

NLP with Deep Learning

Mini Project 3

Name: Poojith Reddy Maligireddy
UID: u1405749

1) I have implemented a simple 4-gram character level model in the file named ngram.py. In this implementation, I have created a dictionary where keys are three character sequence and values are dictionaries with keys as characters following those three character sequences and values are counts of those respective four character sequences. I have considered the length of that dictionary as number of parameters (number of conditional probabilities) and printed as below. But, if I consider the number of parameters as total number of distinct four character sequences then the number of parameters are 793901.

Below is the output after running ngram.py:

The number of parameters of simple 4-gram model is: 147245
The test set perplexity: 17.850721297150038

2) I have implemented and trained an LSTM with number of layers = {1, 2} using weighted losses in the file named main.py. The best hyperparameter (learning rate) for both one-layer and two-layer LSTM networks is 0.0001.

Below are the test perplexities and number of parameters for both one-layer and two-layer LSTM networks:

For one-layer LSTM network:

The test set perplexity: 2.690053476629208
The number of parameters of LSTM network with number of layers 1 is: 338686

For two-layer LSTM network:

The test set perplexity: 2.670453778606268
The number of parameters of LSTM network with number of layers 2 is: 660286

In the above results, we could see that the two-layer LSTM network has lower perplexity than one-layer LSTM network. This suggests that the two-layer LSTM performs better than one-layer LSTM network.

3) Below is the comparison between N-gram and LSTM models:

N-gram models are simple and easy to implement, memory efficient and only suitable for lower context dependencies with faster inference whereas LSTM models are complex, memory intensive and capable of capturing long-term dependencies in sequences with slower inference.

The test set perplexity of the LSTM model is significantly lower than that of the N-gram model. This suggests that the LSTM is outperforming the N-gram model as expected because LSTM models are capable of capturing long-term dependencies. Also, the LSTM networks have more number of parameters than the simple 4-gram model.

4) Below are the outputs of given five seed sequences using a sampling strategy for prediction which can also be found in file named seed_sequences_ans.txt:

(a) The little boy wasessed to the old brought, to cirriged observation, and will be the put characters for a headest no proper sister!\n Isance to pake out of the honishments call clapped a sambilians in the soleried slave

(b) Once upon a time inocently with Toolve, his father consequences are areed by 50% or love Revoluron seems from the great heating daught) with then, to anywhere I may be shall spent of wealthier. Fight she loved them aptai

(c) With the target inoce in them far of his way. In the celection would have it forwards more behind. He saved summies. They pave) occased after an acadd nothing, steady discovered out the creed together procus, seemed: t

(d) Capitals are big cities. For example, followed in the here of the origible of datist. If there were might certain they married, purshing elegated figulative. At there dutifully already made into a few testing in the stage parted charmaric

(e) A cheap alternative tomed interest that I can give new Strask. Henfed you know between the years to and his journey rocks helpened stepped from a risual far, who serve in allar them to man had throat he had been mothered t