Privacy Preserving Decision Tree Prediction On Encrypted Data

Reem Younis, Assia Khateeb, Atheer Abo Foul, Aya Miari Lecturer: Dr. Adi Akavia

Laboratory in Privacy Preserving Machine Learning, University of Haifa

Email: reembyounis@gmail.com, assia.khteb@gmail.com, 19aether6@gmail.com, aia-m-211@hotmail.com

August 18, 2021

Contents

1	Introduction	2
2	ARCHITECTURE	2
3	Single-layer Perceptron	2
4	Multi-layer Perceptron	2
5	RESULTS	2
6	Conclusions	11
7	Platforms	11

Abstract. In part D we implement the prediction of iris species using perceptron neural networks trained on the Iris data set. It includes write-ups of the different types of perceptrons and their accuracy.

Keywords— Perceptron: either a single-layer or mutli-layer feed-forward neural network.

1 Introduction

Reference to our Github repo.

In part D we'll describe the prediction of iris species using two different perceptron neural networks: a single-layer and multi-layer perceptron.

Each perceptron is trained and evaluated on the Iris data set split into train and test sets.

2 ARCHITECTURE

The main architecture of the neural networks are based on a single-layer perceptron and a multi-layer perceptron. Each network was trained using the categorical cross entropy loss function, since the problem consists of multiple classes, and the Adam optimizer, due to its practical advantage over alternatives.

The Iris dataset was split beforehand into train and test sets. It consists of samples of 4 features, sepal length, sepal width, petal length, petal width, with 1 of 3 classes, 1.setosa, 2.versicolor, 3.virginica.

3 Single-layer Perceptron

The single-layer perceptron is modeled with an input layer of 4 nodes, one for each feature, and an output layer of 3 nodes, one for each class, with a softmax activation function, since the problem consists of multiple classes.

This was chosen to be a benchmark to see the performance increase with the multi-layer perceptron.

This single-layer perceptron should have a very low accuracy due to its extreme simplicity.

4 Multi-layer Perceptron

The multi-layer perceptron is modeled with an input layer of 4 nodes, one for each feature, 2 hidden layers of 10 nodes with a ReLU activation function, and an output layer of 3 nodes, one for each class, with a softmax activation function, since the problem consists of multiple classes. ReLU was used due to its practical advantage seen in research papers. Theoretically, the 2 hidden layers should be able to learn higher abstract information that the single-layer perceptron could not model. Thus, it is expected to produce a higher accuracy than the single-layer perceptron.

In addition to the change in the network architecture, the multi-layer perceptron also received proprocessed data, that is, data scaled down to the range of -1 to 1. This was done exclusively for the features as the classes are simply represented as a 0 or 1 and thus do not need any scaling. Theoretically, this change should allow the network to train faster as it does not need to give priority to training features that are high in value.

Finally, after each layer in the network, batch normalization was run to further allow the network to train faster and better.

5 RESULTS

```
Epoch 4/100
Epoch 5/100
Epoch 6/100
Epoch 7/100
Epoch 8/100
Epoch 9/100
Epoch 10/100
Epoch 11/100
Epoch 12/100
Epoch 13/100
Epoch 14/100
Epoch 15/100
Epoch 16/100
Epoch 17/100
Epoch 18/100
Epoch 19/100
Epoch 20/100
Epoch 21/100
Epoch 22/100
Epoch 23/100
Epoch 24/100
Epoch 25/100
Epoch 26/100
Epoch 27/100
Epoch 28/100
Epoch 29/100
```

```
Epoch 30/100
Epoch 31/100
Epoch 32/100
Epoch 33/100
Epoch 34/100
Epoch 35/100
Epoch 36/100
Epoch 37/100
Epoch 38/100
Epoch 39/100
Epoch 40/100
Epoch 41/100
Epoch 42/100
Epoch 43/100
Epoch 44/100
Epoch 45/100
Epoch 46/100
Epoch 47/100
Epoch 48/100
Epoch 49/100
Epoch 50/100
Epoch 51/100
Epoch 52/100
Epoch 53/100
Epoch 54/100
Epoch 55/100
```

```
Epoch 56/100
Epoch 57/100
Epoch 58/100
Epoch 59/100
Epoch 60/100
Epoch 61/100
Epoch 62/100
Epoch 63/100
Epoch 64/100
Epoch 65/100
Epoch 66/100
Epoch 67/100
Epoch 68/100
Epoch 69/100
Epoch 70/100
Epoch 71/100
Epoch 72/100
Epoch 73/100
Epoch 74/100
Epoch 75/100
Epoch 76/100
Epoch 77/100
Epoch 78/100
Epoch 79/100
Epoch 80/100
Epoch 81/100
```

```
Epoch 82/100
Epoch 83/100
Epoch 84/100
Epoch 85/100
Epoch 86/100
Epoch 87/100
Epoch 88/100
Epoch 89/100
Epoch 90/100
Epoch 91/100
Epoch 92/100
Epoch 93/100
Epoch 94/100
Epoch 95/100
Epoch 96/100
Epoch 97/100
Epoch 98/100
Epoch 99/100
Epoch 100/100
Epoch 1/100
Epoch 2/100
Epoch 3/100
Epoch 4/100
Epoch 5/100
Epoch 6/100
Epoch 7/100
```

```
Epoch 8/100
Epoch 9/100
Epoch 10/100
Epoch 11/100
Epoch 12/100
Epoch 13/100
Epoch 14/100
Epoch 15/100
Epoch 16/100
Epoch 17/100
Epoch 18/100
Epoch 19/100
Epoch 20/100
Epoch 21/100
Epoch 22/100
Epoch 23/100
Epoch 24/100
Epoch 25/100
Epoch 26/100
Epoch 27/100
Epoch 28/100
Epoch 29/100
Epoch 30/100
Epoch 31/100
Epoch 32/100
Epoch 33/100
```

```
Epoch 34/100
Epoch 35/100
Epoch 36/100
Epoch 37/100
Epoch 38/100
Epoch 39/100
Epoch 40/100
Epoch 41/100
Epoch 42/100
Epoch 43/100
Epoch 44/100
Epoch 45/100
Epoch 46/100
Epoch 47/100
Epoch 48/100
Epoch 49/100
Epoch 50/100
Epoch 51/100
Epoch 52/100
Epoch 53/100
Epoch 54/100
Epoch 55/100
Epoch 56/100
Epoch 57/100
Epoch 58/100
Epoch 59/100
```

```
Epoch 60/100
Epoch 61/100
Epoch 62/100
Epoch 63/100
Epoch 64/100
Epoch 65/100
Epoch 66/100
Epoch 67/100
Epoch 68/100
Epoch 69/100
Epoch 70/100
Epoch 71/100
Epoch 72/100
Epoch 73/100
Epoch 74/100
Epoch 75/100
Epoch 76/100
Epoch 77/100
Epoch 78/100
Epoch 79/100
Epoch 80/100
Epoch 81/100
Epoch 82/100
Epoch 83/100
Epoch 84/100
Epoch 85/100
```

```
Epoch 86/100
Epoch 87/100
Epoch 88/100
Epoch 89/100
Epoch 90/100
Epoch 91/100
Epoch 92/100
Epoch 93/100
Epoch 94/100
Epoch 95/100
Epoch 96/100
Epoch 97/100
Epoch 98/100
Epoch 99/100
Epoch 100/100
```

Naive Accuracy: 0.6666666865348816

Naive Predictions:

```
[[0.01354977 0.5061691
                         0.48028105]
[0.06246578 0.56894875 0.36858547]
[0.0074184
            0.44270876 \ 0.5498728 ]
[0.01080847 0.4541977
                        0.53499377]
[0.02312316 0.55372
                        0.4231569
[0.00527738 \ 0.39130107 \ 0.60342157]
[0.00576249 0.52103436 0.4732032 ]
[0.00304265 \ 0.31141222 \ 0.68554515]
[0.07634713 0.58648854 0.33716434]
[0.00298612 0.3317028
                        0.6653111
[0.00340858 0.28069374 0.7158977 ]
[0.01619304 0.5350266
                        0.44878045]
[0.00580444 \ 0.32919884 \ 0.66499674]
[0.01750648 0.56174
                        0.4207535 ]
[0.00643218 \ 0.49766505 \ 0.49590284]]
```

Better Accuracy: 0.9333333373069763

Better Predictions:

```
[[0.00862633 0.9620326
                         0.02934105
[0.9790445
            0.01505063 0.005904891
[0.00578253 0.9523178
                        0.0418997 ]
[0.00888783 0.97394866 0.01716346]
[0.07783689 0.7403439
                        0.18181916]
[0.01994696 0.21152395 0.7685291 ]
            0.86186147 0.123285971
[0.0148526
[0.00590803 \ 0.02657974 \ 0.96751225]
[0.9846364
            0.0105518
                        0.00481179]
[0.00601991 \ 0.05104782 \ 0.94293225]
[0.00143864 0.0452151
                        0.95334625]
[0.01584265 0.9249697
                        0.0591877 ]
[0.01459309 0.3523831
                        0.633023741
[0.01764467 0.9287313
                        0.05362389]
[0.01701402 \ 0.77023745 \ 0.21274848]]
```

As seen in the example run, the multi-layer perceptron performed much better than the single-layer perceptron. The accuracy for the single-layer perceptron and multi-layer perceptron was 66.666% and 93.333%, respectively.

Thus, it can easily be stated that the hidden layers and preprocessing of data allowed the multi-layer perceptron to much better learn the data.

The 2 hidden layers allowed the multi-layer perceptron to learn higher abstract features about the Iris data set and thus produced a higher accuracy. Furthermore, the data scaling and batch normalization increased the network's learning rate.

Although a 93.333% accuracy is astounding, this is not likely to continue if the test data set size increased.

6 Conclusions

Overall, the results gathered follow the expectation/hypothesis, that is, that a more complex network, but not too complex, would be able to learn much more effectively than a simple one. This is likely in many scenarios.

However, it is likely that the bias-variance tradeoff effect would come into play here. That is, as the bias of a model increases, its variance will as well, and vise-versa. A simple model has extreme bias, but no variance.

A extremely complex model has high variance, but almost no bias. Thus, an extremely complex model is likely to overfit a data set, while a simple model is likely to underfit a data set.

Therefore, it is evident that caution must be taken to find a point where the bias and variance are minimized while still producing an effective model.

This can be done by using a technique like cross-validation to produce models that are each trained and validated against different "folds" of the train data set. Then, the model that minimizes the loss function (against the sum of error of the validation folds) can be used as the final model and can be run on the test set to produce the resulting accuracy.

7 Platforms

- pycharm, python 3.9
- overleaf.com