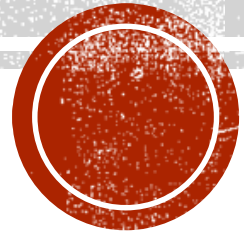
MNIST TRAINING WITH MLP

Chronis Sakos

190744335

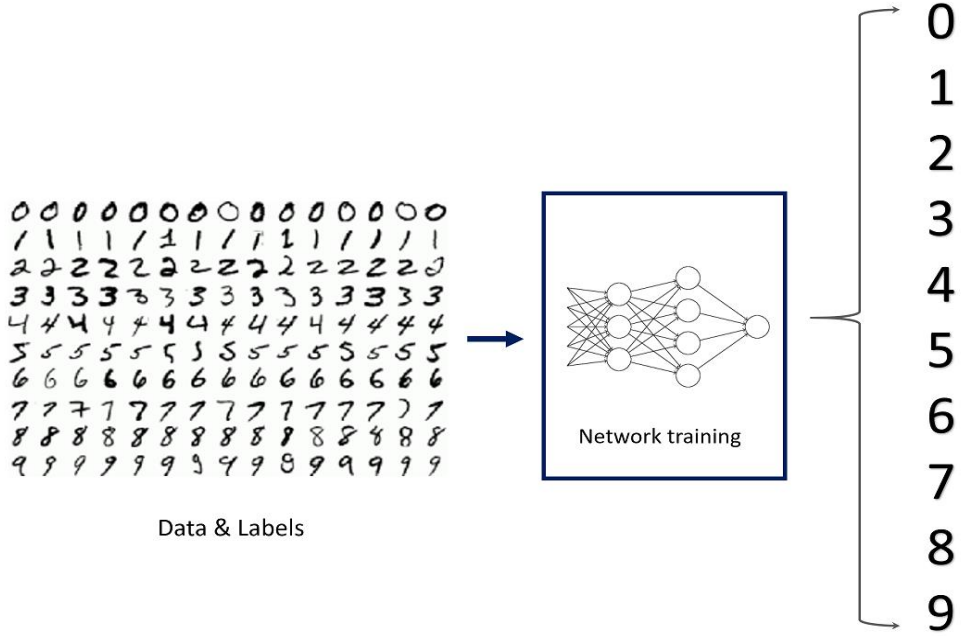


MNIST MLP

The concept of this presentation is to mention how the different methods affect the performance of a MLP network that trains MNIST datasets. The goal is to minimize the error of training and maximize the accuracy while we keep the cost and the training time viable.

Context

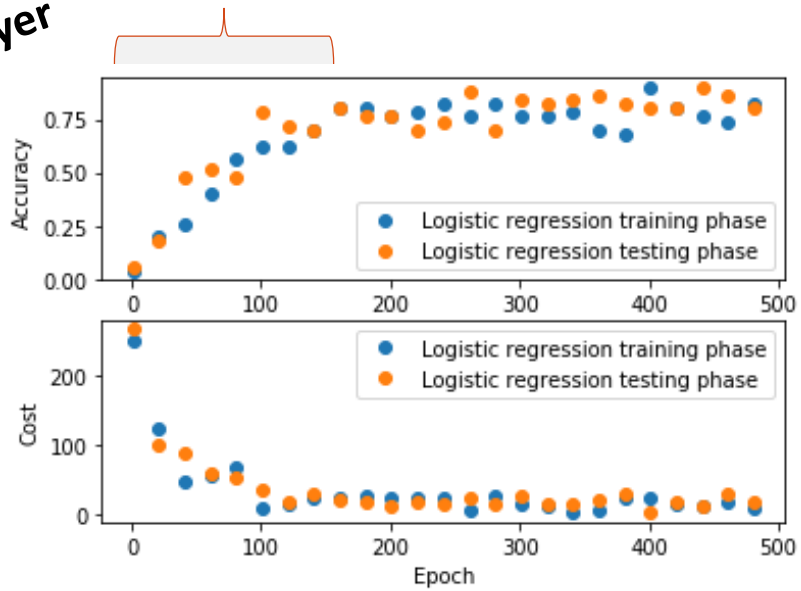
- 1 vs 3 hidden layers MLP
- How learning rate affects performance?
- Batch Training
- Dropout keep probability evaluation
- Summarize



Queen Mary University of London

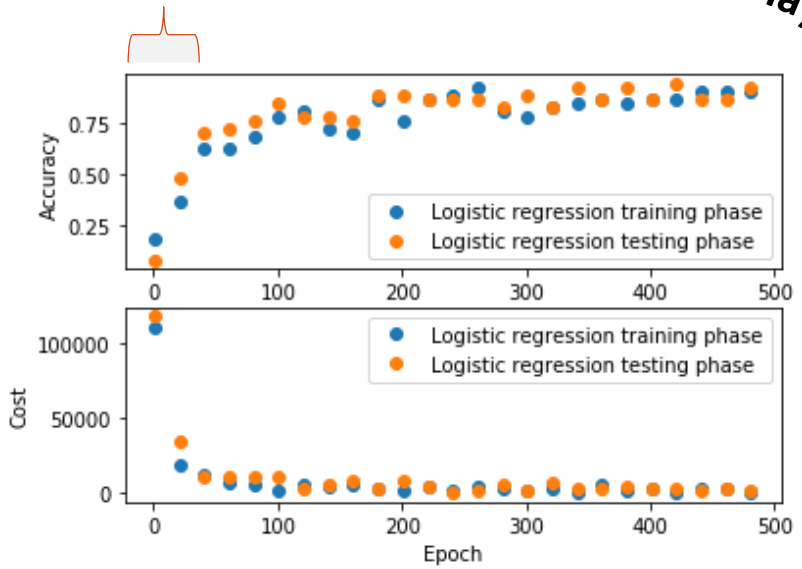
1 VS 3 HIDDEN LAYERS

1 layer



training_epochs = 500
DropOutKeepProb = 0.8
learning_rate = 0.001

3 layers



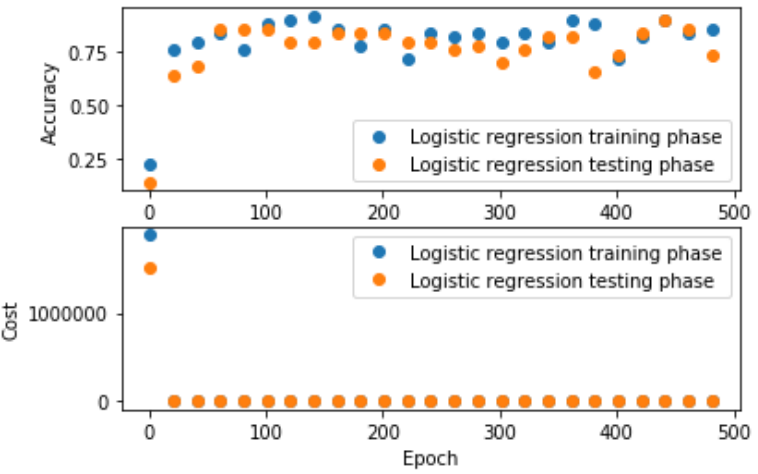

```
accuracy for the train set examples     =  0.82
accuracy for the test set examples      =  0.8
```

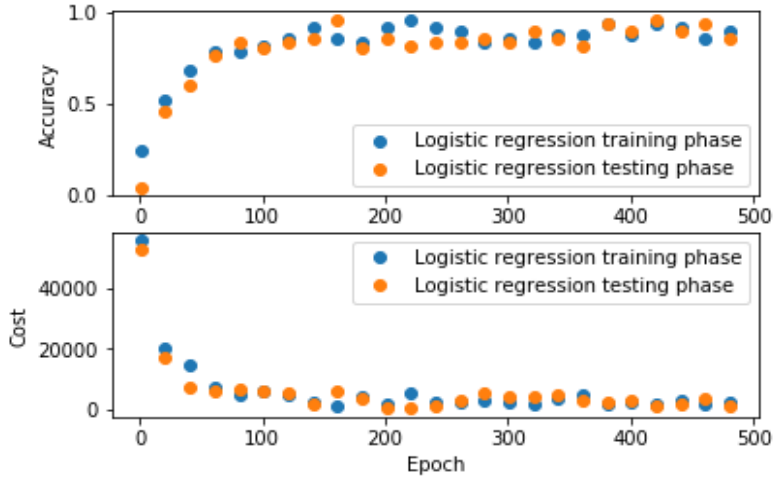
```
accuracy for the train set examples     =  0.9
accuracy for the test set examples      =  0.92
```

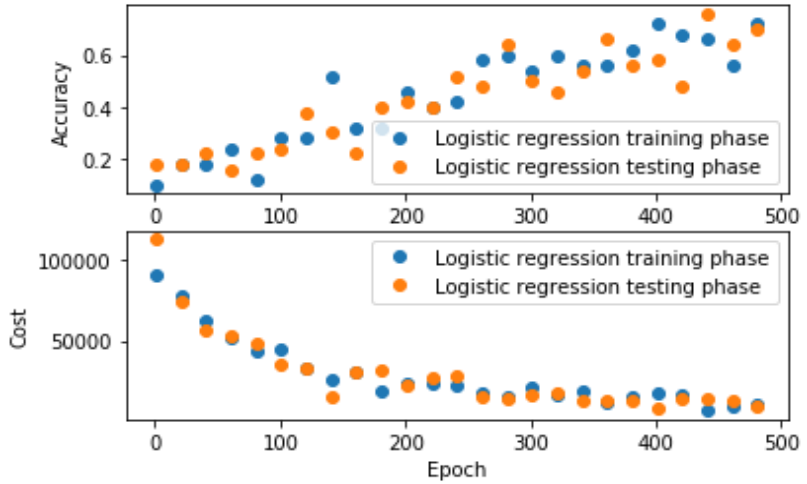
- The multilayer perceptron needs less training epochs to overcome the accuracy percentage of 75% .
- For the same number of epochs i.e. 500 epochs, the accuracy of the MLP with 3 hidden layers is ~10% higher than the accuracy of the 1 layer perceptron.

Queen Mary University of London

LEARNING RATE ⟷ PERFORMANCE







*10 0.01 /100

```
accuracy for the train set examples     =  0.86
accuracy for the test set examples      =  0.74
```

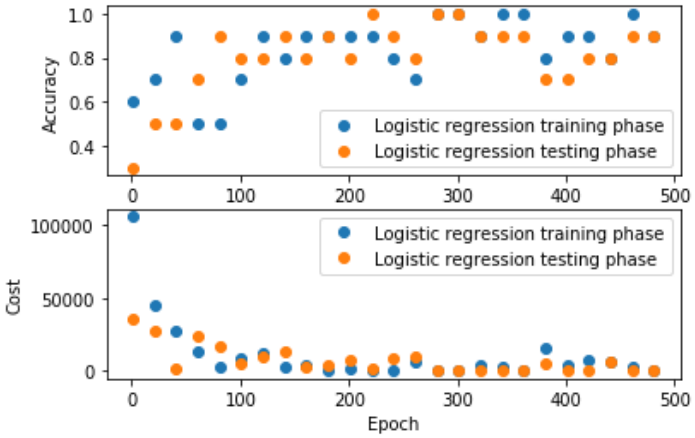
```
accuracy for the train set examples     =  0.9
accuracy for the test set examples      =  0.86
```

```
accuracy for the train set examples     =  0.72
accuracy for the test set examples      =  0.7
```

Learning rate is a hyper-parameter that controls how quickly the model is adapted to the problem. We can conclude that, as we keep lowering the learning rate, the accuracy tends to decrease as well. Therefore, a value at **0.01** performs best.

Queen Mary University of London

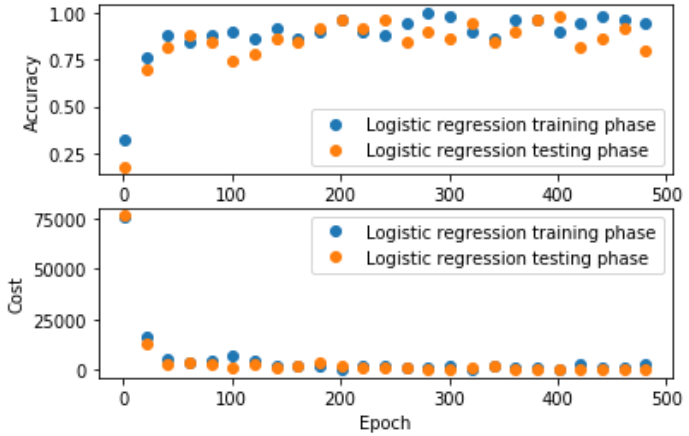
BATCH TRAINING



10 batches

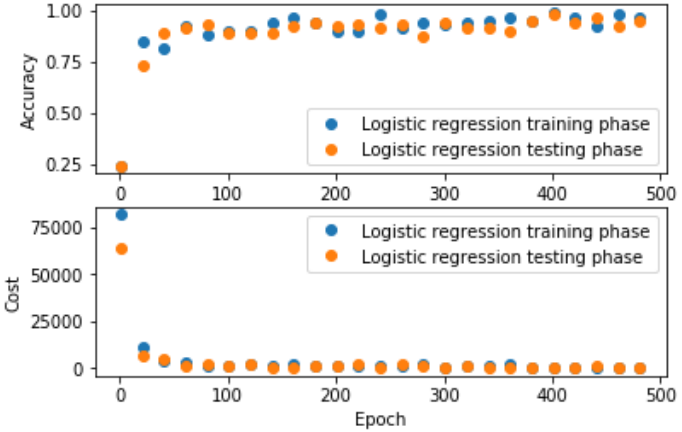
accuracy for the train set examples = 0.9
accuracy for the test set examples = 0.9

50 batches



accuracy for the train set examples = 0.94
accuracy for the test set examples = 0.8

100 batches



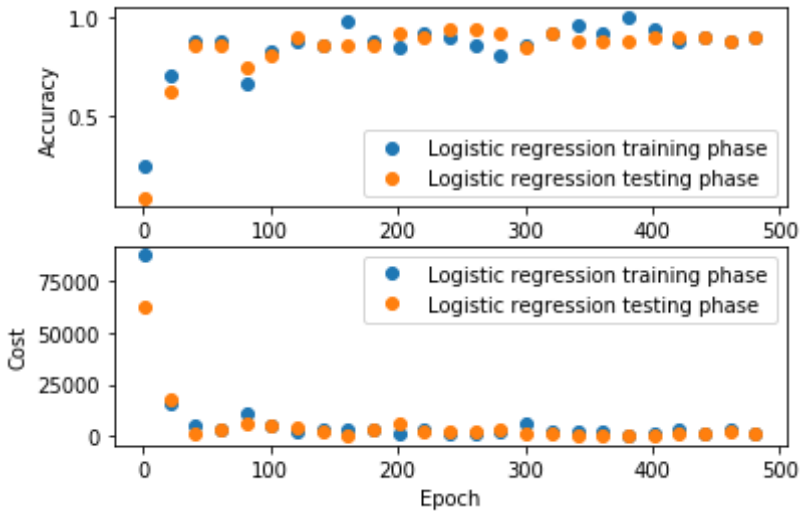
accuracy for the train set examples = 0.96
accuracy for the test set examples = 0.95

We can notice that the higher the number of batches is, namely the smaller the batch size, the closer we get to 100% of accuracy for the training set.

500 batches

accuracy for the train set examples = 0.994
accuracy for the test set examples = 0.946

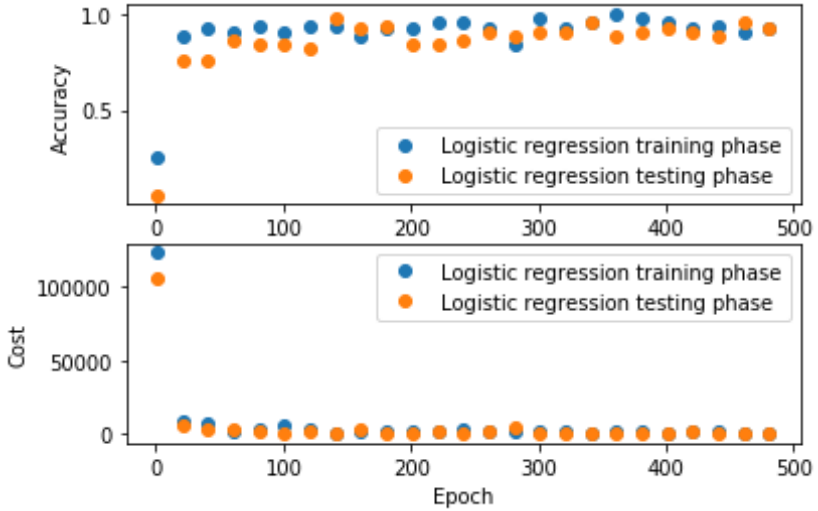
DROPOUT



40%

```
accuracy for the train set examples      =  0.9
accuracy for the test set examples       =  0.9
```

It is obvious that there is no significant difference in our network if we change the dropout keep probability. This happens because the number of training epochs is low, so no overtraining affects the performance. Also, our network is small in relation to the dataset.




```
accuracy for the train set examples      =  0.92
accuracy for the test set examples       =  0.92
```

80%

Queen Mary University of London

SUMMARY

- A multilayer perceptron network with 3 hidden layers is far more accurate than a network with 1 hidden layer, if we compare them for a specific number of training epochs.

- The learning rate affects massively the performance of a MLP network.
  - The smaller the learning rate is, the more time is required to train the network. Smaller rate usually needs more training epochs for the network to become accurate.
  - Less training time, less cost.

- Dividing the dataset into more and smaller batches, makes the accuracy of the training and testing set samples bigger and reaches closer to the 100%.
- The higher the number of batches is, the less epochs are required to reach a satisfying value for accuracy.

- Any change to the Dropout keep probability hardly affects the performance of the network. Any overtraining is eliminated because the network is smaller than the dataset.