CS563: Natural Language Processing

Indian Institute of Technology Patna

Assignment 4

Neural Language Model

REPORT

Group Members:

- Muhammed Sinan C K (1901CS38)
- Vaishakh Sreekanth Menon (1901CS68)
- Varsha Tumburu (1901CS69)

Aim

We target to implement bigram and trigram character-level language models with Feed Forward Neural Networks and predict the next letter of a name, based on the given dataset.

Datasets

Names Dataset

This is a text file with a set of names enlisted line after line. The aim of the dataset is to train the model's already existing names so that the model can predict the possible next letter based on the given sequence of string.

Feed Forward Neural Network

Information flows only in one direction in a feed forward neural network from input nodes through hidden nodes to output nodes. FFNNs consist of an input/output layer, multiple hidden layers, activation functions, weights and bias. Backpropagation is the major factor that helps in training allowing the model to be fit better utilizing its loss and gradient.

FFNNs are commonly used for classification and regression tasks, and can be used in combination with other techniques such as dropout, batch normalization, and early stopping to improve their performance.

For data preprocessing we first split it into 2 grams and into 3 grams we then convert the N-Gram into one hot encodings of size (N, 27), which are flattened into (N*27) size.

For this problem we use a simple feed forward network using fully connected linear layers. We use the ReLu activation function. The loss function used is Cross-Entropy.

The architecture for 2-gram model is as follows:

- Input : (batch_size=32, 2*27 = 54)

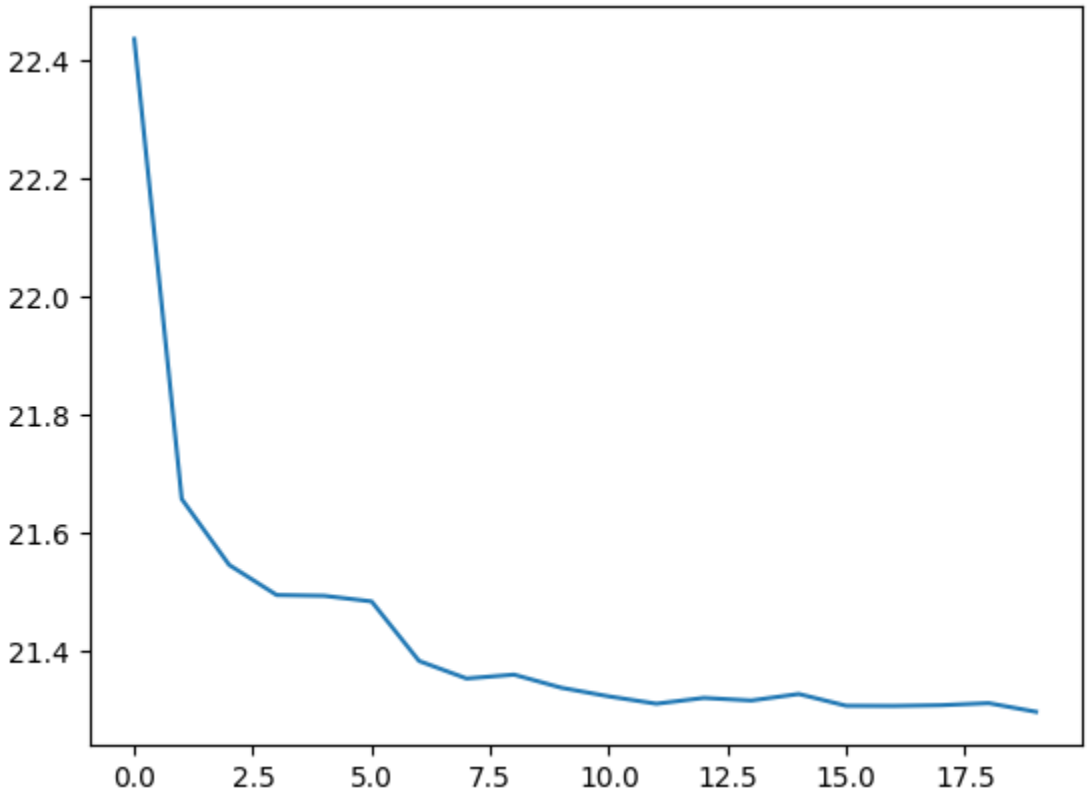
- Hidden layer 1: (128,64)
- Hidden layer 2: (64, 27)
- Output softmax layer

The architecture for 3-gram model is as follows:

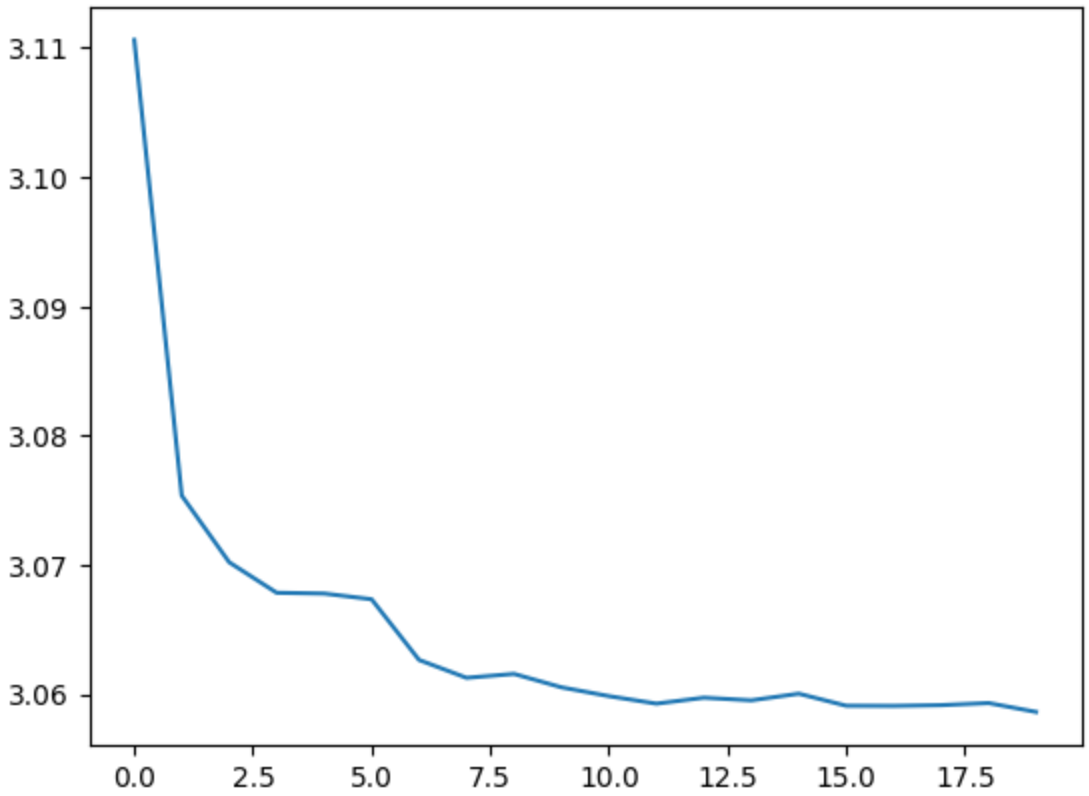
- Input : (batch_size=32, 3*27 = 81)
- Hidden layer 1: (128, 64)
- Hidden layer 2: (64, 27)
- Output softmax layer

Results (2-GRAM)

Perplexity



Train Loss

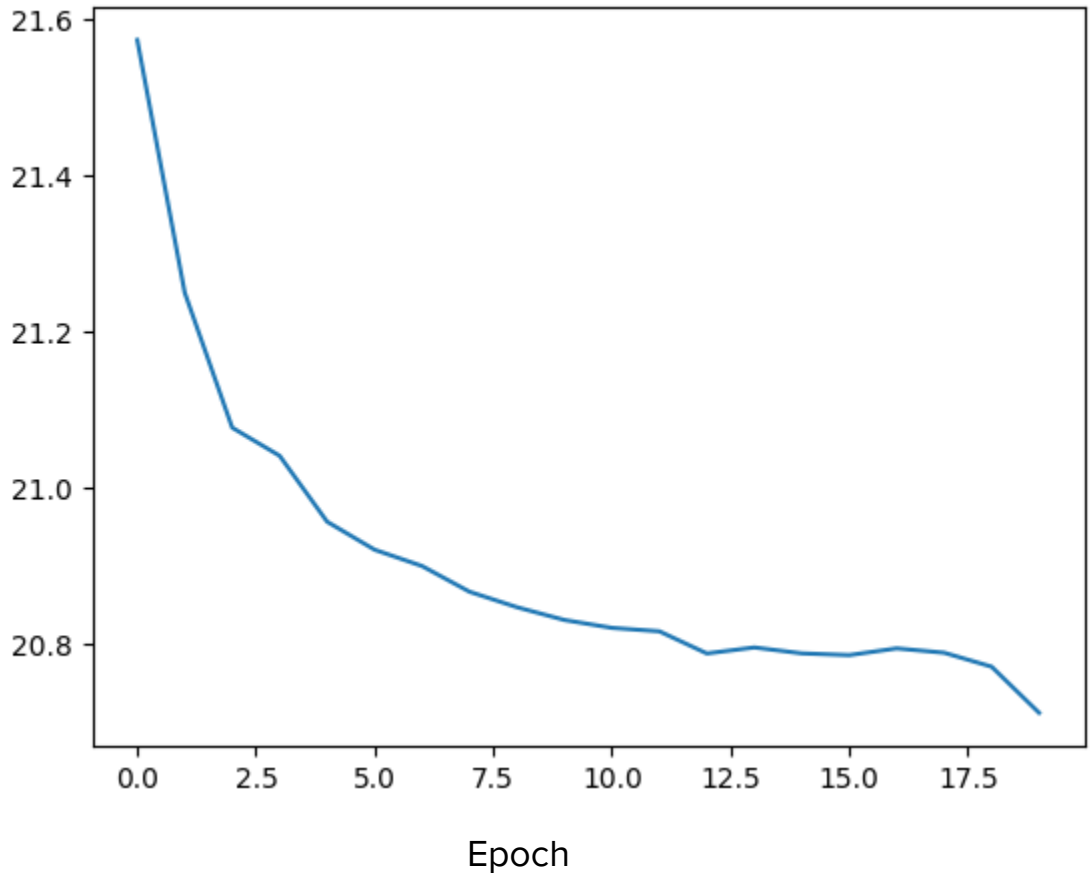


Test set results

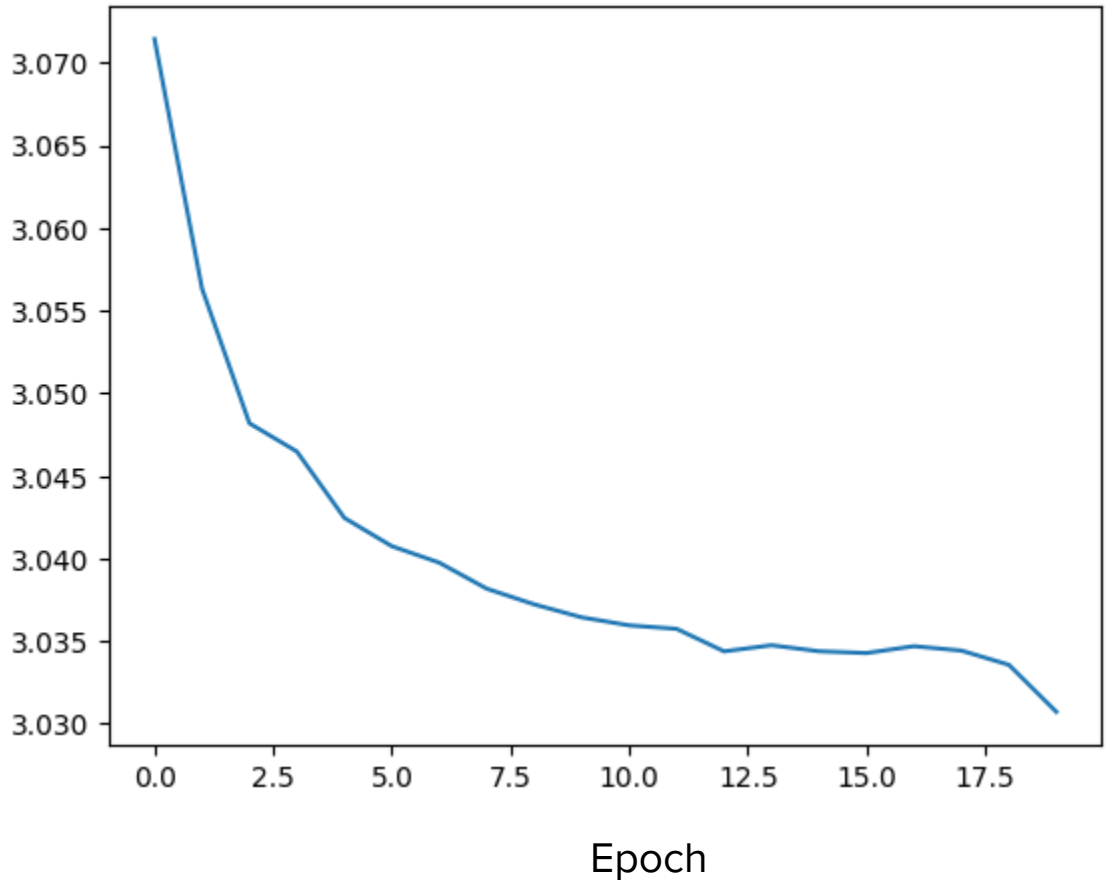
```
Test Loss: 3.0628, Test Accuracy: 29.11%
Test perplexity:  21.38652187585987
```

Results (3-GRAM)

Perplexity



Train Loss



Test set results

```
Test Loss: 3.0310, Test Accuracy: 32.02%
Test perplexity:  20.717172561766183
```