Introduction to NLP

Assignment -2

Name :- Suyash Sethia

Roll No:- 2021114010

Report

POS TAGGING

Hyperparameters

After lots of training and validation on the dev set,I found that the best set of Hyperparameters are

```
EMBEDDING_DIM = 300
HIDDEN_DIM = 300
EPOCHS = 15
LEARNING_RATE = 0.1
```

where,

EMBEDDING_DIM = dimension of the input given to the LSTM model

HIDDEN_DIM = Dimension of the output (POS tag) given by the model

LEARNING_RATE = The learning rate determines how fast the model learns from the data

EPOCHS = Number of times the model is trained on the entire dataset

Scores

LOSS per EPOCH

Epoch: 1 Loss: 0.0005261500482447445
Epoch: 2 Loss: 0.0002194686676375568
Epoch: 3 Loss: 0.0001353075640508905
Epoch: 4 Loss: 9.888481145026162e-05
Epoch: 5 Loss: 7.708981866016984e-05
Epoch: 6 Loss: 6.049320654710755e-05
Epoch: 7 Loss: 4.779941082233563e-05
Epoch: 8 Loss: 3.845760875265114e-05
Epoch: 9 Loss: 3.138031024718657e-05
Epoch: 10 Loss: 2.5986564651248045e-05
Epoch: 11 Loss: 2.1710247892769985e-05
Epoch: 12 Loss: 1.8536509742261842e-05
Epoch: 13 Loss: 1.6271666027023457e-05
Epoch: 14 Loss: 1.4766729691473302e-05
Epoch: 15 Loss: 1.3708801816392224e-05

ACCURACY

accuracy on train set (96.9910595005652)
accuracy on test set (95.77507598784194)
accuracy on the validation set (95.78567128236003)

For Training Data

```
              precision  recall   f1-score   support
     ADJ         0.98      0.94      0.96      1632
     ADP         0.97      0.99      0.98     10791
```

```
      ADV       0.95      0.93      0.94       431
      AUX       0.94      0.97      0.96      1732
    CCONJ       1.00      1.00      1.00       751
      DET       0.84      0.98      0.91      3805
     INTJ       0.98      0.87      0.92       319
     NOUN       1.00      1.00      1.00      8621
      NUM       0.99      1.00      1.00       933
     PART       0.89      0.99      0.94       366
     PRON       0.98      0.80      0.88      3022
    PROPN       1.00      1.00      1.00     11657
     VERB       0.98      0.92      0.95      4595

accuracy                           0.97     48655
macro avg       0.96      0.95      0.96     48655
weighted avg    0.97      0.97      0.97     48655
```

Final F1 Score = 0.97

Final Support = 48655

For Validation Data

```
              precision    recall  f1-score   support
      ADJ       0.97      0.91      0.94       227
      ADP       0.95      0.99      0.97      1415
      ADV       0.91      0.85      0.88        59
      AUX       0.97      0.97      0.97       266
    CCONJ       1.00      1.00      1.00       107
      DET       0.85      0.98      0.91       568
     INTJ       1.00      0.91      0.96        35
     NOUN       0.97      0.99      0.98      1143
      NUM       0.97      0.95      0.96       131
     PART       0.90      0.95      0.92        73
     PRON       0.99      0.78      0.87       414
    PROPN       0.99      0.99      0.99      1551
      SYM       0.00      0.00      0.00         2
     VERB       0.98      0.87      0.92       653

accuracy                           0.96      6644
macro avg       0.89      0.87      0.88      6644
weighted avg    0.96      0.96      0.96      6644
```

Final F1 Score = 0.96

Final Support = 6644

FOR Testing Data

```
              precision    recall  f1-score   support
     ADJ       0.94      0.95      0.95       220
     ADP       0.96      0.99      0.97      1434
     ADV       0.92      0.75      0.83        76
     AUX       0.95      0.96      0.96       256
   CCONJ       1.00      1.00      1.00       109
     DET       0.82      0.98      0.89       512
    INTJ       1.00      0.89      0.94        36
    NOUN       0.97      0.99      0.98      1166
     NUM       0.95      0.94      0.94       127
    PART       0.96      0.98      0.97        56
    PRON       0.97      0.77      0.85       392
   PROPN       0.99      0.99      0.99      1567
    VERB       0.98      0.86      0.92       629

accuracy                           0.96      6580
macro avg      0.96      0.93      0.94      6580
weighted avg   0.96      0.96      0.96      6580
```

Final F1 Score = 0.96

Final Support = 6580

Analysis

Hyperparameter tuning is the process of finding the best set of hyperparameters for a given machine-learning model that maximizes its performance on a particular task

The end goal of hyperparameter tuning in a classification task is to find the set of hyperparameters that maximizes the model's performance

Here we consider the F1 score, we thrive to maximize the F1 score.

F1 score is the harmonic mean of precision and recall and ranges from 0 to 1, with higher values indicating better performance

Tuning the Hyperparameters

1. **Embedding size - T**he embedding size can affect the model's ability to capture important features in the data. Generally, larger embedding sizes can capture more complex relationships but can also increase the computational cost of the model.

That's why the embedding size of 300 was selected as it could capture all features as well as be computationally feasible

2. **Hidden Dimensions =** The number of hidden layers and the number of neurons in each hidden layer determines the complexity of the model. Adding more layers can make the model more expressive but can also increase the risk of overfitting and since the size of the training model was not large enough the hidden dimensions were selected as 300 which does not let the model overfit and also is feasible according to the complexity of POS tagging problem

3. **Learning rate =** The learning rate determines how fast the model learns from the data. A higher learning rate can lead to faster convergence, but it can also cause the model to overshoot the optimal solution and get stuck in a suboptimal solution. A lower learning rate can lead to slower convergence, So a Learning rate of 0.1 was selected through training and Keeping in mind the F1 score and the probability of overfitting.
4. **Epochs:-** Epochs are the Number of times the model is trained on the entire dataset, though more epochs decrease the loss at one point in time the gradient of loss change is very less and also it is computationally expensive so an optimal Epochs of 15 was chosen.

the hyperparameter Tuning which led to the best F1 score was selected keeping all the above factors in mind