Report

💡 Name - Srija Mukhopadhyay
Roll Number - 2021114002

Model Implemented

Neural LM

This model utilises LSTMs to create a language model from the training data, optimising on the hyperparameters using a held out set or a validation/dev corpus and finally giving the results on the final testing corpus.

It was observed that for the Ulysses corpus, during training the following scores were obtained

```
The model has 6078358 trainable parameters
Train Perplexity: 792.088
Valid Perplexity: 487.577
Train Perplexity: 515.921
Valid Perplexity: 471.081
Train Perplexity: 460.453
Valid Perplexity: 412.488
Train Perplexity: 405.571
Valid Perplexity: 375.393
Train Perplexity: 372.424
Valid Perplexity: 354.919
```

For testing corpus, the following was observed

```
The model has 6078358 trainable parameters
  Test Perplexity: 354.654
```

On the other hand for the Pride and Prejudice corpus the following scores were obtained during the training phase

```
The model has 3769801 trainable parameters
  Train Perplexity: 706.628
  Valid Perplexity: 379.854
  Train Perplexity: 407.299
```

```
  Valid Perplexity: 372.861
  Train Perplexity: 383.404
  Valid Perplexity: 364.653
  Train Perplexity: 365.477
  Valid Perplexity: 344.488
  Train Perplexity: 337.760
  Valid Perplexity: 314.200
```

For the testing corpus, the following was obtained

```
The model has 3769801 trainable parameters
  Test Perplexity: 320.177
```

However, it should be observed that the model can be improved significantly with the help of more epochs, and probably a greater time to train the model.

It was seen that for the Pride and Prejudice corpus, a saturation point was reached for around 45 epochs.

Note

Other important details have been mentioned in the README document