Handwritten Digits Classification

Course: CSE 574 – Introduction to Machine Learning

Team:

Shad Ullah Khan (shadulla@buffalo.edu)
Rudra Prasad Baksi (rudrapra@buffalo.edu)
Leila Talebpour (leilatal@buffalo.edu)

1. Introduction

Neural Networks are computational method, which are put to use to analyze a data set just like the neurons or the nerve cells of the human brain. The assignment involves implementing a Multi-Layer Perceptron Neural Network and evaluating its performance in classifying handwritten digits. In the first part of the assignment, a neural network with a single hidden layer is used for the aforementioned job. In the later part, multilayer neural networks are implemented using TensorFlow libraries and their performances are studied.

2. Hyper-parameters for the Neural Network

Hyper-parameters are some of the important variables, which are generally set before the other parameters are optimized. They are used to optimize the performance of the model. Optimizing the Hyper-parameters improves the learning capabilities of the model. The hyper-parameters used in this project are the number of hidden units (n) and the regularization coefficient lambda (λ). To get an optimized result, one parameter is varied while keeping the other fixed and then same thing is done with the other one. The hyper-parameters used in this project are meant to avoid the over-fitting problem of the model.

In this project the number hidden units value is kept fixed at 50 ($n = 50$) and then varied the lambda value (λ). Figure 1 shows the graph of accuracy versus the lambda value (λ) for the training data, test data and the validation data.

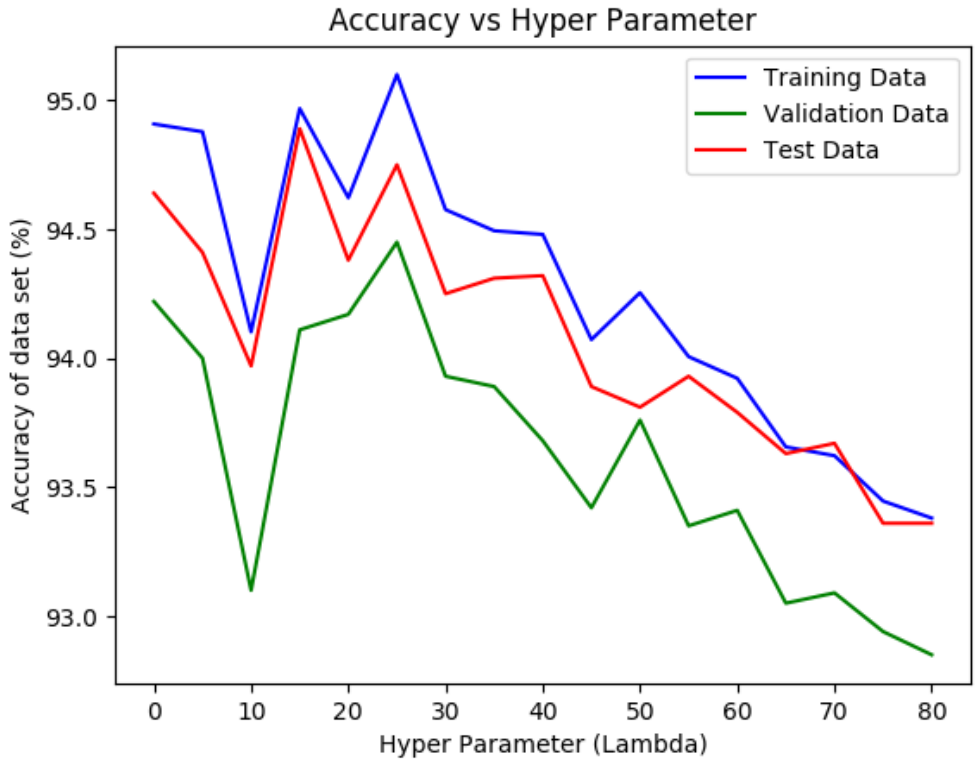

Figure 1: Accuracy vs. Lambda Value for Training, Test and Validation Data Sets

The graph from Figure 1, gives the optimized lambda value (λ = 25), for which the accuracies for the training, test and validation data sets are the highest. Now keeping the lambda value constant (λ = 25), the number of hidden units is varied, which is shown in Figure 2.

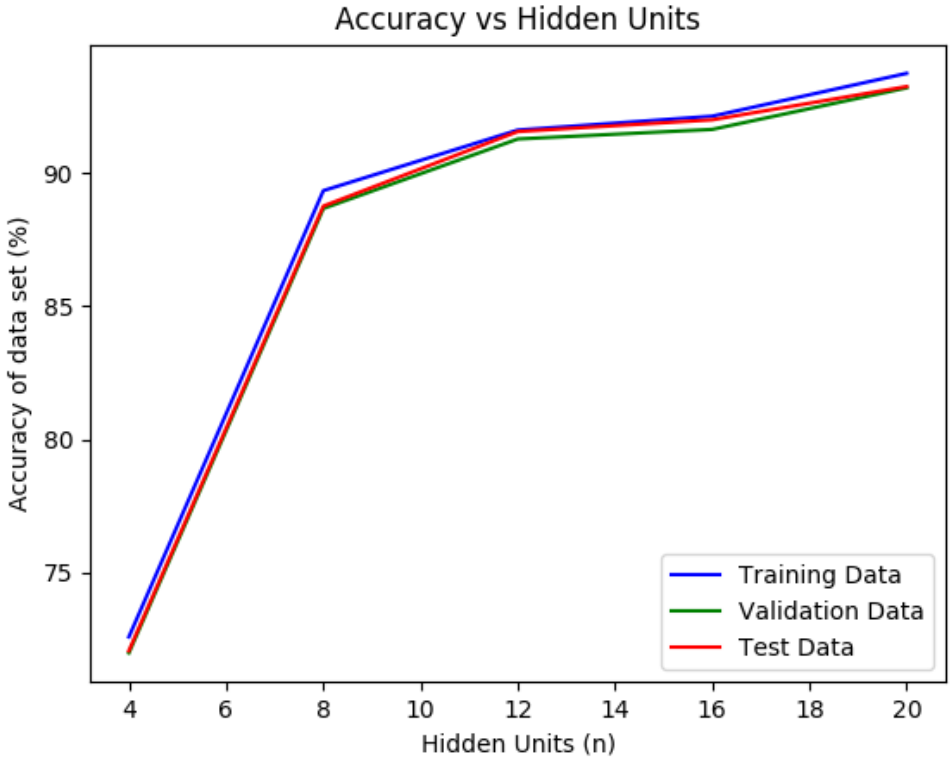

Figure 2: Accuracy vs. Number of Hidden Units for Training, Test and Validation Data Sets

The graph in Figure 2, gives an idea how the accuracy of the model varied with the number of hidden units (n). It was found that as the number of hidden units increased, the accuracy initially increased rapidly. Then for around $n = 20$, the accuracy was quite high. Further increasing the number of hidden units would have increased the accuracy but not that significantly. On the other hand, increasing number of hidden units beyond $n = 20$, would have significantly increased the training time, which is evident from Figure 3.

Figure 3: Training Time vs. Number of Hidden Nodes (n)

3. Handwritten Digits: Accuracy of Classification on Test Data Set

The accuracy achieved by the script for the Training Data Set is 93.422%, for the Validation Data Set is 92.53% and for the Test Data Set is 93.08%. Figure 4 gives the output of the script.

```
Shads-MacBook-Air:NeuralNetwork shadkhan$ python nnScript.py
preprocess done

 Training set Accuracy:93.422%

 Validation set Accuracy:92.53%

 Test set Accuracy:93.08%
```

Figure 4: Accuracy of Classification on the Handwritten Digits Data Set

4. CelebA: Accuracy of Classification on Test Data Set

The accuracy achieved by the script for classification method on the CelebA Data Set is 84.677% for the Training Data Set, 83.452% for the Validation Data Set and 84.444% for the Test Data Set. Figure 5, gives the output of the script for the CelebA Data Set.

```
[metallica {~/Machine_Learning/assignment_1} > python facennScript\ \(1\).py
 Total time taken: 135.6141722202301 seconds

  Training set Accuracy:84.6777251185%

  Validation set Accuracy:83.4521575985%

  Test set Accuracy:84.4436033308%
metallica {~/Machine_Learning/assignment_1} >
```

Figure 5: Accuracy of Classification Method on the CelebA Data Set

5. Neural Network vs. Deep Neural Network (using TensorFlow)

The analysis of the CelebA Data Set was carried out using Neural Networks with multiple hidden layers. Multiple Hidden Layers were achieved in the script by using the TensorFlow libraries. The neural networks using TensorFlow libraries consisted of 3, 5 and 7 layers respectively. Figure 6 gives the accuracies with the aforementioned number of layers.

```
 Last login: Wed Mar  8 01:52:31 on ttys000
[Rudras-MacBook-Pro:~ rudraprasadbaksi$ ssh rudrapra@springsteen.cse.buffalo.edu
[rudrapra@springsteen.cse.buffalo.edu's password:
 Last login: Wed Mar  8 01:52:51 2017 from cpe-98-5-147-99.buffalo.res.rr.com

[springsteen {~} > cd Machine_Learning/
[springsteen {~/Machine_Learning} > cd assignment_1/
[springsteen {~/Machine_Learning/assignment_1} > ls
 deep5nnScript.py  deep7nnScript.py  deepnn3Script.py  face_all.pickle  facennScript (1).py  mnist_all.mat  nnScript.py
[springsteen {~/Machine_Learning/assignment_1} > python deepnn3Script.py
 Optimization Finished!
 Accuracy: 0.789932
 359.225923538208
[springsteen {~/Machine_Learning/assignment_1} > python deep5nnScript.py
 Optimization Finished!
 Accuracy: 0.756624
 227.65772199630737
[springsteen {~/Machine_Learning/assignment_1} > python deep7nnScript.py
 Optimization Finished!
 Accuracy: 0.722937
 367.41997838020325
 springsteen {~/Machine_Learning/assignment_1} >
```

Figure 6: Accuracies for Neural Networks with 3, 5, and 7 Hidden layers respectively

The neural network with 3 hidden layers ran with an accuracy of 78.993%. The neural networks with 5 and 7 hidden layers ran with accuracies 75.9662% and 72.294% respectively. The script with a single hidden layer used earlier ran with an accuracy of 84.667% on the Training Data Set.

The training time taken by the single hidden layer neural network on the CelebA data set is 135.614 seconds. On the other hand the training time for neural networks with 3, 5 and 7 hidden layers are 359.226, 227.658, and 367.42 seconds respectively. The training time decreased as the number of hidden layers increased from 3 to 5. But the training

time increased as the number of layers increased from 5 to 7. The training time for the neural network with 7 hidden layers is the highest. The accuracy on the training data set is highest for the neural network with a single hidden layer.

It is thus observed that increasing the number of hidden layers might not always be good for the accuracy. Thus, going by the Principle of Occam's Razor, it can be safely said, that increasing the complexity may not always be favorable for the performance or the result.