Handwritten Digits Classification Using Neural Networks

Group No. 49

CSE 574

Miraj Kheni (mirajsub)

Rushin Shah (rushinka)

Objective

Implement a Multilayer Perceptron Neural Network and evaluate its performance in classifying handwritten digits. We need to incorporate following tasks in order to build Neural Network:

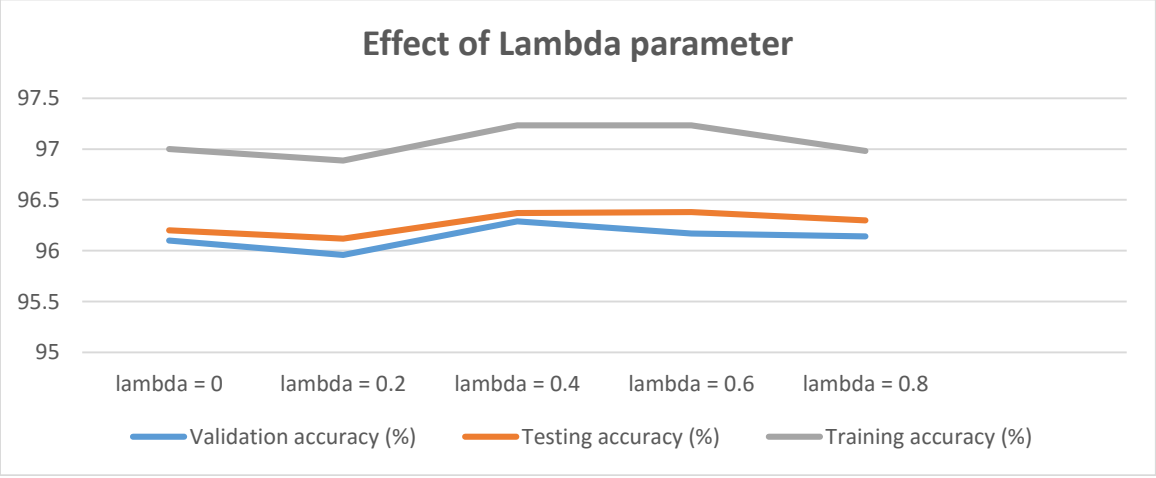
- Include Feed Forward, Back Propagation to implement Neural Network
- Incorporate regularization on the weights
- Use validation set to tune hyper-parameters for Neural Network (number of units in the hidden layer and lambda parameter).

Regularization in Neural Network

In order to avoid overfitting problem (the learning model is best fit with the training data but give poor generalization when test with validation data), we can add a regularization term into our error function to control the magnitude of parameters in Neural Network.

$$\tilde{J}(W^{(1)}, W^{(2)}) = J(W^{(1)}, W^{(2)}) + \frac{\lambda}{2n}\left(\sum_{j=1}^{m}\sum_{i=1}^{d+1}(w_{ji}^{(1)})^2 + \sum_{l=1}^{k}\sum_{j=1}^{m+1}(w_{lj}^{(2)})^2\right)$$

We have tested with different values of lambda ranging from 0 to 1 with increments of 0.2 and observed validation accuracy.



We have kept our hidden units to 20 for finding the best lambda value. With the results we obtained, lambda value 0.4 gives best results.

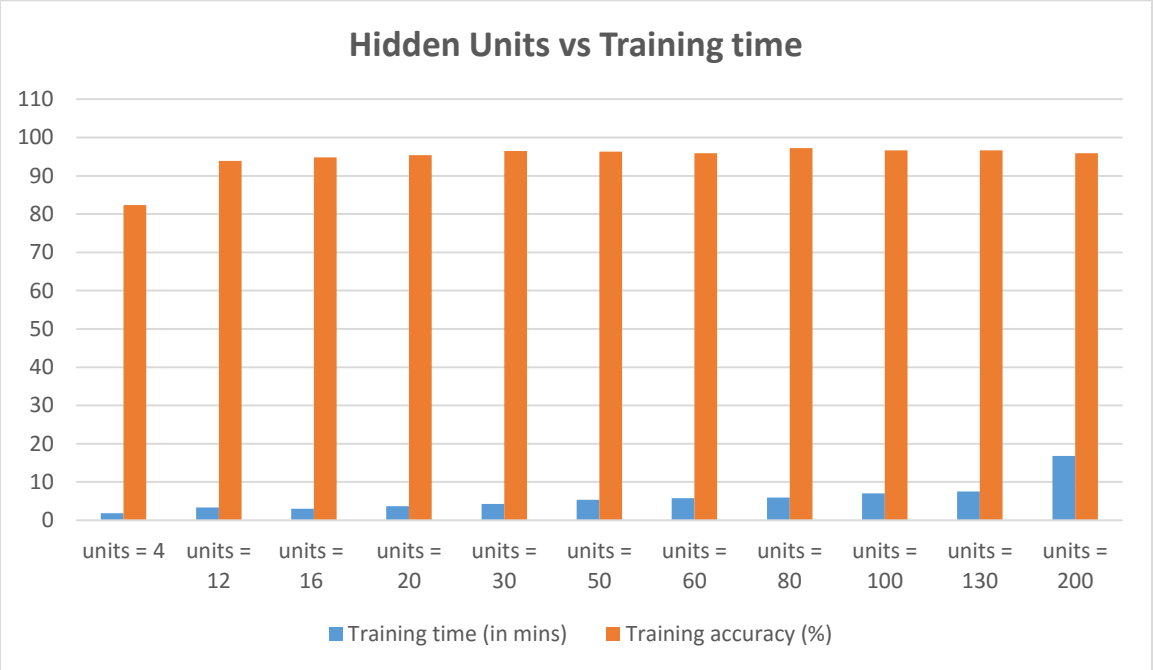
Overfitting vs Underfitting

Overfitting happens when our learning algorithm has learnt to fit the training data perfectly, therefore it predicts with high accuracy on the training example labels (classes). Unfortunately, when new, unseen in the training data, examples are introduced, the learning algorithm might fail to predict correctly their label.

As seen from the graph that with no regularization (lambda=0), our neural network gives high training accuracy but with less validation accuracy. As it can be seen from the formula and results above, lambda value helps in controlling the magnitude of parameters in Neural Network and prevents overfitting.

Performance of Neural network w.r.t Hidden units

We trained our Neural network with different number of hidden units starting with very small e.g. 4 and gradually increasing the number of units.



We observed that increasing hidden unit helps in better training the neural network for the training data but at the marginal cost of training time. Therefore, we should prudently choose number of hidden units which gives enough accuracy with minimal amount of training time.

For the scope of this assignment, optimum values derived for number of hidden units is 80 and lambda value (regularization parameter) is 0.4. For hidden units 80, we receive higher accuracy compared to less number of hidden units with minimal increase in training time. We received **training accuracy of 97.288%, validation accuracy 96.18%,** and **test accuracy of 96.53%** with around 6-7 mins of training time.