

Advanced Natural Language Processing

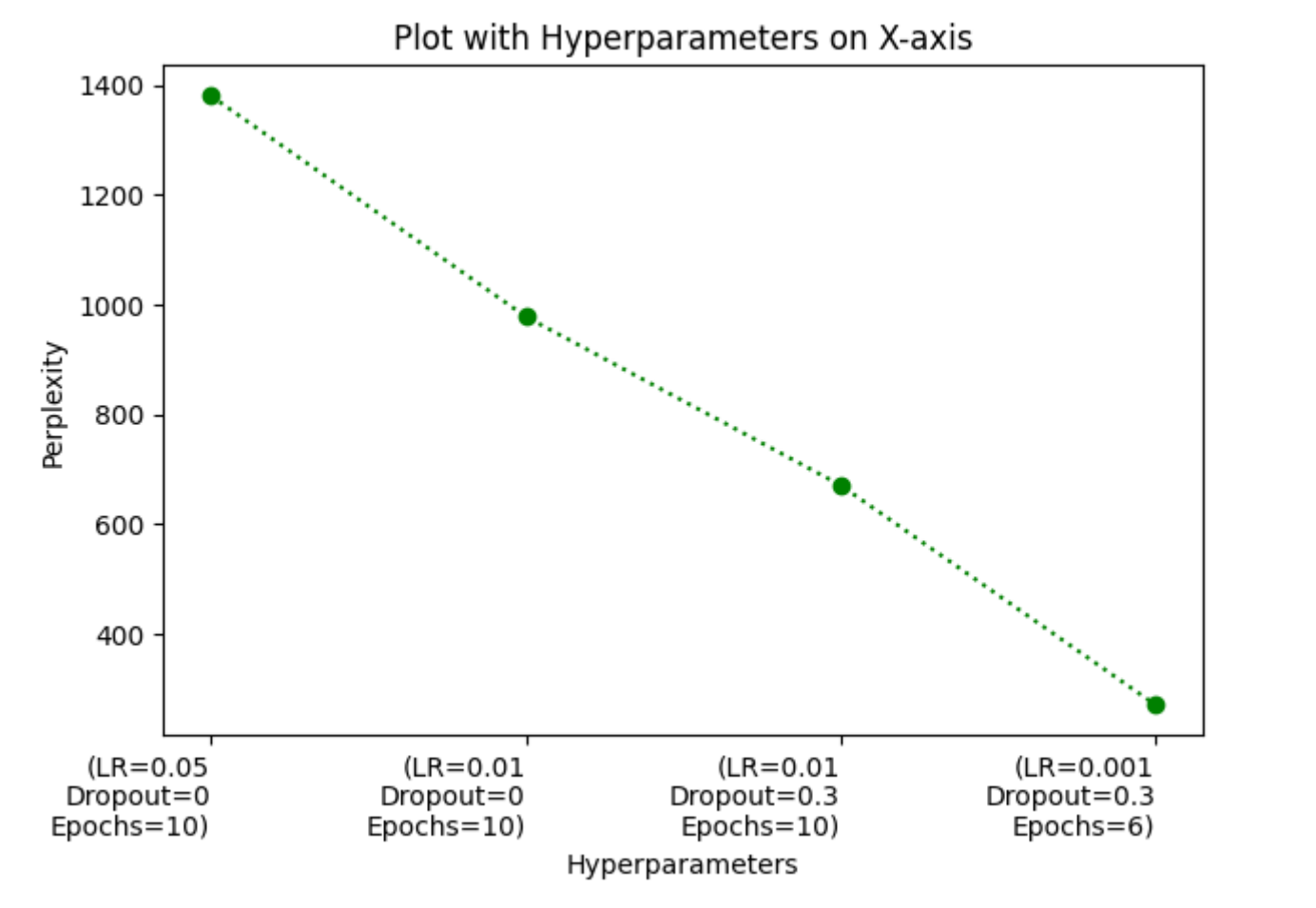
Assignment 1 - Report

Role of Hyperparameters

- **Learning Rate** - The learning rate determines the step size during the optimization process. It influences how much the model parameters are adjusted based on the gradients computed from the loss function. If the learning is too high, it may lead to overfitting of the model while a low learning rate leads to underfitting.
- **Dropout Rate** - Dropout is a regularization technique used to prevent overfitting by randomly setting a fraction of neurons to zero during each training iteration. This helps to ensure that the model does not rely too heavily on any specific neurons. By dropping out nodes, dropout encourages the network to learn redundant representations, thus improving the model's ability to generalize to unseen data.
- **Epochs** - The number of epochs represents how many times the entire dataset is processed through the network during training. It directly impacts how well the model learns from the data. Too few epochs may result in underfitting, where the model fails to capture the underlying patterns in the data. Conversely, too many epochs can lead to overfitting, where the model learns noise and specific patterns in the training data that do not generalize well.

Neural Language Model

- Train Perplexity: 185.4174
- Test Perplexity: 258.2892
- Validation Perplexity: 272.7334
- Values of Hyperparameters used -
 - EMBED_DIM = 500
 - LR = 0.001
 - BATCH_SIZE = 100
 - EPOCHS = 6
 - DROPOUT_RATE = 0.03



```
(main) gargi-shroff@gargi-shroff-Victus-HP-15:~/Assignments_Projects/ANLP/2022114009_assignment1$ python3 nnlm.py
Epoch [1/10], Train Loss: 6.2289
Epoch [1/10], Validation Loss: 5.9992
Perplexity: 403.0940246582031

Epoch [2/10], Train Loss: 5.6524
Epoch [2/10], Validation Loss: 6.0386
Perplexity: 419.29656982421875

Epoch [3/10], Train Loss: 5.4153
Epoch [3/10], Validation Loss: 6.1368
Perplexity: 462.5536804199219

Epoch [4/10], Train Loss: 5.2713
Epoch [4/10], Validation Loss: 6.2944
Perplexity: 541.5361328125

Epoch [5/10], Train Loss: 5.1766
Epoch [5/10], Validation Loss: 6.4367
Perplexity: 624.3388061523438

Epoch [6/10], Train Loss: 5.1058
Epoch [6/10], Validation Loss: 6.5911
Perplexity: 728.5723266601562

Epoch [7/10], Train Loss: 5.0554
Epoch [7/10], Validation Loss: 6.7591
Perplexity: 861.8924560546875

Epoch [8/10], Train Loss: 5.0145
Epoch [8/10], Validation Loss: 6.8922
Perplexity: 984.5819702148438

Epoch [9/10], Train Loss: 4.9858
Epoch [9/10], Validation Loss: 7.0256
Perplexity: 1125.080078125

Epoch [10/10], Train Loss: 4.9626
Epoch [10/10], Validation Loss: 7.1848
Perplexity: 1319.2305908203125

Perplexity on Test Data: 1380.7926025390625
```

```
• (main) gargi-shroff@gargi-shroff-Victus-HP-15:~/Assignments_Projects/ANLP/2022114009_assignment1$ python3 nnlm.py
Epoch [1/10], Train Loss: 6.3337
Epoch [1/10], Validation Loss: 6.1598
Perplexity: 473.341552734375

Epoch [2/10], Train Loss: 5.8223
Epoch [2/10], Validation Loss: 6.1905
Perplexity: 488.0797424316406

Epoch [3/10], Train Loss: 5.6674
Epoch [3/10], Validation Loss: 6.3028
Perplexity: 546.0880126953125

Epoch [4/10], Train Loss: 5.6056
Epoch [4/10], Validation Loss: 6.3445
Perplexity: 569.3499755859375

Epoch [5/10], Train Loss: 5.5809
Epoch [5/10], Validation Loss: 6.3968
Perplexity: 599.9071044921875

Epoch [6/10], Train Loss: 5.5745
Epoch [6/10], Validation Loss: 6.5357
Perplexity: 689.3045043945312

Epoch [7/10], Train Loss: 5.5795
Epoch [7/10], Validation Loss: 6.5749
Perplexity: 716.9041137695312

Epoch [8/10], Train Loss: 5.5859
Epoch [8/10], Validation Loss: 6.6973
Perplexity: 810.2415161132812

Epoch [9/10], Train Loss: 5.5992
Epoch [9/10], Validation Loss: 6.8065
Perplexity: 903.719970703125

Epoch [10/10], Train Loss: 5.6135
Epoch [10/10], Validation Loss: 6.8254
Perplexity: 920.9744873046875

Perplexity on Test Data: 978.4900512695312
```

```
• (main) gargi-shroff@gargi-shroff-Victus-HP-15:~/Assignments_Projects/ANLP/2022114009_assignment1$ python3 nnlm.py
Epoch [1/10], Train Loss: 6.4925
Epoch [1/10], Validation Loss: 6.3488
Perplexity: 571.814697265625

Epoch [2/10], Train Loss: 6.1168
Epoch [2/10], Validation Loss: 6.3503
Perplexity: 572.673095703125

Epoch [3/10], Train Loss: 5.9696
Epoch [3/10], Validation Loss: 6.4109
Perplexity: 608.4652099609375

Epoch [4/10], Train Loss: 5.9012
Epoch [4/10], Validation Loss: 6.4723
Perplexity: 646.9845581054688

Epoch [5/10], Train Loss: 5.8886
Epoch [5/10], Validation Loss: 6.5488
Perplexity: 698.382568359375

Epoch [6/10], Train Loss: 5.8880
Epoch [6/10], Validation Loss: 6.6061
Perplexity: 739.609619140625

Epoch [7/10], Train Loss: 5.8891
Epoch [7/10], Validation Loss: 6.6618
Perplexity: 781.9444580078125

Epoch [8/10], Train Loss: 5.8995
Epoch [8/10], Validation Loss: 6.6752
Perplexity: 792.4957275390625

Epoch [9/10], Train Loss: 5.9148
Epoch [9/10], Validation Loss: 6.7334
Perplexity: 840.0018920898438

Epoch [10/10], Train Loss: 5.9285
Epoch [10/10], Validation Loss: 6.7419
Perplexity: 847.1724243164062

Perplexity on Test Data: 671.245361328125
```

```
Epoch [1/6], Train Loss: 6.2022
Epoch [1/6], Validation Loss: 5.9598
Perplexity: 387.5179138183594

Epoch [2/6], Train Loss: 5.7944
Epoch [2/6], Validation Loss: 5.8596
Perplexity: 350.6007995605469

Epoch [3/6], Train Loss: 5.6802
Epoch [3/6], Validation Loss: 5.8307
Perplexity: 340.6119384765625

Epoch [4/6], Train Loss: 5.6244
Epoch [4/6], Validation Loss: 5.8126
Perplexity: 334.4833984375

Epoch [5/6], Train Loss: 5.5915
Epoch [5/6], Validation Loss: 5.8149
Perplexity: 335.2674255371094

Epoch [6/6], Train Loss: 5.5706
Epoch [6/6], Validation Loss: 5.8061
Perplexity: 332.3160705566406

Perplexity on Test Data: 271.0044250488281
```

RNN-based Language Model

- Train Perplexity: 72.5210
- Test Perplexity: 154.8519
- Validation Perplexity: 160.3617
- Values of Hyperparameters used -
 - EMDB_DIM = 100
 - LR = 0.01
 - BATCH_SIZE = 120
 - EPOCHS = 10
 - DROPOUT RATE = 0.03

Transformer Decoder based Language Model

- Train Perplexity: 72.5210
- Test Perplexity: 154.8519
- Validation Perplexity: 160.3617
- Values of Hyperparameters used -
 - BATCH_SIZE = 128
 - MAX_LEN = 60
 - LAYERS = 5
 - HEADS = 5
 - FFN_DIM = 300
 - EMBD_DIM = 100
 - EPOCHS = 20
 - LR = 0.01

Analysis

- Since Neural Language Models uses a simple architecture with 5 word context window they are unable to capture long term dependencies. The perplexity is higher than RNNs and Transformer models, indicating poorer performance in comparison to those.

- Language is a temporal Phenomenon (continuous input stream). Feed forward neural networks assume simultaneous access to all aspects of their input (by using a sliding context window approach). RNNs (LSTM architecture in this case) have a mechanism that deals directly with the sequential nature of language, allowing them to handle the temporal nature of language without the use of arbitrary fixed-sized windows. Hence they show better perplexity scores than Feed Forward Neural Language Model.
- Transformers can process all tokens in a sequence simultaneously due to their self-attention mechanism. This parallelism significantly speeds up training and inference times, especially with large datasets. Use self-attention also allows to capture relationships between tokens regardless of their distance in the sequence. This allows them to handle long-range dependencies effectively and maintain context across long sequences. Multihead mechanism in transformers allows the model to capture various types of relationships between tokens. Hence, this language model outperforms the other two.