Piradon (Tien) Liengtiraphan
CMPT 496L-111: Deep Learning w Tensor Flow
Assignment Number
Due: 2017-10-23
Instructor: Dr. Pablo Rivas

Midterm

Exercise Definition.

Perplexity *is a measurement of how well a probability distribution or probability model predicts a sample. It may be used to compare probability models. A low perplexity indicates the probability distribution is good at predicting the sample.*

$$2^{H(p)} = 2^{-\sum_x p(x) \log_2 p(x)}$$

Exercise 1a.

Go to the labs and find the one for RNNs entitled "Character Modeling with RNNs/LSTMs". You will be running a few experiments and will answer a few questions. Vary the number of hidden units exponentially as follows, 32, 64, 128, 256, and 512

Response. Perplexity of changing hidden states with a base sequence of 50.

Hidden States	Perplexity
32	1.820
64	1.629
128	1.421
256	1.222
512	0.887

Exercise 1b. *Compare results of each and answer the following questions:*

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

Response.

With a base sequence of 50 we can see that there general downwards trend in regards to the perplexity.

As the amount of hidden states increases. Example output below:

512 Hidden States

1. The truth, we, sit myself to wed: they are attend on m

2. The Tawardly chop then howly fault, I wish is thus. P

3. The Capulets were in complain to thy fair stolds, Sinc

4. The knavour of our deathsul up the city. But we were a

5. The Caples voice of learn,–ave's miltain-lesser; as y

The examples above are the results of the last few epoch for when the RNN was ran with 512 hidden states and a base sequence of 50, this set of parameters produced the lowest perplexity of **0.887**. The sentences being produced at this point seem to follow a trend/pattern of starting with "The" and have significantly less amount of gibberish when compared to the lower spectrum of hidden states. This is most likely due to the model being complex enough or having enough information that increasing the number of neurons would **not lead to over fitting**. The sequence size is also a relatively small number, reducing the chance of diluting the function's effect while processing the data. A reduction is perplexity means less chance of error in the prediction being made; the increased number of hidden states compounded with the relatively small sequence size allowed for a maximization of the RNN's resources, leading to better and more accurate predictions. Another possibility that adding more hidden states would decrease the perplexity is that, by adding more neurons each individual neuron can be more specialized in identifying specifics within the data input. The benefit of this is that by allowing each neuron to specialize, we increase the certainty to which they can identify their specialized field, thereby decreasing the overall perplexity.

Exercise 2a.

Go to the labs and find the one for RNNs entitled "Character Modeling with RNNs/LSTMs". You will be running a few experiments and will answer a few questions. Vary the length of the sequences as follows, 25, 50, and 75

Response. Perplexity of changing sequences with a base hidden state of 128.

Sequence #	Perplexity
25	1.339
50	1.417
75	1.360

Exercise 2b. *Compare results of each and answer the following questions:*

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

Response.

With a base hidden state of 128 we can see that there general upwards trend in regards to the perplexity. As the sequence size increases, RNN is required to process more input with a limited number of hidden states, leading to a dilution of the function's effect on the input.

512 Hidden States

1. The leaful, Cal shame fitper; Not lords was an passing
2. The loving any three as Ome keep, and is not, that abo
3. The holy, As they stay Thou wilts men and red! the lor
4. The bed the case the teared likele's lord!; and, glars
5. The none, follow utterming denied quoters! 'Cousine wa

As with before the results returned from the RNN seem to follow a trend of starting each and every line with "The". However, this is only a general trend with the limited data that we have on hand, that is not to say that this downwards trend will continue when other parameters are applied. The issue with this particular data set is that sequence size seems to have a relatively minute, even inconsistent, effect on the resulting perplexity. Take note the results return for our test case; we see that increasing the sequence size from 25 to 50 seems to have resulted in a upwards tick of 0.1, however, when we increase the sequence size to 75 the resulting perplexity actually goes down. What could be happening here is that by increasing the sequence length on an input that is never that length, the input might be padded with 0's or other characters to compensate for the length. This would interfere with the quality of results returned, possible negatively impacting the overall perplexity. Another possible interpretation of this is that by increasing the sequence size we are giving the RNN more context to infer from, allowing for better predictions, assuming that we have sufficient neurons to process the entirety of the data, otherwise this may lead to under fitting.