Midterm

Sami Ellougani

October 21, 2017

1 Hidden Units Exercise

1.1 32 Hidden Units

Sample: The sent or say inclef thy makesllim syo,
Train_Loss: 1.833
Time/Batch: 0.016

1.2 64 Hidden Units

Sample: The taul them! Pesal earry; As obstreen For thee. —Du
Train_Loss: 1.624
Time/Batch: 0.018

1.3 128 Hidden Units

Sample: The Sick. PETRUCHIO: Your mind, if it crave thy lover.
Train_Loss: 1.405
Time/Batch: 0.070

1.4 256 Hidden Units

Sample: The empty wood to me that time it far with me for the
Train_Loss: 1.202
Time/Batch: 0.162

1.5 512 Hidden Units

Sample: The Volces of my child. VALERIA: Your brother help
Train_Loss: 0.901
Time/Batch: 0.583

1.6 Questions

1.6.1 What happens to the perplexity? Why?

As the amount of hidden units grow, the perplexity of the sentence becomes more and more accurate. Around 128 Hidden units, that the train_loss rate started to drop drastically. I believe the perplexity became better as the hidden units grew because the hidden units created a larger vector for the output layer. This layer then decodes the results into some sort of probability of characters, which became more accurate as more hidden units were added.

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

As more hidden units were added, the sentences became much more similar to the ones that were in the input text. Around 128 hidden units, I noticed that names were being included to show that it was someone else's part in the script. This relates to the previous question because the train_loss rate became much lower as the hidden units increased, causing the sentences to be much more similar to the input text.

2 Sequence Exercise

2.1 25 Sequences

Sample: The waters were thinks enters; Into your parting deeds.
Train_Loss: 1.332
Time/Batch: 0.030

2.2 50 Sequences

Sample: The thee assend the scespy right. Blokens again. The.
Train_Loss: 1.407
Time/Batch: 0.070

2.3 75 Sequences

Sample: The offend to the king! PETRUCHIO: A sir, like the well.
Train_Loss: 1.364
Time/Batch: 0.098

2.4 Questions

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

The perplexity of the sentences are at its best when its sequence length is at its lowest, in this case 25. I believe this is the case because the sequence length dictates how many previous characters are going to be analyzed to determine which character is going to be chosen. However, if you look at the input text, a lot of the sentences are very small, so a sequence length too large might mix different parts of multiple sentences.

2.4.2 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?

The quality of the sentences are at its best when the sequence length is once again, at its lowest, in this case 25. I think relates closely to the perplexity because when a smaller group of previous characters are being used to analyze data, the train_loss rate is lower. Earlier, I said this was the case because the sentences in the input data are small, therefore the longer the sequence length of previous characters, the more of a mixture of sentences you are going to have.