CSE 574 Programming Assignment 1 Report----Group 10

Handwritten Digits Classification Using Neural Networks

Chuohao Jiang 37821320; Fengguo Tian 50097353; Zhe Yang 50169180

1. Main purpose

The main purpose is to implement the three layer Handwritten Digits Classification Neural Network (NN) and train the NN with multiple datasets to find the best lambda value and number of hidden units, and finally to reach high accuracy and reasonably decrease running time.

2. Main function explanation

Preprocess(): First separate MINST training data into a training dataset and a validation dataset, and then separate the test data randomly for each training and test operation, finally create the corresponding data labels which means each 7 vector will be represented by a digit.

Sigmoid(): is a logical function and its output ranges between 0 and 1, increasing monotonically with its input, because it maps a very large input domain to a small range of outputs.

nnObjfunction(): This is the key function to calculate the error of each input dataset and update the error gradient with the regularization value, which helps avoid the overtraining and undertraining, we will use the output of this function as the input of the minimize function to train the neural network. Cause the dataset is very big, we use matrix multiplication and avoid the "for" loop, which greatly reduces the runtime.

nnPredict(): for each dataset, this function was used to calculate the predict labels for each input associated with the updating weights. And the results would be used to calculate the obj_val and compare with the target label to calculate the accuracy of the neural network.

For Optimization: We use the Python scipy function: scipy.optimize.minimize (using the option method='CG' for conjugate gradient descent) with the file **params.pickle**, which contains the learned parameters of Neural Network: optimal n hidden (number of units in hidden layer), w1 (matrix of weight W1), w2 (matrix of weight W2), optimal (regularization).

Feature selection: Based on the requirement, we update the input data matrix by removing unchangeable column for each training round to implement the feature selection.

Maxiter: According to the project guide, the preferred value of the maxiter is 50, however, based on the experiment, we find when the maxiter=100, which means the training run it for 100 iterations, we can get a better accuracy (It does not really reach the minima.).

Hyper-parameter:

Number of hidden units: based on the research, the hidden units should be about Squert_root(No._input*No_output), based on our experiment, we find the value can be 40.

Regularization term (lambda_val): based on the experiment the value should be about 0.4.

3. Training process and parameter selection:

3.1 Lambda values and accuracy

The Fig. 1 shows the relationship between the accuracy and lambda value. We find that different lambda values has a weak impact on accuracy, which is very close to each other with different lambda values and sitting between 92% and 96%. Actually for this project, there would be over fitting, the Network is not even convergence, because we only train 100 times.

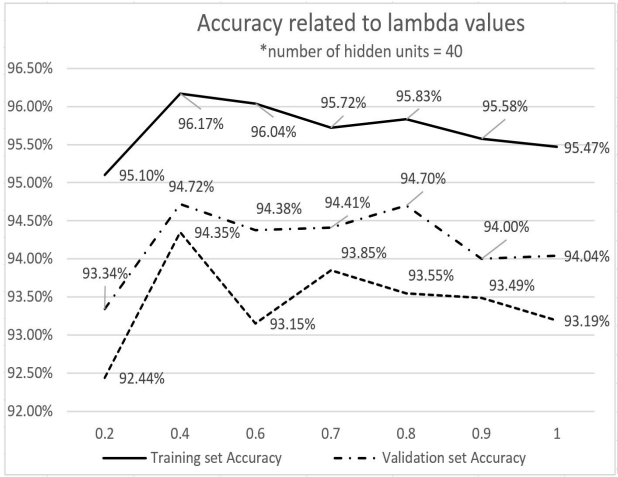

Fig. 1

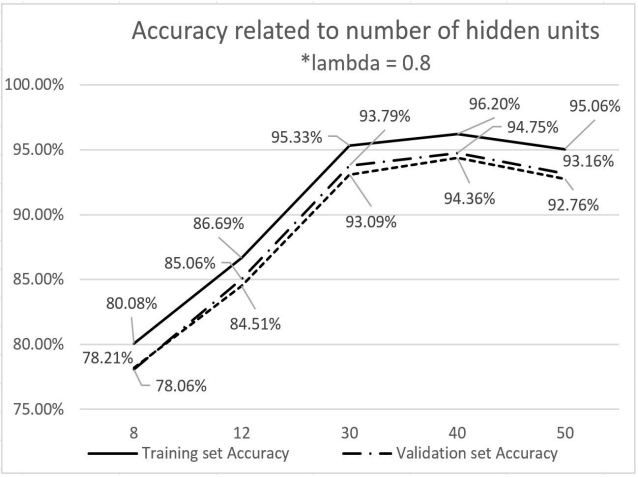

Fig.2

Fig.2 shows the relation between accuracy and number of hidden units, where we can see the number of hidden units and accuracy are positively correlated. Because with more hidden units, the error function can get more information about the useful feature, which impacts the weight of the training samples. From the figure, we can find the optimal **Hyper-parameter** should be **Lambda**=0.8, **hidden_value**=40.

3.2 Hidden Layer and Time consuming

Fig.3 shows that it is obvious that running time increased when more hidden layers were added, where we can see the number of hidden units, time and accuracy are positively correlated. With more hidden units, the accuracy increased significantly from 66% to 96% and cost more time, because with the increasing of hidden units the dimension of the matrix will become very big that would cost lots of time. The numbers of hidden units about 40 will give good accuracy and cost 212.6 Seconds.

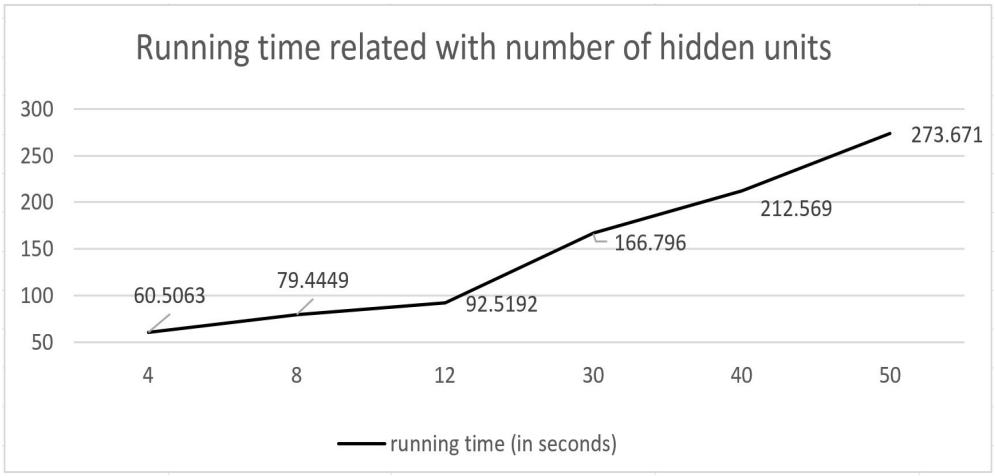

Fig. 3

So overall the parameter of the system as following: **Hyper-parameter** should be **Lambda**=0.8, **hidden_value**=40, maxiter=100, and the training set accuracy is 96.20%, validation set accuracy is 94.75%, and test set accuracy is 94.36%.