model led to higher validation perplexity values.

After training and testing the models, we found that the LSTM models had slightly higher validation perplexity values than the GRU models. We are aware that the difference between the validation value in each one of the types of RNN is also due to the different values of parameters, however, the efficiency of the LSTM performance was better than the efficiency of the GRU in most of the time we ran the training and the testing.