Full name: Thi Ngoc Anh Tran
Student ID: 9102520232

Report Assignment 2

a) Make a MLP with 2 hidden layers (each layer with 256 units) and its training/validation/evaluation code to perform multiclass classification over all digits.
-   Please go to this Link

b) Train for 5 epochs with batch size 64.
-    This is the result after training for 5 epochs with batch size = 64:

With epoch 0: loss: 2.299855 [64/60000] loss: 2.294864 [12864/60000] loss: 2.290794 [25664/60000] loss: 2.288573 [38464/60000] Train: Accuracy: 19.1%, Avg loss: 2.291189 Validation: Accuracy: 28.4%, Avg loss: 2.278475	With epoch 1: loss: 2.278586 [64/60000] loss: 2.273299 [12864/60000] loss: 2.267096 [25664/60000] loss: 2.249541 [38464/60000] Train: Accuracy: 40.1%, Avg loss: 2.263770 Validation: Accuracy: 51.5%, Avg loss: 2.24775
With epoch 2: loss: 2.245789 [64/60000] loss: 2.229317 [12864/60000] loss: 2.223746 [25664/60000] loss: 2.210133 [38464/60000] Train: Accuracy: 57.6%, Avg loss: 2.226064 Validation: Accuracy: 60.2%, Avg loss: 2.202433	With epoch 3: loss: 2.205108 [64/60000] loss: 2.196153 [12864/60000] loss: 2.169593 [25664/60000] loss: 2.136563 [38464/60000] Train: Accuracy: 62.4%, Avg loss: 2.167951 Validation: Accuracy: 63.1%, Avg loss: 2.130547
With epoch 4: loss: 2.156175 [64/60000] loss: 2.115165 [12864/60000] loss: 2.086789 [25664/60000] loss: 2.078814 [38464/60000] Train: Accuracy: 64.5%, Avg loss: 2.075427 Validation: Accuracy: 64.3%, Avg loss: 2.016472	**Test: Accuracy: 64.8%,** **Avg loss: 2.007726**

c) How should we choose the number of iterations to achieve good generalization? Train until you think the model has achieved good generalization.

- We should choose the number of iterations when the loss of validation stops decreasing even though the training loss is still decreasing. The model becomes overfitting after that.

-
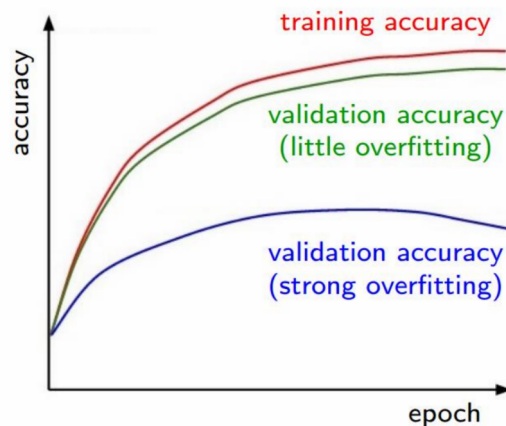
d) What are the hyperparameters you can tune? List all of them.

These are some hyperparameters that I can tune:
- Number of Units per Hidden Layer: 128, 256,...
- Learning Rate: 0.01, 0.001,...
- Batch Size: 32, 64,...
- Number of Epochs: 3, 5, 10,...
- Optimizer: Adam, Adagrad, SGD,...

e) Use the hyperparameter tuning strategy (for at least two iterations) to find a set of hyperparameters (for at least two hyperparameters) that would provide good generalization performance. Report the procedure you took in detail.

➢ I will change two hyperparameters: the learning rate and the number of epochs
➢ I choose learning rate = 0.01, 0.005, 0.001
➢ I choose the number of epochs = 5, 10, 20

Learning rate	Number of epochs	Last validation accuracy	Last test accuracy
0.01	5	Validation: Accuracy: 90.8%, Avg loss: 0.328752	Test: Accuracy: 91.1%, Avg loss: 0.305645
	10	Validation: Accuracy: 92.7%, Avg loss: 0.255329	Test: Accuracy: 93.3%, Avg loss: 0.238912
	20	Validation: Accuracy: 94.8%, Avg loss: 0.176305	Test: Accuracy: 95.4%, Avg loss: 0.160310
0.001	5	Validation: Accuracy:	Test: Accuracy:

		64.3%, Avg loss: 2.016472	64.8%, Avg loss: 2.007726
	10	Validation: Accuracy: 78.9%, Avg loss: 0.961431	Test: Accuracy: 79.6%, Avg loss: 0.933880
	20	Validation: Accuracy: 86.7%, Avg loss: 0.490305	Test: Accuracy: 87.3%, Avg loss: 0.465445
0.005	5	Validation: Accuracy: 88.1%, Avg loss: 0.426373	Test: Accuracy: 89.0%, Avg loss: 0.400292
	10	Validation: Accuracy: 90.6%, Avg loss: 0.327362	Test: Accuracy: 91.2%, Avg loss: 0.305606
	20	Validation: Accuracy: 92.5%, Avg loss: 0.258905	Test: Accuracy: 93.2%, Avg loss: 0.241866

f) What is the final test accuracy?
- Test: Accuracy: 97.5%, Avg loss: 0.085818 (with learning rate = 0.01, number of epochs = 50)