Programming Assignment 2

UNIVERSITY AT BUFFALO

INTRODUCTION TO MACHINE LEARNING CSE 574

JOSE A. BARJAS
MAJD ABU GHAZALEH

Overview

In this project, we implement a Multilayer Perceptron Neural Network and evaluate its performance in classifying handwritten digits from MNIST dataset. We evaluate performance with different values of the regularization parameter and the number of hidden units utilized. The target is to determine the appropriate value for the regularization parameter and the number of hidden nodes that yield high accuracy and reasonable computation time.

Regularization Parameter (λ)

The accuracy and speed of a neural network depend on many factors such as the number of hidden layers, the number of hidden units in each layer, the size of the data set and the regularization -parameter. Figure 1 below shows the values of the regularization parameter (λ) plotted against the accuracy (%) of the network on the validation set with a network of 20 hidden units in its hidden layer. The figure shows that the best value for the regularization parameter λ is 0.3.

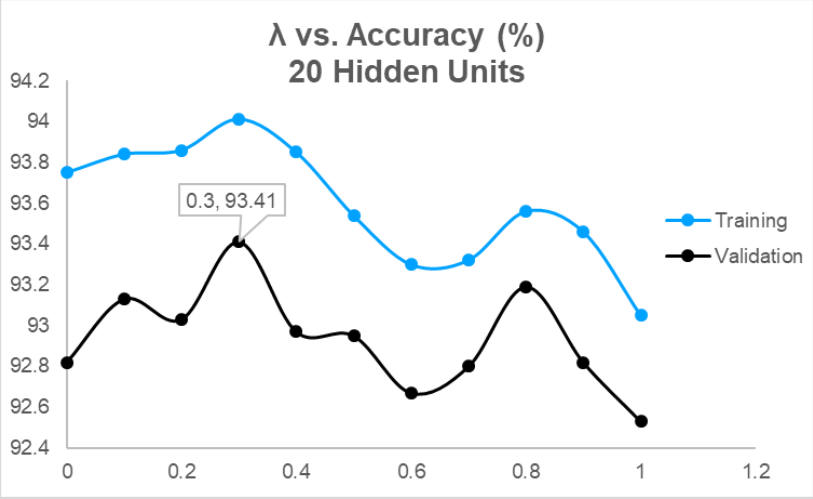

Figure 1: λ vs Accuracy (%) for 20 Hidden Units

Repeating the above procedure with different values for the number of hidden units (8,12,16,35) gives us a better idea of the best hyper-parameter value to choose. Figure 2 (see below) shows that the average λ value comes out to ≈ 0.3.

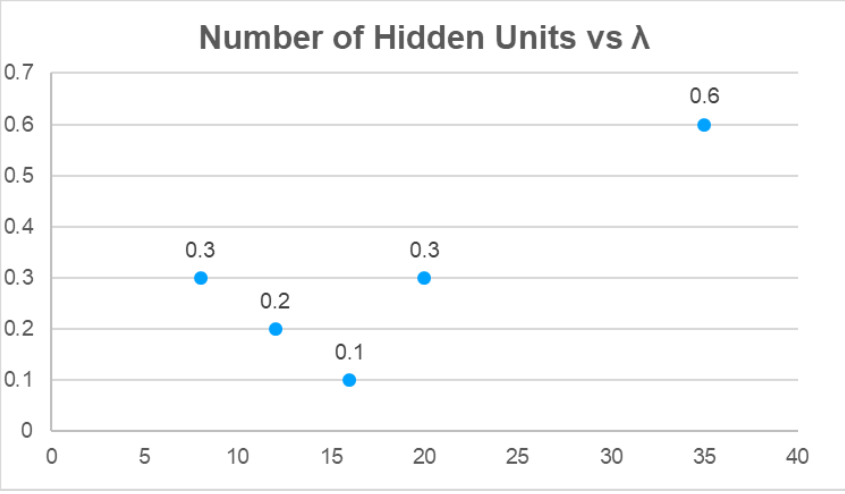

Figure 2: Number of Hidden Units vs. λ

Hidden Layer Units

The number of hidden units was plotted against the test set accuracy results to evaluate the performance of the network with increasing hidden units in its hidden layer. Figure 3 shows that the accuracy generally increased as the number of hidden units increased (note: regularization parameter λ was ideally chosen for each number of hidden unit). The presence of more hidden units means that the network is able to find more minute error and adjust the error accordingly. It is notable that the test accuracy slightly dips at 50 hidden units. This suggests that the accuracy can reach a threshold. The goal is to find a suitable number of hidden units at which performance is acceptable and computation time isn't excessive.

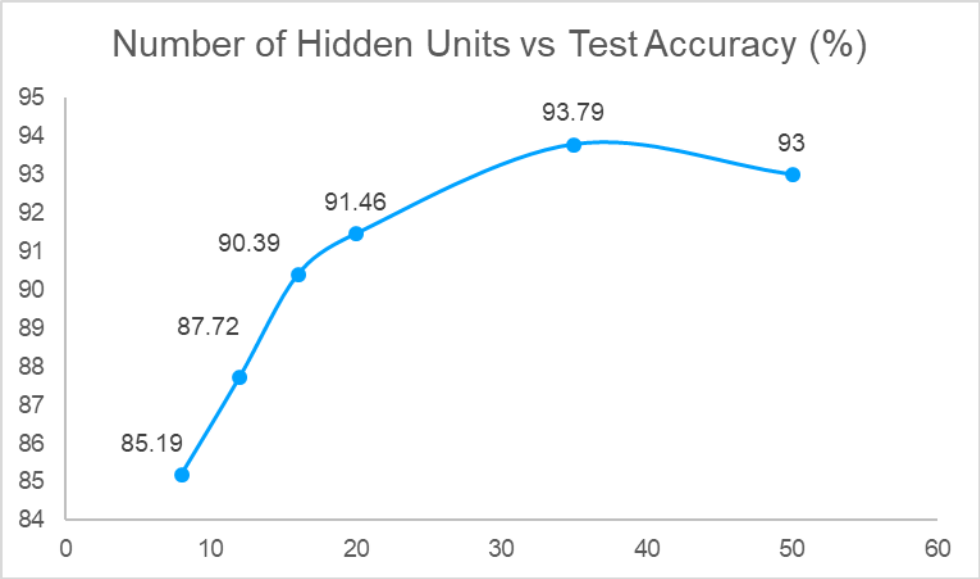

Figure 3: Number of Hidden Units vs Test Set Accuracy (%)

However, as the number of hidden units increases the time taken to train and predict increases as well as seen in Figure 4 below. This makes sense because the network does more computations as the number of hidden units increases. The complexity increases with more hidden units. Finding a good balance between training time and test accuracy is important. From our samples, 35 hidden units satisfies that balance.

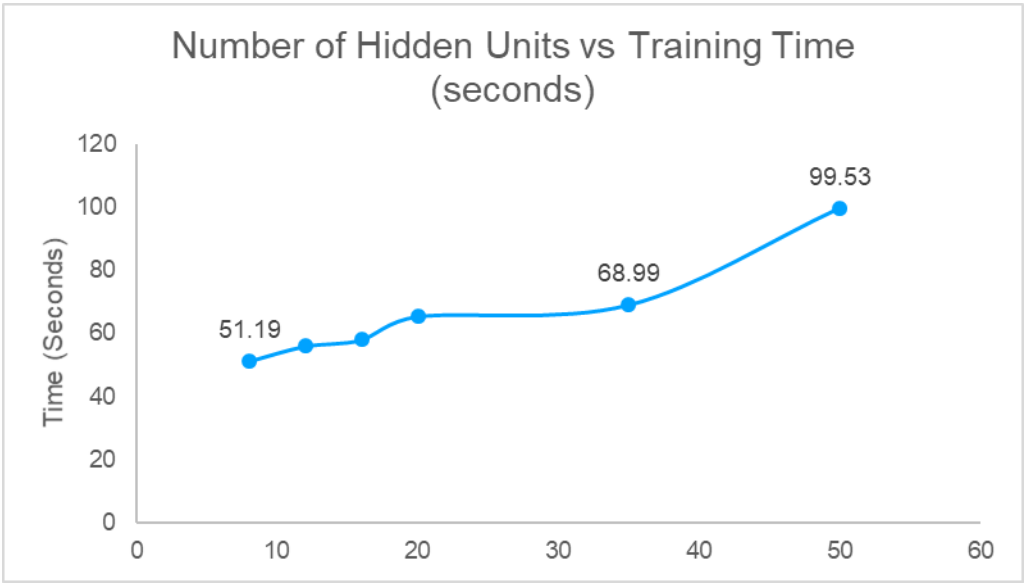

Figure 4: Number of Hidden Units vs Training Time (seconds)