Assignment 4

Michele Cerú
mc3784@nyu.edu

1 nngraph

1. the code for the first question (part a and b) is in (1).
2. See below the graph for the GRU cell.

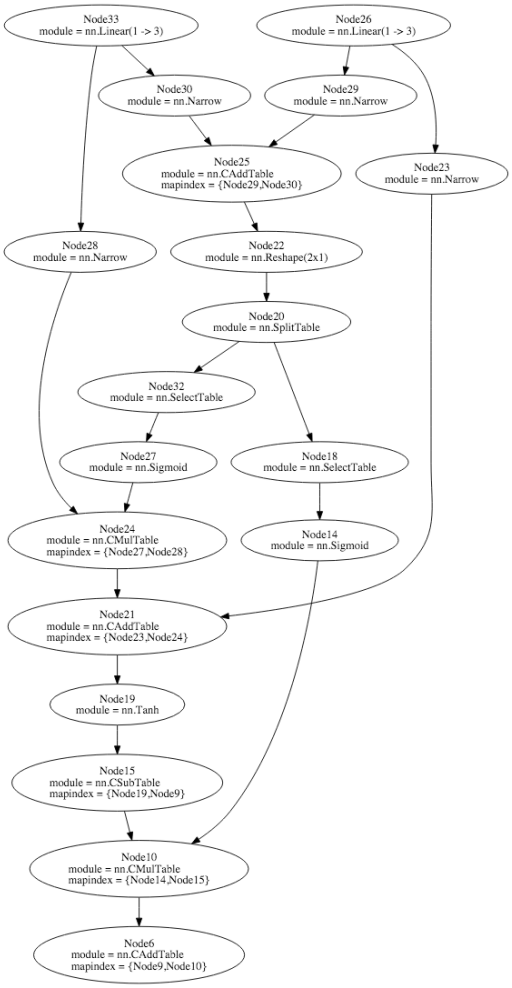

Figure 1: GRU cell graph

2 Language modelling

2.1 Generating sequences

The file query_sentences.lua does the following thing:

1. Load the network of the model
2. Build the vocabulary map and the inverse vocabulary map.
3. Get the number of words to predict and the number of words in the seed.
4. generate the index of each words doing a forward pass on the network, multinomial distribution over the probabilities generated by the logsoftmax layer.
5. Returns the new sentence.

The code is in (2)

2.2 Suggested improvements to your model

To improve the model I have tried a number of different thing:

- Change the number of layers from 2 to 4.
- Change the size of the network from 100 to 700
- Change the dropout parameter from 0 to 0.5
- Change the vocabulary size from 10000 to 12000
- Change the LSTM cell with the GRU cell
- Change gradient clipping

For the LSTM cell, the table below reports the best Perplexity and the parameters used. The table is divided by a line. In the experiments below the line I use dropout in the experiments above the line I didn't use it. When tuning the parameter I have noticed that an improvement in convergence speed and in the value of the perplexity occurred when augmenting the dropout parameter and simultaneously augmenting the size of the network. In particular the best model (3) has rnn_size= 600 and drop_out= 0.4, and it converges in about 8 epochs to a value of Perplexity 87.3.

seq length	layers	rnn size	dropout	vocab size	best Perplexity
20	2	200	0	10000	119.756
30	2	200	0	10000	114.548
15	2	200	0	10000	195.712
30	4	200	0	10000	120.359
40	3	200	0	15000	137.629
40	5	200	0.2	10000	135.020
40	4	400	0.2	10000	107.970
30	2	400	0.2	10000	93.449
30	4	400	0.3	10000	102.013
30	4	400	0.5	10000	113.420
30	2	400	0.5	10000	96.340
30	2	600	0.4	10000	87.368
30	2	500	0.3	10000	89.794

The table below reports the value of the Perplexity and the parameters for models that use the GRU cell. As for the LSTM models I noticed an improvement using dropout along with an increasing size of the network. And I found that the best model for the GRU cell has the same parameter founded for the LSTM: rnn_size= 600 and drop_out= 0.4. This model converges in about 11 epochs to a value of Perplexity 97.0.

seq length	layers	rnn size	dropout	vocab size	best Perplexity
20	2	200	0	10000	182.217
15	2	200	0	10000	195.712
30	2	600	0.4	10000	97.056
30	2	700	0.5	10000	101.021

In the plot below I compered the best model for the LSTM and the best model for GRU plotting the Perplexity vs the Epoch.

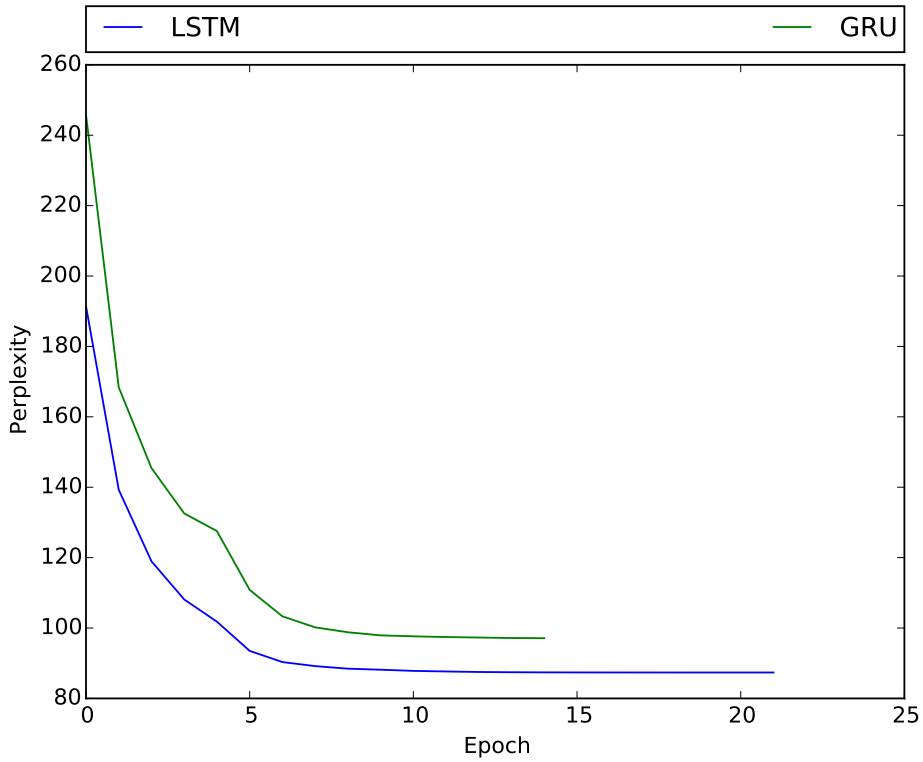

Figure 2: Comparison of the two best performing model for the LSTM cell (in blue) and for the GRU cell (in green)

References

[1] `https://github.com/mc3784/DL-A4/blob/master/submission/nngraph_warmup.lua`

[2] `https://github.com/mc3784/DL-A4/blob/master/submission/query_sentences.lua`

[3] `http://cs.nyu.edu/~mc3784/model.net87.367553621956`

[4] `https://github.com/mc3784/DL-A4/blob/master/submission/result.lua`