

3 Hyperparameter Tuning

3.1 Feed Forward Neural Network POS Tagger

Configuration – 1:

Embedding Dimension = 300

Number of hidden Layer = 5

Hidden layer size = 20

Activation Function = ReLU

Dev Accuracy: 0.9635651912074676

Dev Classification Report:

		precision	recall	f1-score	support
	0	0.98	0.99	0.99	414
	1	0.99	0.98	0.98	266
	2	0.99	0.98	0.98	568
	3	0.98	0.97	0.98	1143
	4	0.94	0.97	0.96	1415
	5	0.97	1.00	0.98	1551
	6	0.98	0.88	0.92	653
	7	0.90	0.94	0.92	131
	8	0.96	0.85	0.90	227
	9	1.00	0.99	1.00	107
	10	0.80	0.75	0.77	59
	11	0.69	0.99	0.81	73
	12	0.95	1.00	0.97	35
	accuracy			0.96	6642
	macro avg	0.93	0.94	0.94	6642
	weighted avg	0.97	0.96	0.96	6642

Dev Confusion Matrix:

[illegible]

```

Accuracy: 0.9635651912074676
Precision (Micro) on Development Set: 0.9635651912074676
Recall (Micro) on Development Set: 0.9635651912074676
F1-score (Micro) on Development Set: 0.9635651912074676
Precision (Macro) on Development Set: 0.9333036005157523
Recall (Macro) on Development Set: 0.9442916990973218
F1-score (Macro) on Development Set: 0.9361772014945017

Test Set Metrics:
Accuracy: 0.9627659574468085
Precision (Micro) on Test Set: 0.9627659574468085
Recall (Micro) on Test Set: 0.9627659574468085
F1-score (Micro) on Test Set: 0.9627659574468085
Precision (Macro) on Test Set: 0.9236882908695393
Recall (Macro) on Test Set: 0.946540302085068
F1-score (Macro) on Test Set: 0.9311911510993374

```

Configuration – 2:

Embedding Dimension = 500

Number of hidden Layer = 10

Hidden layer size = 60

Activation Function = ReLU

```

Dev Accuracy: 0.9682324601023788
Dev Classification Report:

```

		precision	recall	f1-score	support
	0	0.98	1.00	0.99	414
	1	0.99	0.97	0.98	266
	2	1.00	0.97	0.98	568
	3	0.98	0.98	0.98	1143
	4	0.94	0.99	0.96	1415
	5	0.98	0.99	0.99	1551
	6	0.98	0.89	0.93	653
	7	0.98	0.94	0.96	131
	8	0.95	0.84	0.89	227
	9	1.00	1.00	1.00	107
	10	0.68	0.75	0.71	59
	11	0.92	0.95	0.93	73
	12	0.97	0.94	0.96	35
accuracy				0.97	6642
macro avg		0.95	0.94	0.94	6642
weighted avg		0.97	0.97	0.97	6642

```

Dev Confusion Matrix:
[[ 412   0   1   0   1   0   0   0   0   0   0   0   0]
 [  0 257   0   0   0   0   7   0   1   0   0   1   0]
 [  4   0 551   0  12   1   0   0   0   0   0   0   0]
 [  0   0   0 1119   2  17   3   0   2   0   0   0   0]
 [  1   0   1   1 1405   0   1   0   0   0   0   6   0]
 [  0   0   0   3   1 1543   1   0   1   0   2   0   0]
 [  3   0   0   2   69   1 578   0   0   0   0   0   0]
 [  1   0   0   3   0   3   0 123   0   0   1   0   0]
 [  0   1   0  16   2   2   0   0 190   0  16   0   0]
 [  0   0   0   0   0   0   0   0   0 107   0   0   0]
 [  0   1   0   1   4   1   0   2   6   0  44   0   0]
 [  0   0   0   0   3   0   0   0   0   0   0  69   1]
 [  0   1   0   0   0   0   0   0   0   0   1   0 33]]

```

```

Development Set Metrics:
Accuracy: 0.9682324601023788
Precision (Micro) on Development Set: 0.9682324601023788
Recall (Micro) on Development Set: 0.9682324601023788
F1-score (Micro) on Development Set: 0.9682324601023788
Precision (Macro) on Development Set: 0.9494830604187805
Recall (Macro) on Development Set: 0.9379299201547272
F1-score (Macro) on Development Set: 0.9429574731034785

```

```

Test Set Metrics:
Accuracy: 0.9679331306990882
Precision (Micro) on Test Set: 0.9679331306990882
Recall (Micro) on Test Set: 0.9679331306990882
F1-score (Micro) on Test Set: 0.9679331306990882
Precision (Macro) on Test Set: 0.9522796156259042
Recall (Macro) on Test Set: 0.9421440028850038
F1-score (Macro) on Test Set: 0.9464595869338013

```

Configuration – 3:

Embedding Dimension = 200

Number of hidden Layer = 5

Hidden layer size = 60

Activation Function = Tan h

```

Dev Accuracy: 0.9623607347184583
Dev Classification Report:

```

	precision	recall	f1-score	support
0	0.99	0.99	0.99	414
1	0.97	0.96	0.97	266
2	0.98	0.97	0.98	568
3	0.98	0.98	0.98	1143
4	0.94	0.99	0.96	1415
5	0.97	0.99	0.98	1551
6	0.97	0.86	0.91	653
7	0.98	0.92	0.95	131
8	0.97	0.81	0.88	227
9	1.00	0.98	0.99	107
10	0.65	0.73	0.69	59
11	0.87	0.97	0.92	73
12	0.92	1.00	0.96	35
accuracy			0.96	6642
macro avg	0.94	0.93	0.93	6642
weighted avg	0.96	0.96	0.96	6642

```

Dev Confusion Matrix:
[[ 410  0  3  0  0  0  1  0  0  0  0  0  0]
 [  0 255  0  1  0  0 10  0  0  0  0  0  0]
 [  3  0 552  0 10  2  0  0  0  0  1  0  0]
 [  0  1  0 1117  1 18  2  0  0  0  4  0  0]
 [  0  0  5  0 1398  0  0  0  0  0  0 11  1]
 [  0  0  0  4  1 1541  0  1  1  0  2  0  1]
 [  2  5  1  7  77  0 561  0  0  0  0  0  0]
 [  0  0  0  1  1  6  0 121  0  0  1  0  1]
 [  0  0  0 11  0 18  0  0 183  0 15  0  0]
 [  0  0  1  0  0  1  0  0  0 105  0  0  0]
 [  0  1  1  3  0  3  2  1  5  0 43  0  0]
 [  1  0  0  0  1  0  0  0  0  0  0 71  0]
 [  0  0  0  0  0  0  0  0  0  0  0  0 35]]

```

```

Development Set Metrics:
Accuracy: 0.9623607347184583
Precision (Micro) on Development Set: 0.9623607347184583
Recall (Micro) on Development Set: 0.9623607347184583
F1-score (Micro) on Development Set: 0.9623607347184583
Precision (Macro) on Development Set: 0.9375977568924738
Recall (Macro) on Development Set: 0.9347134651986924
F1-score (Macro) on Development Set: 0.934634804564944

```

```

Test Set Metrics:
Accuracy: 0.9665653495440729
Precision (Micro) on Test Set: 0.9665653495440729
Recall (Micro) on Test Set: 0.9665653495440729
F1-score (Micro) on Test Set: 0.9665653495440729
Precision (Macro) on Test Set: 0.9504698566061202
Recall (Macro) on Test Set: 0.9445029764311736
F1-score (Macro) on Test Set: 0.9465153811464975

```

➤ Test evaluation matrix

Evaluation matrix for test set on model configuration that runs best on dev i.e. Configuration – 2

```

Test Accuracy: 0.9679331306990882
Test Classification Report:

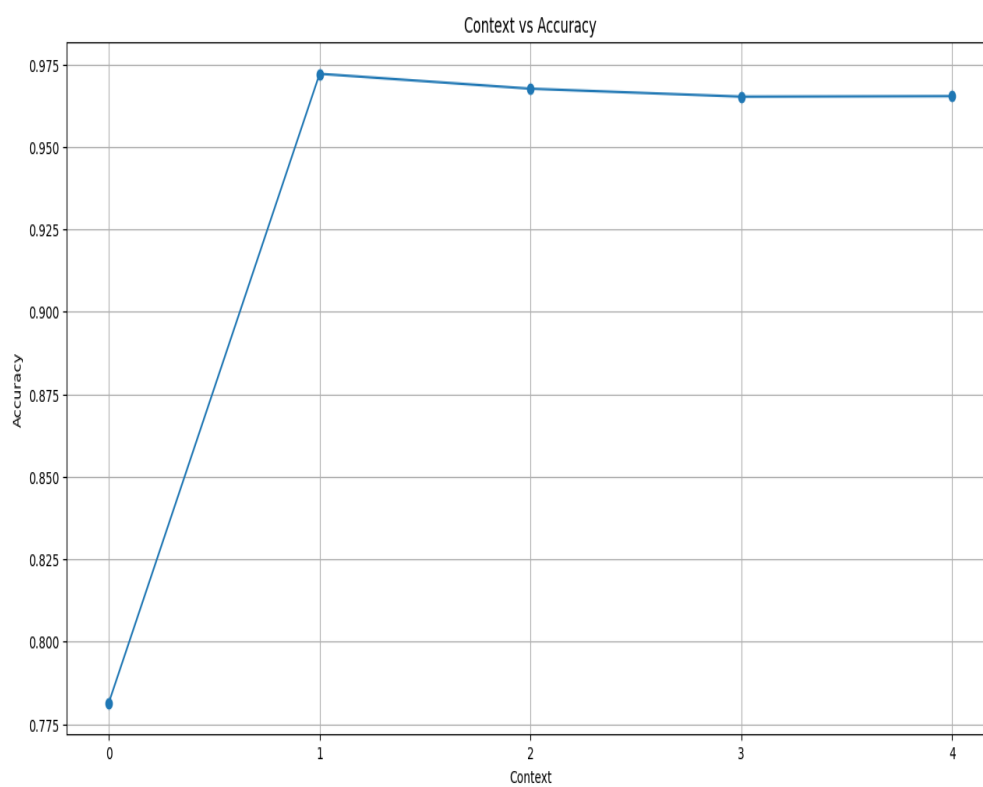
```

	precision	recall	f1-score	support
0	0.97	0.99	0.98	392
1	0.96	0.99	0.98	256
2	0.98	0.98	0.98	512
3	0.97	0.97	0.97	1166
4	0.95	0.99	0.97	1434
5	0.98	0.99	0.99	1567
6	0.99	0.87	0.93	629
7	0.95	0.83	0.88	127
8	0.93	0.91	0.92	220
9	0.99	1.00	1.00	109
10	0.81	0.76	0.78	76
11	0.98	0.98	0.98	56
12	0.92	0.97	0.95	36
accuracy			0.97	6580
macro avg	0.95	0.94	0.95	6580
weighted avg	0.97	0.97	0.97	6580

```

Test Confusion Matrix:
[[ 387  0  4  0  0  0  0  0  1  0  0  0  0  0]
 [  0 254  0  0  0  0  1  0  0  0  0  0  0  1]
 [  3  0 504  0  3  1  0  0  0  0  0  0  0  1]
 [  0  0  1 1134  2 21  2  2  3  0  1  0  0  0]
 [  2  0  5  0 1423  0  1  0  0  0  1  1  1  1]
 [  0  0  0  7  0 1556  0  2  1  0  1  0  0  0]
 [  2  7  0  5  61  1 549  1  1  0  2  0  0  0]
 [  1  0  1 12  0  2  2 105  3  1  0  0  0  0]
 [  1  1  0  6  2  1  0  0 200  0  9  0  0  0]
 [  0  0  0  0  0  0  0  0  0 109  0  0  0  0]
 [  1  2  0  1  1  5  0  0  8  0  58  0  0  0]
 [  0  0  0  0  1  0  0  0  0  0  0  55  0  0]
 [  0  1  0  0  0  0  0  0  0  0  0  0  35  1]

```



3.2 Recurrent Neural Network POS Tagger

Configuration – 1:

Embedding Dimension = 200

Number of layers = 2

Hidden Dimension = 10

Bidirectionality= True

Dev Accuracy: 0.9854287458040891

Dev Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	6880
1	0.99	0.97	0.98	266
2	0.98	0.98	0.98	568
3	0.98	0.98	0.98	1143
4	0.94	0.99	0.97	1415
5	0.99	0.99	0.99	1551
6	0.97	0.89	0.92	653
7	0.97	0.97	0.97	131
8	0.98	0.88	0.93	227
9	1.00	1.00	1.00	107
10	0.82	0.83	0.82	59
11	0.89	0.89	0.89	73
12	1.00	1.00	1.00	35
accuracy			0.99	13108
macro avg	0.96	0.95	0.96	13108
weighted avg	0.99	0.99	0.99	13108

Dev Confusion Matrix:

[illegible]

```

Development Set Metrics:
Accuracy: 0.9854287458040891
Precision (Micro) on Development Set: 0.9854287458040891
Recall (Micro) on Development Set: 0.9854287458040891
F1-score (Micro) on Development Set: 0.9854287458040891
Precision (Macro) on Development Set: 0.9617141056115728
Recall (Macro) on Development Set: 0.9516118806538055
F1-score (Macro) on Development Set: 0.9562585335218348

Test Set Metrics:
Accuracy: 0.986697965571205
Precision (Micro) on Test Set: 0.986697965571205
Recall (Micro) on Test Set: 0.986697965571205
F1-score (Micro) on Test Set: 0.986697965571205
Precision (Macro) on Test Set: 0.9662051629466363
Recall (Macro) on Test Set: 0.9515748507028904
F1-score (Macro) on Test Set: 0.9580793408955043
PS C:\Users\Harsh\Desktop\NLP\Assignment-2>

```

Configuration – 2:

Embedding Dimension = 400

Number of layers = 3

Hidden Dimension = 20

Bidirectionality= True

```

Dev Accuracy: 0.9851235886481537
Dev Classification Report:
              precision    recall  f1-score   support

     0           1.00        1.00        1.00        6880
     1           0.98         0.96         0.97         266
     2           0.99         0.98         0.98         568
     3           0.98         0.98         0.98        1143
     4           0.95         0.99         0.97        1415
     5           0.99         0.99         0.99        1551
     6           0.95         0.89         0.92         653
     7           0.95         0.97         0.96         131
     8           0.95         0.88         0.91         227
     9           1.00         1.00         1.00         107
    10           0.80         0.80         0.80          59
    11           0.90         1.00         0.95          73
    12           0.96         0.74         0.84          35

 accuracy          0.99        13108
 macro avg         0.95         0.94         0.94        13108
 weighted avg      0.99         0.99         0.98        13108

Dev Confusion Matrix:
[[6877  0  2  0  0  0  1  0  0  0  0  0  0  0]
 [  0 255  1  0  0  0 10  0  0  0  0  0  0  0]
 [  1  0 555  0 11  0  0  1  0  0  0  0  0  0]
 [  0  0  0 1120  1 15  5  0  1  0  1  0  0  0]
 [  1  1  1  1 1401  2  0  0  0  0  0  8  0  0]
 [  1  0  0  2  2 1543  2  0  0  0  1  0  0  0]
 [  0  3  0  2  60  1 583  1  3  0  0  0  0  0]
 [  0  0  0  1  1  2  0 127  0  0  0  0  0  0]
 [  0  0  0  13  0  2  0  2 199  0 10  0  0  1]
 [  0  0  0  0  0  0  0  0  0 107  0  0  0  0]
 [  0  1  0  1  0  1  1  2  6  0 47  0  0  0]
 [  0  0  0  0  0  0  0  0  0  0  0 73  0  0]
 [  0  0  0  0  0  0  0  9  0  0  0  0 261  0]

```

Configuration – 3:

Embedding Dimension = 200

Number of layers = 4

Hidden Dimension = 10

Bidirectionality= True

Dev Accuracy: 0.9847421422032346

Dev Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	6880
1	0.98	0.96	0.97	266
2	0.99	0.97	0.98	568
3	0.98	0.98	0.98	1143
4	0.96	0.99	0.97	1415
5	0.98	0.99	0.99	1551
6	0.96	0.91	0.93	653
7	0.98	0.85	0.91	131
8	0.96	0.87	0.91	227
9	0.99	0.98	0.99	107
10	0.71	0.86	0.78	59
11	0.86	0.99	0.92	73
12	0.92	1.00	0.96	35
accuracy			0.98	13108
macro avg	0.94	0.95	0.95	13108
weighted avg	0.99	0.98	0.98	13108

Dev Confusion Matrix:

[illegible]

➤ Test evaluation matrix

Evaluation matrix for test set on model configuration that runs best on dev i.e.
Configuration – 1

```
Test Accuracy: 0.986697965571205
Test Classification Report:
              precision    recall  f1-score   support

     0           1.00       1.00       1.00       6592
     1           0.97       1.00       0.99        256
     2           0.97       0.99       0.98        512
     3           0.99       0.99       0.99       1166
     4           0.96       0.99       0.98       1434
     5           0.99       0.99       0.99       1567
     6           0.98       0.89       0.93        629
     7           0.94       0.91       0.93        127
     8           0.93       0.94       0.93        220
     9           1.00       0.99       1.00        109
    10           0.87       0.72       0.79         76
    11           0.96       0.98       0.97         56
    12           1.00       0.97       0.99         36

 accuracy              0.99       12780
 macro avg           0.97       0.95       0.96       12780
 weighted avg        0.99       0.99       0.99       12780

Test Confusion Matrix:
[[6585   0    6   0   0   0   0   1   0   0   0   0   0]
 [   0  256   0   0   0   0   0   0   0   0   0   0   0]
 [   1    0  507   0   1   2   1   0   0   0   0   0   0]
 [   0    0    0 1150   0   9   4   2   1   0   0   0   0]
 [   0    0    8    1 1424   0   0   0   0   0   0   1   0]
 [   0    0    3    6    0 1554   1   1   1   0   1   0   0]
 [   0    6    0    5   56    1  558   2   0   0   1   0   0]
 [   2    0    0    0   0    3    3  116   2   0   0   1   0]
 [   0    0    1    2   0    2    1    1  207   0   6   0   0]
 [   0    0    0    0   0    1    0    0    0  108   0   0   0]
 [   0    1    0    1    3    4    0    0   12    0   55   0   0]
 [   0    0    0    0    1    0    0    0    0    0   0   55   0]
 [   0    0    0    0    0    0    1    0    0    0   0   0   35]]
```

Epochs V/S Dev Accuracy Plots For RNN

Config – 1

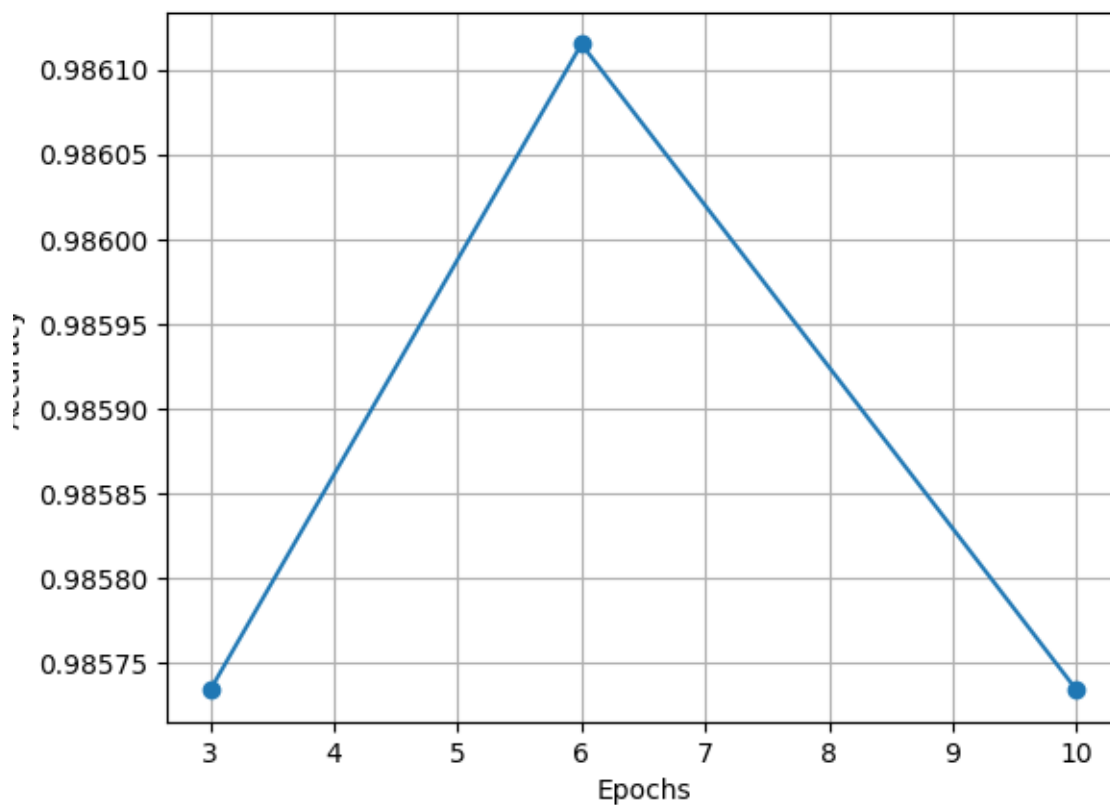
Embedding Dim = 200

Num of Layer = 2

Hidden Dim = 10

Bidirectional = True

For configuration : Embedding Dim = 200, Num of Layer = 2,
Hidden Dim = 10, Bidirectional = True



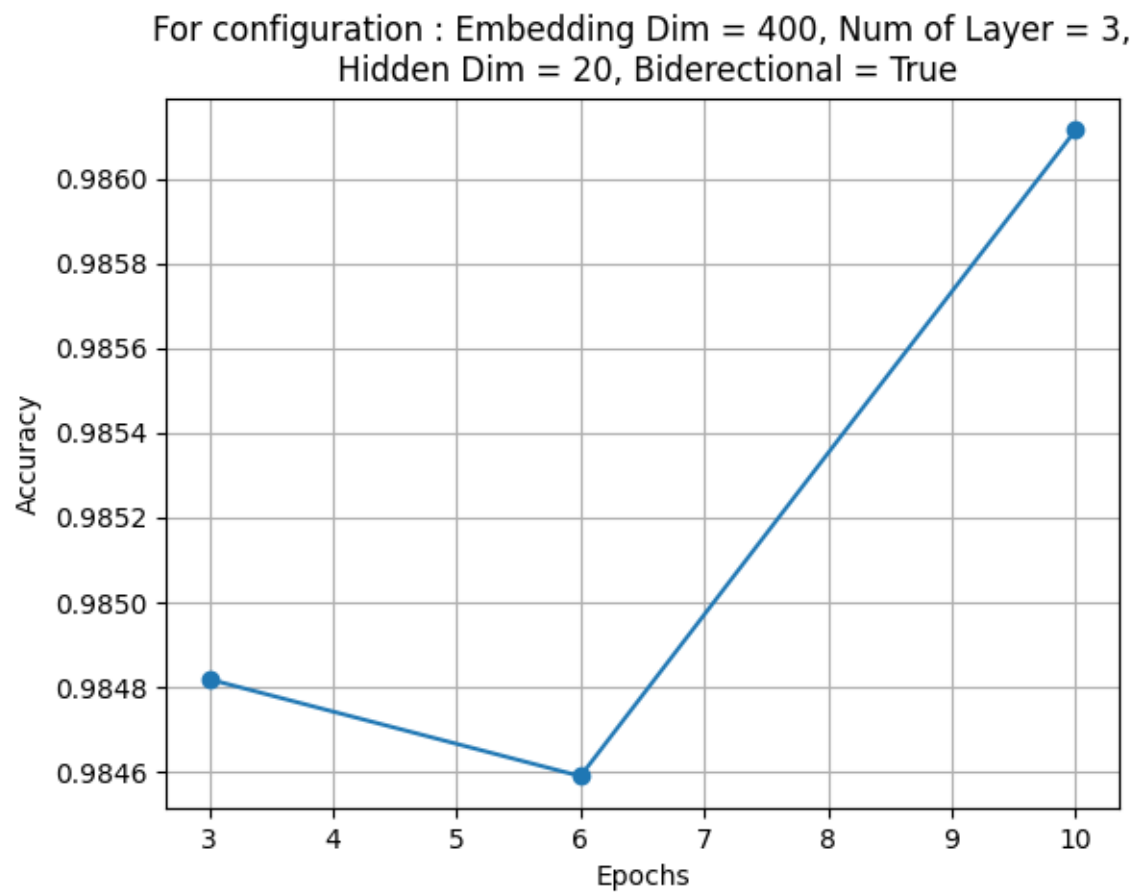
Config – 2

Embedding Dim = 400

Num of Layer = 3

Hidden Dim = 20

Bidirectional = True



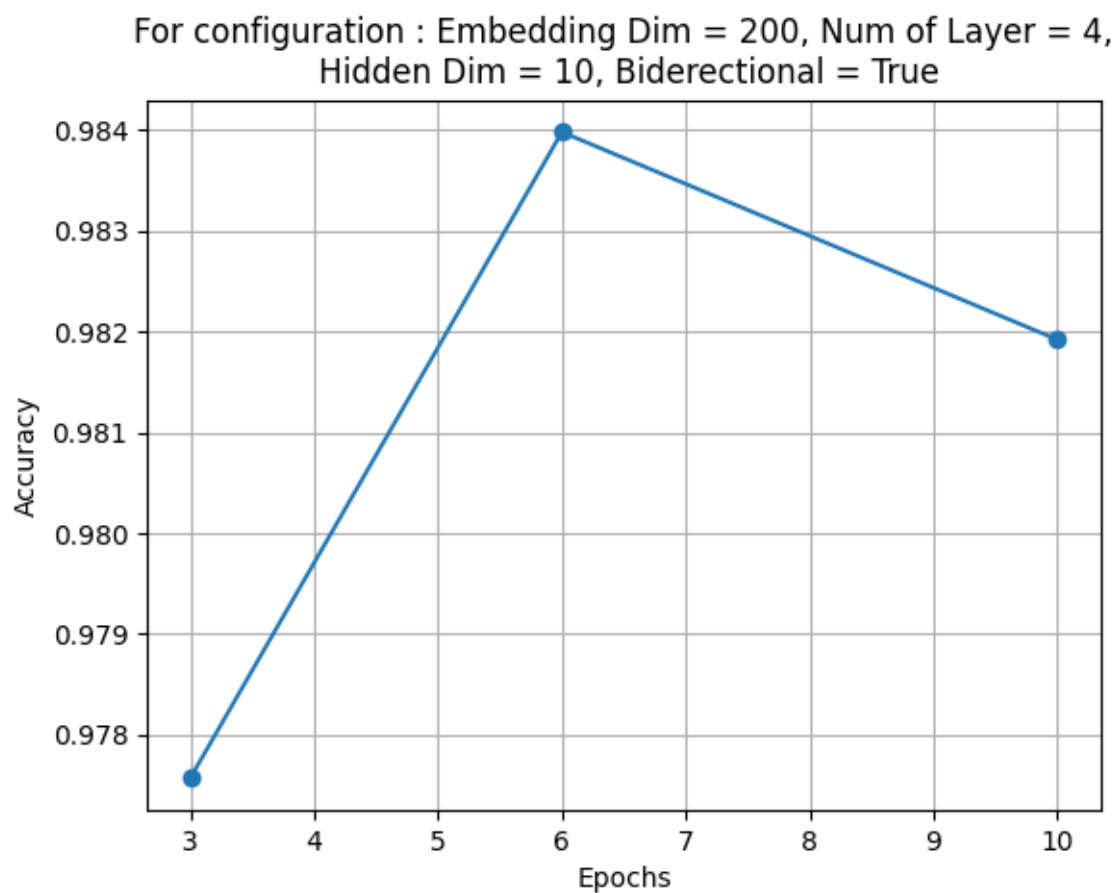
Config – 3

Embedding Dim = 200

Num of Layer = 4

Hidden Dim = 10

Bidirectional = True



5 Analysis

5.1 Feed Forward Neural Network POS Tagger

- **Increasing Model Complexity:**
 - As we increase the complexity of the model by raising the embedding dimension number of hidden layers and the size of hidden layers, we observe a slight increase in development accuracy, macro average, weighted average.
- **Test Accuracy Improvement:**
 - Additionally, the test accuracy improves for the complex model.
- **Activation Function Impact:**
 - If we utilize the Tanh activation function instead of ReLU, there is a slight decrease in accuracy.
- **Effect of Context window:**
 - The accuracy of the model is lowest when both p and s are set to 0, while it is highest when both are set to 1.
 - As we continue to increase p and s , for example, from $\{2, 4\}$, we observe a slight decrease in accuracy.

5.2 Recurrent Neural Network POS Tagger

- **Effect of Number of Layers:**
 - As we continue to increase the number of layers, there is a substantial increase in the time required to train the model.
 - However, despite the increase in layers, the accuracy does not show a significant improvement.
- **Effect of Bidirectionality:**
 - Enabling the bidirectional parameter leads to a substantial increase in accuracy.

- **Effect of Number of Epochs:**

- Despite increasing the number of epochs, the accuracy did not improve significantly due to bidirectionality.

- **Activation Function:**

- In LSTM, the built-in Tanh function is utilized, and at the output layer, the Softmax function has been applied.