

Deep_Dand_Mid_Term_Deep_Learning

Part 1: Vary the number of hidden units exponentially as follows, 32, 64, 128, 256, and 512; compare results of each and answer the following questions:

What happens to the perplexity? Why do you think that is?

As we increase the number of hidden units, the perplexity has reduced from 1.830 to 0.88 with respect to number of units from 32 to 512. From the sentence output, it makes more sense as the number of units are increased. While the number of hidden units are more, the network has more activation function and so is able to help the output layer produce more accurate results.

Number of Units	Perplexity	Sentence
32	1.830	46374/46375 (epoch 124), train_loss = 1.830, time/batch = 0.026 >> sample mode: The right, I do, mear. Plitithune meated, we but flan
64	1.619	46374/46375 (epoch 124), train_loss = 1.619, time/batch = 0.062 >> sample mode: The kingdous alious. BIONDELLO: How prince fit id del
128	1.421	The art as need thy son: dispering us Enavise thee of
256	1.209	46374/46375 (epoch 124), train_loss = 1.209, time/batch = 0.354 >> sample mode: The to forehio.

		BIONDELLO: O sir, marry it so, it is
512	0.886	46374/46375 (epoch 124), train_loss = 0.886, time/batch = 0.565 >> sample mode: The slave's take my leave. DUKE OF YORK: My lord, I'll

What happens to the sentences that it produces? Why do you think that is? How does this relate to the previous question about perplexity?

The sentences make better sense at the end of 512 units. Since the perplexity is reduced the results get better.

(feel free to include any tables or plots to support your answers)

Part 2: Vary the length of the sequences as follows, 25, 50, and 75; compare results of each and answer the following questions:

(feel free to include any tables or plots to support your answers)

Sequence Length	Perplexity	Sentence
25	1.329	92874/92875 (epoch 124), train_loss = 1.329, time/batch = 0.026 >> sample mode: The seat he will not; Toon. LERDENI: Why, an is world
50	1.421	The art as need thy son:

		dispering us Enavise thee of
75	1.378	30874/30875 (epoch 124), train_loss = 1.378, time/batch = 0.131 >> sample mode: The thy should I may having'st out shall peace. First

What happens to the perplexity? Why do you think that is?

The perplexity doesn't change too much significantly. It does show the changes because as the length of sequence changes(increase or decrease), the number of operations performed by every unit is more/less respectively. With the experiments above it doesn't make much difference with the length and it is difficult to related if increased length is increasing perplexity because with the length of 75, perplexity is 0.05 lesser than the length of 50.

What happens to the quality of the sentences that it produces? Why do you think that is?
How does this relate to the previous question about perplexity?

With length of 75 it makes very little sense but with length of 50 and 25, the sentences does make some sense. Even the perplexity is better with length 25, the sentence formed doesn't make sense.

For every word, it passes through 2 layers so it is going to be processed. Since LSTM has a memory it is always having more context and with the regular RNN, sequence length would matter since it has little context and probability of guessing the character is lesser and so the perplexity is higher. With number of experiments and nature of data, the best match of sequence length would differ for each network.