AMS 691.02: Natural Language Processing – Fall 2024
Assignment 2

REPORT

1) Neural Networks for Part-of-Speech Tagging

1.1) A Baseline Neural Network Tagger

I have trained a feed-forward neural network classifier to predict the POS tag of a word in its context. The input taken was the word embedding for the center word concatenated with the word embeddings for words in a context window. I have defined a context window as the sequence of words containing w words to either side of the center word and including the center word itself, so the context window contains 1 + 2w words in total. For example, if w = 1 and the word embedding dimensionality is d, the total dimensionality of the input will be 3d. For words near the sentence boundaries, I have padded the sentence with beginning-of-sentence and end-of-sentence characters (<s> and </s>). The word embeddings were randomly initialized and learned along with all other parameters in the model.

Functional Architecture: The input used is the concatenation of word embeddings in the context window, with the word to be tagged in the center. I have used a single hidden layer of width 128 with a tanh nonlinearity. Then the hidden layer was fed to an affine transformation which produced scores for all possible POS tags.

Learning: I have used SGD optimizer.

Initialization: I have randomly initialized all parameters, including word embeddings, and trained them. I have used an initialization range of -0.01 to 0.01 for all word embedding parameters. I have trained on TRAIN data and performed early stopping and preliminary testing on DEV. Final evaluation is done on DEVTEST data.

Result:

For w = 0:-

```
Epoch 1/10, Loss: 26006.6902
Accuracy: 0.7094
Epoch 2/10, Loss: 12139.1604
Accuracy: 0.7693
Epoch 3/10, Loss: 7735.5150
Accuracy: 0.7725
Epoch 4/10, Loss: 6091.0452
Accuracy: 0.7737
Epoch 5/10, Loss: 5323.4264
Accuracy: 0.7743
Epoch 6/10, Loss: 4886.4637
Accuracy: 0.7729
Epoch 7/10, Loss: 4596.4812
Accuracy: 0.7727
Epoch 8/10, Loss: 4382.1922
Accuracy: 0.7727
Epoch 9/10, Loss: 4208.7536
Accuracy: 0.7731
Epoch 10/10, Loss: 4059.0412
Accuracy: 0.7745
Evaluating on DEVTEST set...
Accuracy: 0.7894
```

For w = 1:-

```
Epoch 1/10, Loss: 26248.9624
Accuracy: 0.7264
Epoch 2/10, Loss: 10571.9282
Accuracy: 0.7930
Epoch 3/10, Loss: 5715.2257
Accuracy: 0.7953
Epoch 4/10, Loss: 3615.7730
Accuracy: 0.7955
Epoch 5/10, Loss: 2390.1751
Accuracy: 0.7944
Epoch 6/10, Loss: 1653.6959
Accuracy: 0.7994
Epoch 7/10, Loss: 1250.9881
Accuracy: 0.8007
Epoch 8/10, Loss: 946.4691
Accuracy: 0.8017
Epoch 9/10, Loss: 769.2347
Accuracy: 0.7957
Epoch 10/10, Loss: 602.4454
Accuracy: 0.7982
Evaluating on DEVTEST set...
Accuracy: 0.8176
```

With w = 0, I saw the best DEV accuracy of 77.45% and with w = 1 it improved to 80.17%. I set the size (dimensionality) of word embeddings to 50 and used SGD with a fixed step size of 0.02 and each mini-batch contained one word to be tagged. I trained for 10 epochs and evaluated on DEV once per epoch.

1.2) Feature Engineering

I have added features to the model by concatenating my own feature function outputs to the word embedding concatenation used above. I have defined feature functions based on looking at the training data. For example, I have added binary features if the center word contains certain special characters or capitalization patterns, a feature that returns the number of characters in the center word, features for particular prefixes, suffixes, and other character patterns in the center word, etc. Also implemented these features for context words.

Result:

For w = 0:-

```
Epoch 1/10, Loss: 22638.2898
Accuracy: 0.7185
Epoch 2/10, Loss: 11121.6414
Accuracy: 0.7741
Epoch 3/10, Loss: 7399.7439
Accuracy: 0.7770
Epoch 4/10, Loss: 5998.6403
Accuracy: 0.7795
Epoch 5/10, Loss: 5255.0829
Accuracy: 0.7793
Epoch 6/10, Loss: 4798.8504
Accuracy: 0.7785
Epoch 7/10, Loss: 4493.9533
Accuracy: 0.7785
Epoch 8/10, Loss: 4269.9913
Accuracy: 0.7791
Epoch 9/10, Loss: 4094.4390
Accuracy: 0.7787
Epoch 10/10, Loss: 3947.8282
Accuracy: 0.7787
Evaluating on DEVTEST set...
Accuracy: 0.7948
```

For w = 1:-

```
Epoch 1/10, Loss: 21479.1382
Accuracy: 0.7434
Epoch 2/10, Loss: 9435.1631
Accuracy: 0.8048
Epoch 3/10, Loss: 5385.6582
Accuracy: 0.8079
Epoch 4/10, Loss: 3513.5458
Accuracy: 0.8119
Epoch 5/10, Loss: 2397.7036
Accuracy: 0.8191
Epoch 6/10, Loss: 1657.3680
Accuracy: 0.8189
Epoch 7/10, Loss: 1181.6974
Accuracy: 0.8177
Epoch 8/10, Loss: 904.4941
Accuracy: 0.8166
Epoch 9/10, Loss: 724.6088
Accuracy: 0.8171
Epoch 10/10, Loss: 609.8508
Accuracy: 0.8154
Evaluating on DEVTEST set...
Accuracy: 0.8263
```

Here after adding features, there is not much improvement in the accuracy for w = 0. But there is an improvement for w = 1 from 80.17 to 81.91.

1.3) Pretrained Embeddings

I have initialized word embeddings using the pre-trained embeddings from twitter-embeddings.txt. For words in the tagging datasets that are not in the pre-trained embeddings, I have used the unknown word embedding (i.e., the embedding for the word “UUUNKKK”).

• I have experimented with updating (fine-tuning) the pre-trained embeddings for both w = 0 and w = 1.

Result:

For w = 0:-

```
Epoch 1/10, Loss: 14436.3236
Accuracy: 0.8102
Epoch 2/10, Loss: 10423.5516
Accuracy: 0.8168
Epoch 3/10, Loss: 10012.9584
Accuracy: 0.8177
Epoch 4/10, Loss: 9749.0976
Accuracy: 0.8195
Epoch 5/10, Loss: 9518.9686
Accuracy: 0.8200
Epoch 6/10, Loss: 9292.5743
Accuracy: 0.8245
Epoch 7/10, Loss: 9071.1312
Accuracy: 0.8247
Epoch 8/10, Loss: 8866.4081
Accuracy: 0.8224
Epoch 9/10, Loss: 8685.2950
Accuracy: 0.8222
Epoch 10/10, Loss: 8527.8389
Accuracy: 0.8245
Evaluating on DEVTEST set...
Accuracy: 0.8230
```

For w = 1:-

```
Epoch 1/10, Loss: 13674.0126
Accuracy: 0.8243
Epoch 2/10, Loss: 8810.1603
Accuracy: 0.8345
Epoch 3/10, Loss: 8107.2609
Accuracy: 0.8351
Epoch 4/10, Loss: 7649.6606
Accuracy: 0.8355
Epoch 5/10, Loss: 7267.2322
Accuracy: 0.8365
Epoch 6/10, Loss: 6925.3962
Accuracy: 0.8403
Epoch 7/10, Loss: 6609.0819
Accuracy: 0.8428
Epoch 8/10, Loss: 6309.7090
Accuracy: 0.8471
Epoch 9/10, Loss: 6020.9735
Accuracy: 0.8488
Epoch 10/10, Loss: 5737.4976
Accuracy: 0.8490
Evaluating on DEVTEST set...
Accuracy: 0.8534
```

Here, we can see that for w = 0, the accuracy has improved from 77.45% to 82.47%, and for w = 1, the accuracy has improved from 80.17% to 84.90%.

• With w = 1, I empirically compared updating the pre-trained word embeddings during training and keeping them fixed.

Result:

```
Training with updating embeddings:
Epoch 1/10, Loss: 13640.3876
Accuracy: 0.8251
Epoch 2/10, Loss: 8768.6606
Accuracy: 0.8330
Epoch 3/10, Loss: 8081.7865
Accuracy: 0.8359
Epoch 4/10, Loss: 7640.4794
Accuracy: 0.8376
Epoch 5/10, Loss: 7272.5088
Accuracy: 0.8365
Epoch 6/10, Loss: 6929.9067
Accuracy: 0.8390
Epoch 7/10, Loss: 6594.4760
Accuracy: 0.8411
Epoch 8/10, Loss: 6262.6499
Accuracy: 0.8434
Epoch 9/10, Loss: 5936.6733
Accuracy: 0.8477
Epoch 10/10, Loss: 5619.5313
Accuracy: 0.8494
Evaluating on DEVTEST set...
Accuracy: 0.8467
```

```
Training with fixed embeddings:
Epoch 1/10, Loss: 13596.9028
Accuracy: 0.8251
Epoch 2/10, Loss: 8778.0269
Accuracy: 0.8324
Epoch 3/10, Loss: 8111.7985
Accuracy: 0.8343
Epoch 4/10, Loss: 7687.1712
Accuracy: 0.8343
Epoch 5/10, Loss: 7327.7407
Accuracy: 0.8357
Epoch 6/10, Loss: 6988.6198
Accuracy: 0.8365
Epoch 7/10, Loss: 6654.0582
Accuracy: 0.8401
Epoch 8/10, Loss: 6325.0811
Accuracy: 0.8446
Epoch 9/10, Loss: 6003.6220
Accuracy: 0.8471
Epoch 10/10, Loss: 5690.7811
Accuracy: 0.8463
Evaluating on DEVTEST set...
Accuracy: 0.8506
```

• I have combined the features from Section 1.2 with the use of pre-trained embeddings.

Result:

For w = 0:-

```
Epoch 1/10, Loss: 13832.6978
Accuracy: 0.8197
Epoch 2/10, Loss: 9782.1717
Accuracy: 0.8227
Epoch 3/10, Loss: 9373.8821
Accuracy: 0.8243
Epoch 4/10, Loss: 9122.7273
Accuracy: 0.8256
Epoch 5/10, Loss: 8911.2973
Accuracy: 0.8278
Epoch 6/10, Loss: 8711.0557
Accuracy: 0.8270
Epoch 7/10, Loss: 8516.9677
Accuracy: 0.8283
Epoch 8/10, Loss: 8332.3968
Accuracy: 0.8305
Epoch 9/10, Loss: 8162.2448
Accuracy: 0.8314
Epoch 10/10, Loss: 8008.4303
Accuracy: 0.8349
Evaluating on DEVTEST set...
Accuracy: 0.8319
```

For w = 1:-

```
Epoch 1/10, Loss: 12976.2816
Accuracy: 0.8293
Epoch 2/10, Loss: 8230.7429
Accuracy: 0.8405
Epoch 3/10, Loss: 7572.0660
Accuracy: 0.8426
Epoch 4/10, Loss: 7149.6051
Accuracy: 0.8444
Epoch 5/10, Loss: 6799.0688
Accuracy: 0.8461
Epoch 6/10, Loss: 6477.5862
Accuracy: 0.8475
Epoch 7/10, Loss: 6168.9701
Accuracy: 0.8490
Epoch 8/10, Loss: 5866.2953
Accuracy: 0.8525
Epoch 9/10, Loss: 5566.7842
Accuracy: 0.8544
Epoch 10/10, Loss: 5270.8547
Accuracy: 0.8583
Evaluating on DEVTEST set...
Accuracy: 0.8629
```

Here adding features has increased the accuracy for both w = 0 and w = 1. For w = 0, the accuracy has improved from 82.47% to 83.49%, and for w = 1, the accuracy has improved from 84.90% to 85.83%.

1.4) Architecture Engineering

Here, I have explored the space of neural architectures to see if I can improve the tagger further. Some implementations that I have done are below:

• I have compared the use of 0, 1, and 2 hidden layers. For each number of hidden layers, I have tried two different layer widths that differ by a factor of 2 (e.g., 256 and 512).

• By Keeping the number of layers and layer sizes fixed, I have experimented with different nonlinearities, e.g., identity (g(a) = a), tanh, ReLU, and logistic sigmoid.

• Also experimented with w = 2 and compare the results to w = 0 and 1.

```
Running experiment with window_size=0, hidden_layers=0, layer_widths=[128], activation=tanh
Epoch 1/10, Loss: 20067.9755
Accuracy: 0.7764
Epoch 2/10, Loss: 12808.6664
Accuracy: 0.8075
Epoch 3/10, Loss: 11624.6174
Accuracy: 0.8129
Epoch 4/10, Loss: 11073.4463
Accuracy: 0.8175
Epoch 5/10, Loss: 10740.1414
Accuracy: 0.8181
Epoch 6/10, Loss: 10511.0815
Accuracy: 0.8214
Epoch 7/10, Loss: 10341.3145
Accuracy: 0.8210
Epoch 8/10, Loss: 10209.0107
Accuracy: 0.8214
Epoch 9/10, Loss: 10102.1276
Accuracy: 0.8212
Epoch 10/10, Loss: 10013.4223
Accuracy: 0.8239
```

```
Running experiment with window_size=0, hidden_layers=0, layer_widths=[128], activation=relu
Epoch 1/10, Loss: 20037.1366
Accuracy: 0.7766
Epoch 2/10, Loss: 12805.4272
Accuracy: 0.8073
Epoch 3/10, Loss: 11622.7861
Accuracy: 0.8127
Epoch 4/10, Loss: 11071.6904
Accuracy: 0.8179
Epoch 5/10, Loss: 10738.4446
Accuracy: 0.8179
Epoch 6/10, Loss: 10509.4469
Accuracy: 0.8206
Epoch 7/10, Loss: 10339.7156
Accuracy: 0.8208
Epoch 8/10, Loss: 10207.4179
Accuracy: 0.8214
Epoch 9/10, Loss: 10100.5181
Accuracy: 0.8214
Epoch 10/10, Loss: 10011.7867
Accuracy: 0.8243
```

```
Running experiment with window_size=0, hidden_layers=0, layer_widths=[128], activation=sigmoid
Epoch 1/10, Loss: 19985.5570
Accuracy: 0.7768
Epoch 2/10, Loss: 12802.8071
Accuracy: 0.8077
Epoch 3/10, Loss: 11622.1811
Accuracy: 0.8135
Epoch 4/10, Loss: 11072.5509
Accuracy: 0.8168
Epoch 5/10, Loss: 10739.9986
Accuracy: 0.8181
Epoch 6/10, Loss: 10511.3348
Accuracy: 0.8206
Epoch 7/10, Loss: 10341.7706
Accuracy: 0.8214
Epoch 8/10, Loss: 10209.5696
Accuracy: 0.8216
Epoch 9/10, Loss: 10102.7365
Accuracy: 0.8210
Epoch 10/10, Loss: 10014.0555
Accuracy: 0.8237
```

```
Running experiment with window_size=0, hidden_layers=0, layer_widths=[128], activation=identity
Epoch 1/10, Loss: 20036.1036
Accuracy: 0.7776
Epoch 2/10, Loss: 12815.9507
Accuracy: 0.8065
Epoch 3/10, Loss: 11627.1143
Accuracy: 0.8131
Epoch 4/10, Loss: 11074.2094
Accuracy: 0.8175
Epoch 5/10, Loss: 10739.8966
Accuracy: 0.8177
Epoch 6/10, Loss: 10510.2428
Accuracy: 0.8210
Epoch 7/10, Loss: 10340.1209
Accuracy: 0.8210
Epoch 8/10, Loss: 10207.6049
Accuracy: 0.8212
Epoch 9/10, Loss: 10100.5921
Accuracy: 0.8212
Epoch 10/10, Loss: 10011.8111
Accuracy: 0.8239
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[128], activation=tanh
Epoch 1/10, Loss: 14760.3487
Accuracy: 0.7953
Epoch 2/10, Loss: 10553.2942
Accuracy: 0.8042
Epoch 3/10, Loss: 10089.5054
Accuracy: 0.8063
Epoch 4/10, Loss: 9798.4148
Accuracy: 0.8081
Epoch 5/10, Loss: 9553.8669
Accuracy: 0.8100
Epoch 6/10, Loss: 9330.6323
Accuracy: 0.8133
Epoch 7/10, Loss: 9126.7138
Accuracy: 0.8177
Epoch 8/10, Loss: 8941.4685
Accuracy: 0.8191
Epoch 9/10, Loss: 8773.2296
Accuracy: 0.8187
Epoch 10/10, Loss: 8621.0925
Accuracy: 0.8187
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[128], activation=relu
Epoch 1/10, Loss: 15074.6760
Accuracy: 0.8102
Epoch 2/10, Loss: 9787.2184
Accuracy: 0.8237
Epoch 3/10, Loss: 8993.5459
Accuracy: 0.8301
Epoch 4/10, Loss: 8483.5991
Accuracy: 0.8357
Epoch 5/10, Loss: 8137.2013
Accuracy: 0.8401
Epoch 6/10, Loss: 7878.3842
Accuracy: 0.8405
Epoch 7/10, Loss: 7666.2922
Accuracy: 0.8397
Epoch 8/10, Loss: 7488.3234
Accuracy: 0.8409
Epoch 9/10, Loss: 7326.8300
Accuracy: 0.8407
Epoch 10/10, Loss: 7183.5146
Accuracy: 0.8411
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[128], activation=sigmoid
Epoch 1/10, Loss: 26794.3498
Accuracy: 0.7295
Epoch 2/10, Loss: 13676.9611
Accuracy: 0.7575
Epoch 3/10, Loss: 11778.9328
Accuracy: 0.7886
Epoch 4/10, Loss: 10979.8898
Accuracy: 0.8112
Epoch 5/10, Loss: 10546.8859
Accuracy: 0.8168
Epoch 6/10, Loss: 10263.6371
Accuracy: 0.8179
Epoch 7/10, Loss: 10052.9431
Accuracy: 0.8187
Epoch 8/10, Loss: 9882.9833
Accuracy: 0.8193
Epoch 9/10, Loss: 9737.9635
Accuracy: 0.8210
Epoch 10/10, Loss: 9609.1610
Accuracy: 0.8222
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[128], activation=identity
Epoch 1/10, Loss: 14712.9020
Accuracy: 0.7949
Epoch 2/10, Loss: 10863.7214
Accuracy: 0.8013
Epoch 3/10, Loss: 10519.1170
Accuracy: 0.8042
Epoch 4/10, Loss: 10343.6080
Accuracy: 0.8034
Epoch 5/10, Loss: 10234.0902
Accuracy: 0.8032
Epoch 6/10, Loss: 10162.0497
Accuracy: 0.8019
Epoch 7/10, Loss: 10108.1409
Accuracy: 0.8023
Epoch 8/10, Loss: 10070.8568
Accuracy: 0.8021
Epoch 9/10, Loss: 10038.7107
Accuracy: 0.8015
Epoch 10/10, Loss: 10014.9459
Accuracy: 0.8019
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[256], activation=tanh
Epoch 1/10, Loss: 14162.9025
Accuracy: 0.7986
Epoch 2/10, Loss: 10680.1302
Accuracy: 0.8034
Epoch 3/10, Loss: 10280.7793
Accuracy: 0.8050
Epoch 4/10, Loss: 10045.7103
Accuracy: 0.8050
Epoch 5/10, Loss: 9874.7612
Accuracy: 0.8056
Epoch 6/10, Loss: 9732.2265
Accuracy: 0.8063
Epoch 7/10, Loss: 9599.0864
Accuracy: 0.8067
Epoch 8/10, Loss: 9465.7000
Accuracy: 0.8061
Epoch 9/10, Loss: 9328.0706
Accuracy: 0.8071
Epoch 10/10, Loss: 9186.4084
Accuracy: 0.8077
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[256], activation=relu
Epoch 1/10, Loss: 14416.3691
Accuracy: 0.8156
Epoch 2/10, Loss: 9606.9205
Accuracy: 0.8297
Epoch 3/10, Loss: 8806.7875
Accuracy: 0.8353
Epoch 4/10, Loss: 8310.9747
Accuracy: 0.8413
Epoch 5/10, Loss: 7941.6206
Accuracy: 0.8413
Epoch 6/10, Loss: 7654.7220
Accuracy: 0.8421
Epoch 7/10, Loss: 7431.3309
Accuracy: 0.8419
Epoch 8/10, Loss: 7240.5955
Accuracy: 0.8413
Epoch 9/10, Loss: 7073.7178
Accuracy: 0.8417
Epoch 10/10, Loss: 6931.8570
Accuracy: 0.8424
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[256], activation=sigmoid
Epoch 1/10, Loss: 26202.1579
Accuracy: 0.7349
Epoch 2/10, Loss: 13682.7214
Accuracy: 0.7689
Epoch 3/10, Loss: 11952.4122
Accuracy: 0.7824
Epoch 4/10, Loss: 11209.0426
Accuracy: 0.8036
Epoch 5/10, Loss: 10792.3519
Accuracy: 0.8125
Epoch 6/10, Loss: 10510.1245
Accuracy: 0.8160
Epoch 7/10, Loss: 10296.4494
Accuracy: 0.8171
Epoch 8/10, Loss: 10123.1856
Accuracy: 0.8173
Epoch 9/10, Loss: 9975.5568
Accuracy: 0.8183
Epoch 10/10, Loss: 9845.0556
Accuracy: 0.8195
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[256], activation=identity
Epoch 1/10, Loss: 14120.0923
Accuracy: 0.7965
Epoch 2/10, Loss: 10815.7520
Accuracy: 0.8015
Epoch 3/10, Loss: 10496.5661
Accuracy: 0.8021
Epoch 4/10, Loss: 10333.7273
Accuracy: 0.8027
Epoch 5/10, Loss: 10230.5314
Accuracy: 0.8021
Epoch 6/10, Loss: 10163.5514
Accuracy: 0.8013
Epoch 7/10, Loss: 10111.6441
Accuracy: 0.8017
Epoch 8/10, Loss: 10074.5138
Accuracy: 0.8019
Epoch 9/10, Loss: 10042.5901
Accuracy: 0.8017
Epoch 10/10, Loss: 10020.1842
Accuracy: 0.8021
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[512], activation=tanh
Epoch 1/10, Loss: 13495.2746
Accuracy: 0.7944
Epoch 2/10, Loss: 10708.4203
Accuracy: 0.8011
Epoch 3/10, Loss: 10388.0098
Accuracy: 0.8021
Epoch 4/10, Loss: 10199.4085
Accuracy: 0.8025
Epoch 5/10, Loss: 10060.3103
Accuracy: 0.8017
Epoch 6/10, Loss: 9946.1572
Accuracy: 0.8019
Epoch 7/10, Loss: 9842.7878
Accuracy: 0.8036
Epoch 8/10, Loss: 9740.6961
Accuracy: 0.8036
Epoch 9/10, Loss: 9637.2745
Accuracy: 0.8044
Epoch 10/10, Loss: 9533.9372
Accuracy: 0.8081
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[512], activation=relu
Epoch 1/10, Loss: 13711.2448
Accuracy: 0.8171
Epoch 2/10, Loss: 9432.5469
Accuracy: 0.8312
Epoch 3/10, Loss: 8628.9601
Accuracy: 0.8368
Epoch 4/10, Loss: 8128.6149
Accuracy: 0.8407
Epoch 5/10, Loss: 7771.6171
Accuracy: 0.8428
Epoch 6/10, Loss: 7493.6364
Accuracy: 0.8432
Epoch 7/10, Loss: 7274.7580
Accuracy: 0.8440
Epoch 8/10, Loss: 7079.1563
Accuracy: 0.8434
Epoch 9/10, Loss: 6912.9441
Accuracy: 0.8432
Epoch 10/10, Loss: 6772.2932
Accuracy: 0.8419
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[512], activation=sigmoid
Epoch 1/10, Loss: 26588.1829
Accuracy: 0.7343
Epoch 2/10, Loss: 14298.8114
Accuracy: 0.7501
Epoch 3/10, Loss: 12591.4170
Accuracy: 0.7791
Epoch 4/10, Loss: 11827.2160
Accuracy: 0.7957
Epoch 5/10, Loss: 11362.7397
Accuracy: 0.7955
Epoch 6/10, Loss: 11030.3917
Accuracy: 0.8034
Epoch 7/10, Loss: 10772.6101
Accuracy: 0.8067
Epoch 8/10, Loss: 10562.4013
Accuracy: 0.8102
Epoch 9/10, Loss: 10384.5656
Accuracy: 0.8117
Epoch 10/10, Loss: 10229.9285
Accuracy: 0.8125
```

```
Running experiment with window_size=0, hidden_layers=1, layer_widths=[512], activation=identity
Epoch 1/10, Loss: 13456.4171
Accuracy: 0.7942
Epoch 2/10, Loss: 10776.9319
Accuracy: 0.7984
Epoch 3/10, Loss: 10488.1364
Accuracy: 0.8011
Epoch 4/10, Loss: 10336.5897
Accuracy: 0.8025
Epoch 5/10, Loss: 10240.6720
Accuracy: 0.8021
Epoch 6/10, Loss: 10174.2137
Accuracy: 0.8015
Epoch 7/10, Loss: 10124.7467
Accuracy: 0.8019
Epoch 8/10, Loss: 10087.4569
Accuracy: 0.8019
Epoch 9/10, Loss: 10057.1184
Accuracy: 0.8023
Epoch 10/10, Loss: 10034.6585
Accuracy: 0.8029
```

```
Running experiment with window_size=0, hidden_layers=2, layer_widths=[128, 256], activation=tanh
Epoch 1/10, Loss: 14360.2499
Accuracy: 0.7874
Epoch 2/10, Loss: 10888.3708
Accuracy: 0.7955
Epoch 3/10, Loss: 10222.5964
Accuracy: 0.8023
Epoch 4/10, Loss: 9705.0763
Accuracy: 0.8050
Epoch 5/10, Loss: 9305.7656
Accuracy: 0.8083
Epoch 6/10, Loss: 8991.1883
Accuracy: 0.8112
Epoch 7/10, Loss: 8724.7808
Accuracy: 0.8141
Epoch 8/10, Loss: 8492.1851
Accuracy: 0.8166
Epoch 9/10, Loss: 8286.8527
Accuracy: 0.8162
Epoch 10/10, Loss: 8106.6331
Accuracy: 0.8181
```

```
Running experiment with window_size=0, hidden_layers=2, layer_widths=[128, 256], activation=relu
Epoch 1/10, Loss: 14708.4204
Accuracy: 0.8112
Epoch 2/10, Loss: 9415.5522
Accuracy: 0.8278
Epoch 3/10, Loss: 8570.6643
Accuracy: 0.8374
Epoch 4/10, Loss: 8046.3583
Accuracy: 0.8392
Epoch 5/10, Loss: 7673.2374
Accuracy: 0.8413
Epoch 6/10, Loss: 7365.7058
Accuracy: 0.8430
Epoch 7/10, Loss: 7132.2032
Accuracy: 0.8409
Epoch 8/10, Loss: 6919.0703
Accuracy: 0.8415
Epoch 9/10, Loss: 6743.8438
Accuracy: 0.8399
Epoch 10/10, Loss: 6592.0752
Accuracy: 0.8411
```

```
Running experiment with window_size=0, hidden_layers=2, layer_widths=[128, 256], activation=sigmoid
Epoch 1/10, Loss: 45224.7552
Accuracy: 0.3669
Epoch 2/10, Loss: 26815.7304
Accuracy: 0.6646
Epoch 3/10, Loss: 17245.0678
Accuracy: 0.7138
Epoch 4/10, Loss: 14239.8540
Accuracy: 0.7364
Epoch 5/10, Loss: 13073.9333
Accuracy: 0.7517
Epoch 6/10, Loss: 12392.0044
Accuracy: 0.7660
Epoch 7/10, Loss: 11850.9704
Accuracy: 0.7872
Epoch 8/10, Loss: 11389.6475
Accuracy: 0.7978
Epoch 9/10, Loss: 10999.7712
Accuracy: 0.8038
Epoch 10/10, Loss: 10673.3082
Accuracy: 0.8063
```

```
Running experiment with window_size=0, hidden_layers=2, layer_widths=[128, 256], activation=identity
Epoch 1/10, Loss: 14718.0822
Accuracy: 0.7828
Epoch 2/10, Loss: 11652.2461
Accuracy: 0.7843
Epoch 3/10, Loss: 11373.7540
Accuracy: 0.7861
Epoch 4/10, Loss: 11249.9211
Accuracy: 0.7847
Epoch 5/10, Loss: 11190.3019
Accuracy: 0.7841
Epoch 6/10, Loss: 11139.4838
Accuracy: 0.7843
Epoch 7/10, Loss: 11129.7797
Accuracy: 0.7837
Epoch 8/10, Loss: 11098.1950
Accuracy: 0.7837
Epoch 9/10, Loss: 11099.8937
Accuracy: 0.7837
Epoch 10/10, Loss: 11068.3276
Accuracy: 0.7828
```

```
Running experiment with window_size=0, hidden_layers=2, layer_widths=[256, 512], activation=tanh
Epoch 1/10, Loss: 14105.6129
Accuracy: 0.7855
Epoch 2/10, Loss: 11189.6507
Accuracy: 0.7909
Epoch 3/10, Loss: 10562.7267
Accuracy: 0.7965
Epoch 4/10, Loss: 10067.6246
Accuracy: 0.8034
Epoch 5/10, Loss: 9643.0088
Accuracy: 0.8071
Epoch 6/10, Loss: 9285.6889
Accuracy: 0.8106
Epoch 7/10, Loss: 8993.5242
Accuracy: 0.8131
Epoch 8/10, Loss: 8755.2646
Accuracy: 0.8177
Epoch 9/10, Loss: 8545.8409
Accuracy: 0.8175
Epoch 10/10, Loss: 8373.4845
Accuracy: 0.8179
```

```
Running experiment with window_size=0, hidden_layers=2, layer_widths=[256, 512], activation=relu
Epoch 1/10, Loss: 14073.6959
Accuracy: 0.8075
Epoch 2/10, Loss: 9200.4445
Accuracy: 0.8241
Epoch 3/10, Loss: 8340.4805
Accuracy: 0.8345
Epoch 4/10, Loss: 7806.3354
Accuracy: 0.8334
Epoch 5/10, Loss: 7422.5117
Accuracy: 0.8345
Epoch 6/10, Loss: 7120.9444
Accuracy: 0.8332
Epoch 7/10, Loss: 6871.1489
Accuracy: 0.8353
Epoch 8/10, Loss: 6643.4667
Accuracy: 0.8339
Epoch 9/10, Loss: 6443.7178
Accuracy: 0.8349
Epoch 10/10, Loss: 6283.5889
Accuracy: 0.8332
```

```
Running experiment with window_size=0, hidden_layers=2, layer_widths=[256, 512], activation=sigmoid
Epoch 1/10, Loss: 45470.8868
Accuracy: 0.3470
Epoch 2/10, Loss: 27847.4358
Accuracy: 0.6333
Epoch 3/10, Loss: 18797.5065
Accuracy: 0.7104
Epoch 4/10, Loss: 14973.2392
Accuracy: 0.7349
Epoch 5/10, Loss: 13401.6337
Accuracy: 0.7467
Epoch 6/10, Loss: 12572.6968
Accuracy: 0.7639
Epoch 7/10, Loss: 11969.3072
Accuracy: 0.7749
Epoch 8/10, Loss: 11470.9553
Accuracy: 0.7934
Epoch 9/10, Loss: 11059.1441
Accuracy: 0.8005
Epoch 10/10, Loss: 10724.8817
Accuracy: 0.8044
```

```
Running experiment with window_size=0, hidden_layers=2, layer_widths=[256, 512], activation=identity
Epoch 1/10, Loss: 14187.6137
Accuracy: 0.7818
Epoch 2/10, Loss: 11662.8133
Accuracy: 0.7853
Epoch 3/10, Loss: 11398.7683
Accuracy: 0.7857
Epoch 4/10, Loss: 11274.6044
Accuracy: 0.7847
Epoch 5/10, Loss: 11215.9254
Accuracy: 0.7843
Epoch 6/10, Loss: 11161.4008
Accuracy: 0.7847
Epoch 7/10, Loss: 11151.6494
Accuracy: 0.7834
Epoch 8/10, Loss: 11114.4849
Accuracy: 0.7832
Epoch 9/10, Loss: 11114.8357
Accuracy: 0.7834
Epoch 10/10, Loss: 11081.4846
Accuracy: 0.7828
```

```
Running experiment with window_size=1, hidden_layers=0, layer_widths=[128], activation=tanh
Epoch 1/10, Loss: 17857.8405
Accuracy: 0.8085
Epoch 2/10, Loss: 10872.8789
Accuracy: 0.8326
Epoch 3/10, Loss: 9638.0909
Accuracy: 0.8407
Epoch 4/10, Loss: 9016.1730
Accuracy: 0.8436
Epoch 5/10, Loss: 8618.5681
Accuracy: 0.8457
Epoch 6/10, Loss: 8334.3881
Accuracy: 0.8473
Epoch 7/10, Loss: 8117.5272
Accuracy: 0.8496
Epoch 8/10, Loss: 7944.6575
Accuracy: 0.8496
Epoch 9/10, Loss: 7802.4872
Accuracy: 0.8496
Epoch 10/10, Loss: 7682.7970
Accuracy: 0.8507
```

```
Running experiment with window_size=1, hidden_layers=0, layer_widths=[128], activation=relu
Epoch 1/10, Loss: 17855.4740
Accuracy: 0.8081
Epoch 2/10, Loss: 10879.9402
Accuracy: 0.8328
Epoch 3/10, Loss: 9641.9486
Accuracy: 0.8401
Epoch 4/10, Loss: 9017.9847
Accuracy: 0.8446
Epoch 5/10, Loss: 8619.2248
Accuracy: 0.8455
Epoch 6/10, Loss: 8334.3773
Accuracy: 0.8484
Epoch 7/10, Loss: 8117.1206
Accuracy: 0.8500
Epoch 8/10, Loss: 7944.0133
Accuracy: 0.8500
Epoch 9/10, Loss: 7801.7020
Accuracy: 0.8500
Epoch 10/10, Loss: 7681.9328
Accuracy: 0.8511
```

```
Running experiment with window_size=1, hidden_layers=0, layer_widths=[128], activation=sigmoid
Epoch 1/10, Loss: 17870.1719
Accuracy: 0.8083
Epoch 2/10, Loss: 10880.4101
Accuracy: 0.8332
Epoch 3/10, Loss: 9643.4880
Accuracy: 0.8407
Epoch 4/10, Loss: 9020.0776
Accuracy: 0.8434
Epoch 5/10, Loss: 8621.4350
Accuracy: 0.8448
Epoch 6/10, Loss: 8336.4704
Accuracy: 0.8477
Epoch 7/10, Loss: 8119.0175
Accuracy: 0.8492
Epoch 8/10, Loss: 7945.7057
Accuracy: 0.8490
Epoch 9/10, Loss: 7803.2061
Accuracy: 0.8498
Epoch 10/10, Loss: 7683.2703
Accuracy: 0.8511
```

```
Running experiment with window_size=1, hidden_layers=0, layer_widths=[128], activation=identity
Epoch 1/10, Loss: 17883.0425
Accuracy: 0.8088
Epoch 2/10, Loss: 10884.1520
Accuracy: 0.8324
Epoch 3/10, Loss: 9643.6940
Accuracy: 0.8409
Epoch 4/10, Loss: 9019.3398
Accuracy: 0.8434
Epoch 5/10, Loss: 8620.4785
Accuracy: 0.8455
Epoch 6/10, Loss: 8335.5554
Accuracy: 0.8475
Epoch 7/10, Loss: 8118.2456
Accuracy: 0.8496
Epoch 8/10, Loss: 7945.1073
Accuracy: 0.8494
Epoch 9/10, Loss: 7802.7800
Accuracy: 0.8502
Epoch 10/10, Loss: 7683.0016
Accuracy: 0.8511
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[128], activation=tanh
Epoch 1/10, Loss: 14061.1612
Accuracy: 0.8125
Epoch 2/10, Loss: 9116.3252
Accuracy: 0.8224
Epoch 3/10, Loss: 8357.5334
Accuracy: 0.8224
Epoch 4/10, Loss: 7860.5119
Accuracy: 0.8235
Epoch 5/10, Loss: 7458.1087
Accuracy: 0.8303
Epoch 6/10, Loss: 7099.6513
Accuracy: 0.8330
Epoch 7/10, Loss: 6763.5037
Accuracy: 0.8357
Epoch 8/10, Loss: 6438.9511
Accuracy: 0.8372
Epoch 9/10, Loss: 6123.4620
Accuracy: 0.8380
Epoch 10/10, Loss: 5817.6344
Accuracy: 0.8368
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[128], activation=relu
Epoch 1/10, Loss: 14710.1739
Accuracy: 0.8075
Epoch 2/10, Loss: 8644.3212
Accuracy: 0.8276
Epoch 3/10, Loss: 7480.9402
Accuracy: 0.8380
Epoch 4/10, Loss: 6668.0831
Accuracy: 0.8426
Epoch 5/10, Loss: 5986.4658
Accuracy: 0.8453
Epoch 6/10, Loss: 5412.6386
Accuracy: 0.8467
Epoch 7/10, Loss: 4927.2956
Accuracy: 0.8463
Epoch 8/10, Loss: 4477.5952
Accuracy: 0.8513
Epoch 9/10, Loss: 4056.6335
Accuracy: 0.8498
Epoch 10/10, Loss: 3701.1903
Accuracy: 0.8504
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[128], activation=sigmoid
Epoch 1/10, Loss: 26517.5311
Accuracy: 0.7438
Epoch 2/10, Loss: 12410.1221
Accuracy: 0.7774
Epoch 3/10, Loss: 10265.6484
Accuracy: 0.8021
Epoch 4/10, Loss: 9268.8186
Accuracy: 0.8187
Epoch 5/10, Loss: 8691.5974
Accuracy: 0.8272
Epoch 6/10, Loss: 8302.6752
Accuracy: 0.8328
Epoch 7/10, Loss: 8007.7093
Accuracy: 0.8368
Epoch 8/10, Loss: 7765.3736
Accuracy: 0.8374
Epoch 9/10, Loss: 7555.0935
Accuracy: 0.8403
Epoch 10/10, Loss: 7365.4906
Accuracy: 0.8434
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[128], activation=identity
Epoch 1/10, Loss: 14107.9069
Accuracy: 0.8083
Epoch 2/10, Loss: 9435.7258
Accuracy: 0.8175
Epoch 3/10, Loss: 8851.1843
Accuracy: 0.8216
Epoch 4/10, Loss: 8546.6422
Accuracy: 0.8220
Epoch 5/10, Loss: 8355.5599
Accuracy: 0.8216
Epoch 6/10, Loss: 8225.1545
Accuracy: 0.8212
Epoch 7/10, Loss: 8128.1661
Accuracy: 0.8208
Epoch 8/10, Loss: 8058.8803
Accuracy: 0.8210
Epoch 9/10, Loss: 7997.9809
Accuracy: 0.8216
Epoch 10/10, Loss: 7955.1368
Accuracy: 0.8210
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[128], activation=sigmoid
Epoch 1/10, Loss: 26517.5311
Accuracy: 0.7438
Epoch 2/10, Loss: 12410.1221
Accuracy: 0.7774
Epoch 3/10, Loss: 10265.6484
Accuracy: 0.8021
Epoch 4/10, Loss: 9268.8186
Accuracy: 0.8187
Epoch 5/10, Loss: 8691.5974
Accuracy: 0.8272
Epoch 6/10, Loss: 8302.6752
Accuracy: 0.8328
Epoch 7/10, Loss: 8007.7093
Accuracy: 0.8368
Epoch 8/10, Loss: 7765.3736
Accuracy: 0.8374
Epoch 9/10, Loss: 7555.0935
Accuracy: 0.8403
Epoch 10/10, Loss: 7365.4906
Accuracy: 0.8434
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[128], activation=identity
Epoch 1/10, Loss: 14107.9069
Accuracy: 0.8083
Epoch 2/10, Loss: 9435.7258
Accuracy: 0.8175
Epoch 3/10, Loss: 8851.1843
Accuracy: 0.8216
Epoch 4/10, Loss: 8546.6422
Accuracy: 0.8220
Epoch 5/10, Loss: 8355.5599
Accuracy: 0.8216
Epoch 6/10, Loss: 8225.1545
Accuracy: 0.8212
Epoch 7/10, Loss: 8128.1661
Accuracy: 0.8208
Epoch 8/10, Loss: 8058.8803
Accuracy: 0.8210
Epoch 9/10, Loss: 7997.9809
Accuracy: 0.8216
Epoch 10/10, Loss: 7955.1368
Accuracy: 0.8210
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[256], activation=tanh
Epoch 1/10, Loss: 13703.8385
Accuracy: 0.8104
Epoch 2/10, Loss: 9237.1044
Accuracy: 0.8171
Epoch 3/10, Loss: 8574.7806
Accuracy: 0.8235
Epoch 4/10, Loss: 8153.3281
Accuracy: 0.8237
Epoch 5/10, Loss: 7829.9323
Accuracy: 0.8280
Epoch 6/10, Loss: 7554.2865
Accuracy: 0.8297
Epoch 7/10, Loss: 7304.7356
Accuracy: 0.8328
Epoch 8/10, Loss: 7066.8619
Accuracy: 0.8353
Epoch 9/10, Loss: 6832.2029
Accuracy: 0.8382
Epoch 10/10, Loss: 6598.0830
Accuracy: 0.8378
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[256], activation=relu
Epoch 1/10, Loss: 14253.3466
Accuracy: 0.8133
Epoch 2/10, Loss: 8449.6352
Accuracy: 0.8330
Epoch 3/10, Loss: 7222.7769
Accuracy: 0.8421
Epoch 4/10, Loss: 6340.5036
Accuracy: 0.8490
Epoch 5/10, Loss: 5618.5405
Accuracy: 0.8540
Epoch 6/10, Loss: 4965.5986
Accuracy: 0.8538
Epoch 7/10, Loss: 4425.7277
Accuracy: 0.8598
Epoch 8/10, Loss: 3927.6409
Accuracy: 0.8590
Epoch 9/10, Loss: 3489.0541
Accuracy: 0.8594
Epoch 10/10, Loss: 3113.8688
Accuracy: 0.8604
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[256], activation=sigmoid
Epoch 1/10, Loss: 26099.0288
Accuracy: 0.7430
Epoch 2/10, Loss: 12394.5454
Accuracy: 0.7714
Epoch 3/10, Loss: 10343.5819
Accuracy: 0.7934
Epoch 4/10, Loss: 9382.3514
Accuracy: 0.8083
Epoch 5/10, Loss: 8821.9363
Accuracy: 0.8195
Epoch 6/10, Loss: 8438.1482
Accuracy: 0.8264
Epoch 7/10, Loss: 8143.5401
Accuracy: 0.8301
Epoch 8/10, Loss: 7900.0866
Accuracy: 0.8328
Epoch 9/10, Loss: 7688.8756
Accuracy: 0.8361
Epoch 10/10, Loss: 7499.4049
Accuracy: 0.8378
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[256], activation=identity
Epoch 1/10, Loss: 13686.5220
Accuracy: 0.8092
Epoch 2/10, Loss: 9384.9836
Accuracy: 0.8166
Epoch 3/10, Loss: 8830.7551
Accuracy: 0.8195
Epoch 4/10, Loss: 8532.7562
Accuracy: 0.8216
Epoch 5/10, Loss: 8346.1998
Accuracy: 0.8212
Epoch 6/10, Loss: 8219.4606
Accuracy: 0.8195
Epoch 7/10, Loss: 8125.1600
Accuracy: 0.8206
Epoch 8/10, Loss: 8057.5284
Accuracy: 0.8212
Epoch 9/10, Loss: 7999.0612
Accuracy: 0.8214
Epoch 10/10, Loss: 7957.9802
Accuracy: 0.8218
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[512], activation=tanh
Epoch 1/10, Loss: 13225.9459
Accuracy: 0.8114
Epoch 2/10, Loss: 9292.8288
Accuracy: 0.8181
Epoch 3/10, Loss: 8730.5761
Accuracy: 0.8204
Epoch 4/10, Loss: 8409.0850
Accuracy: 0.8212
Epoch 5/10, Loss: 8190.0794
Accuracy: 0.8218
Epoch 6/10, Loss: 8017.2937
Accuracy: 0.8216
Epoch 7/10, Loss: 7866.8275
Accuracy: 0.8212
Epoch 8/10, Loss: 7724.5388
Accuracy: 0.8212
Epoch 9/10, Loss: 7583.7507
Accuracy: 0.8214
Epoch 10/10, Loss: 7441.2934
Accuracy: 0.8202
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[512], activation=relu
Epoch 1/10, Loss: 13835.0015
Accuracy: 0.8125
Epoch 2/10, Loss: 8309.9226
Accuracy: 0.8320
Epoch 3/10, Loss: 7014.9464
Accuracy: 0.8421
Epoch 4/10, Loss: 6070.8188
Accuracy: 0.8486
Epoch 5/10, Loss: 5280.0422
Accuracy: 0.8525
Epoch 6/10, Loss: 4567.8292
Accuracy: 0.8525
Epoch 7/10, Loss: 3984.5669
Accuracy: 0.8554
Epoch 8/10, Loss: 3453.7513
Accuracy: 0.8567
Epoch 9/10, Loss: 2998.3193
Accuracy: 0.8556
Epoch 10/10, Loss: 2623.2436
Accuracy: 0.8587
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[512], activation=sigmoid
Epoch 1/10, Loss: 26795.8074
Accuracy: 0.7133
Epoch 2/10, Loss: 12870.5215
Accuracy: 0.7455
Epoch 3/10, Loss: 10751.6913
Accuracy: 0.7671
Epoch 4/10, Loss: 9761.8420
Accuracy: 0.7949
Epoch 5/10, Loss: 9176.4948
Accuracy: 0.8063
Epoch 6/10, Loss: 8769.6882
Accuracy: 0.8164
Epoch 7/10, Loss: 8456.8090
Accuracy: 0.8229
Epoch 8/10, Loss: 8199.7816
Accuracy: 0.8270
Epoch 9/10, Loss: 7978.9271
Accuracy: 0.8301
Epoch 10/10, Loss: 7783.0128
Accuracy: 0.8312
```

```
Running experiment with window_size=1, hidden_layers=1, layer_widths=[512], activation=identity
Epoch 1/10, Loss: 13155.8521
Accuracy: 0.8100
Epoch 2/10, Loss: 9364.4362
Accuracy: 0.8137
Epoch 3/10, Loss: 8843.4481
Accuracy: 0.8193
Epoch 4/10, Loss: 8555.6171
Accuracy: 0.8197
Epoch 5/10, Loss: 8376.1269
Accuracy: 0.8193
Epoch 6/10, Loss: 8251.3780
Accuracy: 0.8204
Epoch 7/10, Loss: 8157.2557
Accuracy: 0.8204
Epoch 8/10, Loss: 8089.9396
Accuracy: 0.8214
Epoch 9/10, Loss: 8030.7892
Accuracy: 0.8204
Epoch 10/10, Loss: 7987.6047
Accuracy: 0.8202
```

```
Running experiment with window_size=1, hidden_layers=2, layer_widths=[128, 256], activation=tanh
Epoch 1/10, Loss: 14445.4242
Accuracy: 0.7675
Epoch 2/10, Loss: 9807.7829
Accuracy: 0.7859
Epoch 3/10, Loss: 8910.1050
Accuracy: 0.8038
Epoch 4/10, Loss: 8241.7325
Accuracy: 0.8100
Epoch 5/10, Loss: 7672.6735
Accuracy: 0.8154
Epoch 6/10, Loss: 7146.5606
Accuracy: 0.8173
Epoch 7/10, Loss: 6654.4095
Accuracy: 0.8183
Epoch 8/10, Loss: 6241.0791
Accuracy: 0.8100
Epoch 9/10, Loss: 5900.8212
Accuracy: 0.8227
Epoch 10/10, Loss: 5496.0826
Accuracy: 0.8247
```

```
Running experiment with window_size=1, hidden_layers=2, layer_widths=[128, 256], activation=relu
Epoch 1/10, Loss: 15528.6578
Accuracy: 0.7778
Epoch 2/10, Loss: 8703.3317
Accuracy: 0.8152
Epoch 3/10, Loss: 7251.0951
Accuracy: 0.8287
Epoch 4/10, Loss: 6233.3714
Accuracy: 0.8382
Epoch 5/10, Loss: 5482.3447
Accuracy: 0.8417
Epoch 6/10, Loss: 4853.5811
Accuracy: 0.8438
Epoch 7/10, Loss: 4308.4535
Accuracy: 0.8448
Epoch 8/10, Loss: 3837.4174
Accuracy: 0.8392
Epoch 9/10, Loss: 3566.9382
Accuracy: 0.8436
Epoch 10/10, Loss: 3229.5748
Accuracy: 0.8397
```

```
Running experiment with window_size=1, hidden_layers=2, layer_widths=[128, 256], activation=sigmoid
Epoch 1/10, Loss: 45345.0556
Accuracy: 0.2377
Epoch 2/10, Loss: 29459.2776
Accuracy: 0.6067
Epoch 3/10, Loss: 17920.3553
Accuracy: 0.7146
Epoch 4/10, Loss: 13757.0168
Accuracy: 0.7426
Epoch 5/10, Loss: 12218.6566
Accuracy: 0.7594
Epoch 6/10, Loss: 11361.6780
Accuracy: 0.7768
Epoch 7/10, Loss: 10663.6859
Accuracy: 0.7893
Epoch 8/10, Loss: 10077.3716
Accuracy: 0.8019
Epoch 9/10, Loss: 9609.7787
Accuracy: 0.8073
Epoch 10/10, Loss: 9229.0535
Accuracy: 0.8135
```

```
Running experiment with window_size=1, hidden_layers=2, layer_widths=[128, 256], activation=identity
Epoch 1/10, Loss: 14584.9486
Accuracy: 0.7639
Epoch 2/10, Loss: 10551.6001
Accuracy: 0.7720
Epoch 3/10, Loss: 10116.7097
Accuracy: 0.7868
Epoch 4/10, Loss: 9944.6880
Accuracy: 0.7888
Epoch 5/10, Loss: 9839.3168
Accuracy: 0.7909
Epoch 6/10, Loss: 9805.7813
Accuracy: 0.7915
Epoch 7/10, Loss: 9764.9362
Accuracy: 0.7895
Epoch 8/10, Loss: 9757.0805
Accuracy: 0.7899
Epoch 9/10, Loss: 9737.3703
Accuracy: 0.7899
Epoch 10/10, Loss: 9745.8836
Accuracy: 0.7907
```

```
Running experiment with window_size=1, hidden_layers=2, layer_widths=[256, 512], activation=tanh
Epoch 1/10, Loss: 14032.5734
Accuracy: 0.7666
Epoch 2/10, Loss: 10122.7892
Accuracy: 0.7803
Epoch 3/10, Loss: 9402.1891
Accuracy: 0.7990
Epoch 4/10, Loss: 8853.0933
Accuracy: 0.8056
Epoch 5/10, Loss: 8347.8717
Accuracy: 0.8112
Epoch 6/10, Loss: 7836.4239
Accuracy: 0.8127
Epoch 7/10, Loss: 7336.4218
Accuracy: 0.8173
Epoch 8/10, Loss: 6968.3626
Accuracy: 0.8108
Epoch 9/10, Loss: 6580.5613
Accuracy: 0.8168
Epoch 10/10, Loss: 6264.7198
Accuracy: 0.8187
```

```
Running experiment with window_size=1, hidden_layers=2, layer_widths=[256, 512], activation=relu
Epoch 1/10, Loss: 14898.2357
Accuracy: 0.7878
Epoch 2/10, Loss: 8374.3464
Accuracy: 0.8146
Epoch 3/10, Loss: 6865.9665
Accuracy: 0.8351
Epoch 4/10, Loss: 5763.4193
Accuracy: 0.8326
Epoch 5/10, Loss: 4881.3112
Accuracy: 0.8469
Epoch 6/10, Loss: 4112.8546
Accuracy: 0.8384
Epoch 7/10, Loss: 3559.8454
Accuracy: 0.8411
Epoch 8/10, Loss: 3026.6451
Accuracy: 0.8527
Epoch 9/10, Loss: 2878.0220
Accuracy: 0.8446
Epoch 10/10, Loss: 2694.7922
Accuracy: 0.8438
```

```
Running experiment with window_size=1, hidden_layers=2, layer_widths=[256, 512], activation=sigmoid
Epoch 1/10, Loss: 45854.7842
Accuracy: 0.2246
Epoch 2/10, Loss: 33599.1540
Accuracy: 0.5447
Epoch 3/10, Loss: 20521.3039
Accuracy: 0.6909
Epoch 4/10, Loss: 14853.3589
Accuracy: 0.7370
Epoch 5/10, Loss: 12717.1456
Accuracy: 0.7567
Epoch 6/10, Loss: 11604.0450
Accuracy: 0.7731
Epoch 7/10, Loss: 10821.2008
Accuracy: 0.7897
Epoch 8/10, Loss: 10270.2049
Accuracy: 0.7965
Epoch 9/10, Loss: 9859.3391
Accuracy: 0.8034
Epoch 10/10, Loss: 9523.5573
Accuracy: 0.8061
```

```
Running experiment with window_size=1, hidden_layers=2, layer_widths=[256, 512], activation=identity
Epoch 1/10, Loss: 14157.1848
Accuracy: 0.7619
Epoch 2/10, Loss: 10571.1264
Accuracy: 0.7695
Epoch 3/10, Loss: 10156.8804
Accuracy: 0.7864
Epoch 4/10, Loss: 9977.7239
Accuracy: 0.7878
Epoch 5/10, Loss: 9875.0558
Accuracy: 0.7888
Epoch 6/10, Loss: 9838.8850
Accuracy: 0.7884
Epoch 7/10, Loss: 9796.7155
Accuracy: 0.7884
Epoch 8/10, Loss: 9791.9675
Accuracy: 0.7882
Epoch 9/10, Loss: 9774.3923
Accuracy: 0.7859
Epoch 10/10, Loss: 9780.1931
Accuracy: 0.7880
```

```
Running experiment with window_size=2, hidden_layers=0, layer_widths=[128], activation=tanh
Epoch 1/10, Loss: 17622.7117
Accuracy: 0.8092
Epoch 2/10, Loss: 10575.8520
Accuracy: 0.8309
Epoch 3/10, Loss: 9264.7936
Accuracy: 0.8357
Epoch 4/10, Loss: 8587.3021
Accuracy: 0.8374
Epoch 5/10, Loss: 8147.2015
Accuracy: 0.8399
Epoch 6/10, Loss: 7829.0335
Accuracy: 0.8405
Epoch 7/10, Loss: 7584.0456
Accuracy: 0.8415
Epoch 8/10, Loss: 7387.3196
Accuracy: 0.8438
Epoch 9/10, Loss: 7224.5238
Accuracy: 0.8434
Epoch 10/10, Loss: 7086.7163
Accuracy: 0.8434
```

```
Running experiment with window_size=2, hidden_layers=0, layer_widths=[128], activation=relu
Epoch 1/10, Loss: 17611.4354
Accuracy: 0.8108
Epoch 2/10, Loss: 10562.5200
Accuracy: 0.8305
Epoch 3/10, Loss: 9255.5382
Accuracy: 0.8357
Epoch 4/10, Loss: 8580.6246
Accuracy: 0.8392
Epoch 5/10, Loss: 8142.1696
Accuracy: 0.8388
Epoch 6/10, Loss: 7825.0579
Accuracy: 0.8415
Epoch 7/10, Loss: 7580.7665
Accuracy: 0.8428
Epoch 8/10, Loss: 7384.5139
Accuracy: 0.8432
Epoch 9/10, Loss: 7222.0526
Accuracy: 0.8440
Epoch 10/10, Loss: 7084.4911
Accuracy: 0.8444
```

```
Running experiment with window_size=2, hidden_layers=0, layer_widths=[128], activation=sigmoid
Epoch 1/10, Loss: 17624.7949
Accuracy: 0.8102
Epoch 2/10, Loss: 10572.6936
Accuracy: 0.8297
Epoch 3/10, Loss: 9263.3678
Accuracy: 0.8339
Epoch 4/10, Loss: 8586.2811
Accuracy: 0.8382
Epoch 5/10, Loss: 8146.3273
Accuracy: 0.8399
Epoch 6/10, Loss: 7828.2042
Accuracy: 0.8407
Epoch 7/10, Loss: 7583.2058
Accuracy: 0.8417
Epoch 8/10, Loss: 7386.4353
Accuracy: 0.8432
Epoch 9/10, Loss: 7223.5799
Accuracy: 0.8436
Epoch 10/10, Loss: 7085.7116
Accuracy: 0.8436
```

```
Running experiment with window_size=2, hidden_layers=0, layer_widths=[128], activation=identity
Epoch 1/10, Loss: 17660.6093
Accuracy: 0.8102
Epoch 2/10, Loss: 10578.9646
Accuracy: 0.8297
Epoch 3/10, Loss: 9265.2947
Accuracy: 0.8345
Epoch 4/10, Loss: 8587.1378
Accuracy: 0.8392
Epoch 5/10, Loss: 8146.8087
Accuracy: 0.8413
Epoch 6/10, Loss: 7828.5402
Accuracy: 0.8415
Epoch 7/10, Loss: 7583.5033
Accuracy: 0.8426
Epoch 8/10, Loss: 7386.7477
Accuracy: 0.8434
Epoch 9/10, Loss: 7223.9333
Accuracy: 0.8438
Epoch 10/10, Loss: 7086.1127
Accuracy: 0.8446
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[128], activation=tanh
Epoch 1/10, Loss: 14523.4103
Accuracy: 0.8139
Epoch 2/10, Loss: 9063.6011
Accuracy: 0.8253
Epoch 3/10, Loss: 8073.8406
Accuracy: 0.8274
Epoch 4/10, Loss: 7407.0993
Accuracy: 0.8268
Epoch 5/10, Loss: 6846.4414
Accuracy: 0.8314
Epoch 6/10, Loss: 6334.0819
Accuracy: 0.8287
Epoch 7/10, Loss: 5842.5290
Accuracy: 0.8307
Epoch 8/10, Loss: 5365.9442
Accuracy: 0.8314
Epoch 9/10, Loss: 4903.3395
Accuracy: 0.8295
Epoch 10/10, Loss: 4451.5614
Accuracy: 0.8305
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[128], activation=relu
Epoch 1/10, Loss: 15127.5918
Accuracy: 0.8065
Epoch 2/10, Loss: 8616.2857
Accuracy: 0.8243
Epoch 3/10, Loss: 7159.1408
Accuracy: 0.8351
Epoch 4/10, Loss: 6165.1243
Accuracy: 0.8365
Epoch 5/10, Loss: 5305.3954
Accuracy: 0.8368
Epoch 6/10, Loss: 4595.0321
Accuracy: 0.8363
Epoch 7/10, Loss: 3888.8012
Accuracy: 0.8392
Epoch 8/10, Loss: 3294.4430
Accuracy: 0.8390
Epoch 9/10, Loss: 2765.9814
Accuracy: 0.8363
Epoch 10/10, Loss: 2312.6094
Accuracy: 0.8403
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[128], activation=sigmoid
Epoch 1/10, Loss: 26442.3520
Accuracy: 0.7548
Epoch 2/10, Loss: 12295.2856
Accuracy: 0.7866
Epoch 3/10, Loss: 10037.2183
Accuracy: 0.8119
Epoch 4/10, Loss: 8945.5458
Accuracy: 0.8285
Epoch 5/10, Loss: 8295.9821
Accuracy: 0.8363
Epoch 6/10, Loss: 7845.1627
Accuracy: 0.8399
Epoch 7/10, Loss: 7493.5419
Accuracy: 0.8415
Epoch 8/10, Loss: 7197.9198
Accuracy: 0.8428
Epoch 9/10, Loss: 6936.9002
Accuracy: 0.8442
Epoch 10/10, Loss: 6698.4906
Accuracy: 0.8459
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[128], activation=identity
Epoch 1/10, Loss: 14490.3293
Accuracy: 0.8106
Epoch 2/10, Loss: 9440.1019
Accuracy: 0.8156
Epoch 3/10, Loss: 8697.9764
Accuracy: 0.8195
Epoch 4/10, Loss: 8291.7409
Accuracy: 0.8173
Epoch 5/10, Loss: 8031.8953
Accuracy: 0.8177
Epoch 6/10, Loss: 7847.0016
Accuracy: 0.8162
Epoch 7/10, Loss: 7707.9982
Accuracy: 0.8154
Epoch 8/10, Loss: 7601.7594
Accuracy: 0.8160
Epoch 9/10, Loss: 7513.0294
Accuracy: 0.8144
Epoch 10/10, Loss: 7448.0557
Accuracy: 0.8129
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[256], activation=tanh
Epoch 1/10, Loss: 14226.0300
Accuracy: 0.8137
Epoch 2/10, Loss: 9213.2687
Accuracy: 0.8189
Epoch 3/10, Loss: 8350.4508
Accuracy: 0.8181
Epoch 4/10, Loss: 7798.9268
Accuracy: 0.8179
Epoch 5/10, Loss: 7353.3800
Accuracy: 0.8216
Epoch 6/10, Loss: 6952.7712
Accuracy: 0.8224
Epoch 7/10, Loss: 6571.7411
Accuracy: 0.8200
Epoch 8/10, Loss: 6194.0158
Accuracy: 0.8200
Epoch 9/10, Loss: 5817.1188
Accuracy: 0.8195
Epoch 10/10, Loss: 5434.1170
Accuracy: 0.8210
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[256], activation=relu
Epoch 1/10, Loss: 14829.0884
Accuracy: 0.8104
Epoch 2/10, Loss: 8427.6677
Accuracy: 0.8328
Epoch 3/10, Loss: 6850.3186
Accuracy: 0.8440
Epoch 4/10, Loss: 5708.8259
Accuracy: 0.8430
Epoch 5/10, Loss: 4746.2570
Accuracy: 0.8448
Epoch 6/10, Loss: 3852.5889
Accuracy: 0.8432
Epoch 7/10, Loss: 3118.9606
Accuracy: 0.8467
Epoch 8/10, Loss: 2521.6285
Accuracy: 0.8467
Epoch 9/10, Loss: 1971.8248
Accuracy: 0.8502
Epoch 10/10, Loss: 1635.7225
Accuracy: 0.8405
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[256], activation=sigmoid
Epoch 1/10, Loss: 26606.6951
Accuracy: 0.7519
Epoch 2/10, Loss: 12411.4481
Accuracy: 0.7808
Epoch 3/10, Loss: 10195.0816
Accuracy: 0.8023
Epoch 4/10, Loss: 9118.9242
Accuracy: 0.8200
Epoch 5/10, Loss: 8469.3281
Accuracy: 0.8291
Epoch 6/10, Loss: 8013.2737
Accuracy: 0.8345
Epoch 7/10, Loss: 7655.5024
Accuracy: 0.8365
Epoch 8/10, Loss: 7354.8102
Accuracy: 0.8388
Epoch 9/10, Loss: 7090.6257
Accuracy: 0.8399
Epoch 10/10, Loss: 6851.1041
Accuracy: 0.8409
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[256], activation=identity
Epoch 1/10, Loss: 14210.2938
Accuracy: 0.8092
Epoch 2/10, Loss: 9404.1922
Accuracy: 0.8137
Epoch 3/10, Loss: 8678.3133
Accuracy: 0.8156
Epoch 4/10, Loss: 8277.1105
Accuracy: 0.8150
Epoch 5/10, Loss: 8017.5859
Accuracy: 0.8160
Epoch 6/10, Loss: 7837.6156
Accuracy: 0.8148
Epoch 7/10, Loss: 7698.2440
Accuracy: 0.8144
Epoch 8/10, Loss: 7600.2831
Accuracy: 0.8164
Epoch 9/10, Loss: 7504.6157
Accuracy: 0.8139
Epoch 10/10, Loss: 7453.7500
Accuracy: 0.8119
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[512], activation=tanh
Epoch 1/10, Loss: 13848.9687
Accuracy: 0.8123
Epoch 2/10, Loss: 9302.3052
Accuracy: 0.8158
Epoch 3/10, Loss: 8542.1414
Accuracy: 0.8175
Epoch 4/10, Loss: 8085.6423
Accuracy: 0.8171
Epoch 5/10, Loss: 7752.9468
Accuracy: 0.8164
Epoch 6/10, Loss: 7479.7942
Accuracy: 0.8175
Epoch 7/10, Loss: 7239.0911
Accuracy: 0.8168
Epoch 8/10, Loss: 7013.5888
Accuracy: 0.8168
Epoch 9/10, Loss: 6788.6555
Accuracy: 0.8177
Epoch 10/10, Loss: 6551.4881
Accuracy: 0.8183
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[512], activation=relu
Epoch 1/10, Loss: 14509.2911
Accuracy: 0.8144
Epoch 2/10, Loss: 8264.3165
Accuracy: 0.8336
Epoch 3/10, Loss: 6621.1967
Accuracy: 0.8444
Epoch 4/10, Loss: 5379.2777
Accuracy: 0.8448
Epoch 5/10, Loss: 4314.5784
Accuracy: 0.8463
Epoch 6/10, Loss: 3400.0859
Accuracy: 0.8451
Epoch 7/10, Loss: 2594.5252
Accuracy: 0.8463
Epoch 8/10, Loss: 1981.4804
Accuracy: 0.8486
Epoch 9/10, Loss: 1603.4430
Accuracy: 0.8494
Epoch 10/10, Loss: 1232.7211
Accuracy: 0.8486
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[512], activation=sigmoid
Epoch 1/10, Loss: 27614.5239
Accuracy: 0.7374
Epoch 2/10, Loss: 12848.3195
Accuracy: 0.7662
Epoch 3/10, Loss: 10540.7330
Accuracy: 0.7890
Epoch 4/10, Loss: 9413.8629
Accuracy: 0.8069
Epoch 5/10, Loss: 8736.7081
Accuracy: 0.8166
Epoch 6/10, Loss: 8261.0065
Accuracy: 0.8243
Epoch 7/10, Loss: 7890.4621
Accuracy: 0.8299
Epoch 8/10, Loss: 7582.4769
Accuracy: 0.8320
Epoch 9/10, Loss: 7315.1235
Accuracy: 0.8328
Epoch 10/10, Loss: 7075.8915
Accuracy: 0.8349
```

```
Running experiment with window_size=2, hidden_layers=1, layer_widths=[512], activation=identity
Epoch 1/10, Loss: 13876.4998
Accuracy: 0.8088
Epoch 2/10, Loss: 9400.7100
Accuracy: 0.8114
Epoch 3/10, Loss: 8698.0301
Accuracy: 0.8104
Epoch 4/10, Loss: 8303.3820
Accuracy: 0.8112
Epoch 5/10, Loss: 8046.9653
Accuracy: 0.8139
Epoch 6/10, Loss: 7868.4167
Accuracy: 0.8137
Epoch 7/10, Loss: 7727.3670
Accuracy: 0.8146
Epoch 8/10, Loss: 7629.8587
Accuracy: 0.8152
Epoch 9/10, Loss: 7533.1447
Accuracy: 0.8135
Epoch 10/10, Loss: 7483.4002
Accuracy: 0.8119
```

```
Running experiment with window_size=2, hidden_layers=2, layer_widths=[128, 256], activation=tanh
Epoch 1/10, Loss: 15126.1002
Accuracy: 0.7704
Epoch 2/10, Loss: 9854.7298
Accuracy: 0.7938
Epoch 3/10, Loss: 8780.1966
Accuracy: 0.8040
Epoch 4/10, Loss: 7998.5980
Accuracy: 0.8061
Epoch 5/10, Loss: 7268.9637
Accuracy: 0.8056
Epoch 6/10, Loss: 6604.2038
Accuracy: 0.8077
Epoch 7/10, Loss: 5951.6983
Accuracy: 0.8090
Epoch 8/10, Loss: 5419.7685
Accuracy: 0.8044
Epoch 9/10, Loss: 4999.3830
Accuracy: 0.8160
Epoch 10/10, Loss: 4613.4847
Accuracy: 0.8096
```

```
Running experiment with window_size=2, hidden_layers=2, layer_widths=[128, 256], activation=relu
Epoch 1/10, Loss: 16404.3839
Accuracy: 0.7797
Epoch 2/10, Loss: 9041.9586
Accuracy: 0.8158
Epoch 3/10, Loss: 7255.2857
Accuracy: 0.8220
Epoch 4/10, Loss: 6101.3034
Accuracy: 0.8363
Epoch 5/10, Loss: 5231.0643
Accuracy: 0.8341
Epoch 6/10, Loss: 4475.1448
Accuracy: 0.8365
Epoch 7/10, Loss: 3705.5392
Accuracy: 0.8417
Epoch 8/10, Loss: 3635.0404
Accuracy: 0.8303
Epoch 9/10, Loss: 3297.2430
Accuracy: 0.8361
Epoch 10/10, Loss: 3138.6820
Accuracy: 0.8268
```

```
Running experiment with window_size=2, hidden_layers=2, layer_widths=[128, 256], activation=sigmoid
Epoch 1/10, Loss: 45530.8811
Accuracy: 0.2286
Epoch 2/10, Loss: 31157.3283
Accuracy: 0.5907
Epoch 3/10, Loss: 19248.7363
Accuracy: 0.7096
Epoch 4/10, Loss: 14417.4578
Accuracy: 0.7527
Epoch 5/10, Loss: 12467.3093
Accuracy: 0.7689
Epoch 6/10, Loss: 11450.4468
Accuracy: 0.7799
Epoch 7/10, Loss: 10659.2563
Accuracy: 0.7930
Epoch 8/10, Loss: 9999.1880
Accuracy: 0.8017
Epoch 9/10, Loss: 9466.1900
Accuracy: 0.8098
Epoch 10/10, Loss: 9022.2374
Accuracy: 0.8152
```

```
Running experiment with window_size=2, hidden_layers=2, layer_widths=[128, 256], activation=identity
Epoch 1/10, Loss: 15438.9207
Accuracy: 0.7613
Epoch 2/10, Loss: 10768.2714
Accuracy: 0.7700
Epoch 3/10, Loss: 10215.0795
Accuracy: 0.7679
Epoch 4/10, Loss: 10020.5387
Accuracy: 0.7633
Epoch 5/10, Loss: 9983.0680
Accuracy: 0.7604
Epoch 6/10, Loss: 9937.2896
Accuracy: 0.7695
Epoch 7/10, Loss: 9947.5737
Accuracy: 0.7563
Epoch 8/10, Loss: 9978.4020
Accuracy: 0.7530
Epoch 9/10, Loss: 10066.8808
Accuracy: 0.7511
Epoch 10/10, Loss: 10114.3206
Accuracy: 0.7527
```

```
Running experiment with window_size=2, hidden_layers=2, layer_widths=[256, 512], activation=tanh
Epoch 1/10, Loss: 14864.3786
Accuracy: 0.7664
Epoch 2/10, Loss: 10244.7729
Accuracy: 0.7752
Epoch 3/10, Loss: 9349.9930
Accuracy: 0.7772
Epoch 4/10, Loss: 8699.8269
Accuracy: 0.7756
Epoch 5/10, Loss: 8067.0438
Accuracy: 0.7861
Epoch 6/10, Loss: 7499.0170
Accuracy: 0.7963
Epoch 7/10, Loss: 7010.9434
Accuracy: 0.7944
Epoch 8/10, Loss: 6498.3746
Accuracy: 0.8007
Epoch 9/10, Loss: 6075.2399
Accuracy: 0.7940
Epoch 10/10, Loss: 5818.8702
Accuracy: 0.7998
```

```
Running experiment with window_size=2, hidden_layers=2, layer_widths=[256, 512], activation=relu
Epoch 1/10, Loss: 15770.2576
Accuracy: 0.7847
Epoch 2/10, Loss: 8654.4397
Accuracy: 0.8135
Epoch 3/10, Loss: 6795.7003
Accuracy: 0.8266
Epoch 4/10, Loss: 5402.0143
Accuracy: 0.8372
Epoch 5/10, Loss: 4333.4106
Accuracy: 0.8324
Epoch 6/10, Loss: 3607.5626
Accuracy: 0.8324
Epoch 7/10, Loss: 3142.6915
Accuracy: 0.8365
Epoch 8/10, Loss: 3001.1622
Accuracy: 0.8324
Epoch 9/10, Loss: 2331.0846
Accuracy: 0.8376
Epoch 10/10, Loss: 2134.9771
Accuracy: 0.8390
```

```
Running experiment with window_size=2, hidden_layers=2, layer_widths=[256, 512], activation=sigmoid
Epoch 1/10, Loss: 45810.3208
Accuracy: 0.2209
Epoch 2/10, Loss: 36052.6947
Accuracy: 0.5115
Epoch 3/10, Loss: 22552.0571
Accuracy: 0.6729
Epoch 4/10, Loss: 15910.3687
Accuracy: 0.7397
Epoch 5/10, Loss: 13135.0104
Accuracy: 0.7586
Epoch 6/10, Loss: 11905.7308
Accuracy: 0.7733
Epoch 7/10, Loss: 11057.3383
Accuracy: 0.7866
Epoch 8/10, Loss: 10413.1107
Accuracy: 0.7944
Epoch 9/10, Loss: 9903.7360
Accuracy: 0.7982
Epoch 10/10, Loss: 9483.2444
Accuracy: 0.8023
```

```
Running experiment with window_size=2, hidden_layers=2, layer_widths=[256, 512], activation=identity
Epoch 1/10, Loss: 15084.1372
Accuracy: 0.7583
Epoch 2/10, Loss: 10822.8753
Accuracy: 0.7660
Epoch 3/10, Loss: 10271.7863
Accuracy: 0.7662
Epoch 4/10, Loss: 10094.6579
Accuracy: 0.7619
Epoch 5/10, Loss: 10042.5179
Accuracy: 0.7615
Epoch 6/10, Loss: 10039.0284
Accuracy: 0.7571
Epoch 7/10, Loss: 10056.2066
Accuracy: 0.7542
Epoch 8/10, Loss: 10144.6782
Accuracy: 0.7498
Epoch 9/10, Loss: 10157.6015
Accuracy: 0.7532
Epoch 10/10, Loss: 10267.6899
Accuracy: 0.7501
```

Summary of all the experimentation:

- Increasing the window size generally improves accuracy by providing more context.
- Adding hidden layers improves performance, but gains can be minimal beyond 1 layer.
- Wider layer widths (e.g., 512 neurons) usually yield better results, provided that overfitting is managed.
- ReLU activation is often the most effective for POS tagging tasks due to its efficiency in handling gradients.