# Report of HW3 - Neural Network

Student ID: 109062211 Name: 張惇媛

I. Describe what problems you encountered and how did you solve them when implementing the basic and advanced functions.

- The local minimum for gradient descent
  The gradient descent with fixed learning_rate may get stuck in the local minimum, but we want to find global minimum.
  Solution: change the learning_rate when encountering the local minimum.
- The tradeoff between cost and accuracy
  Although our aim is to lower the cost in gradient descent, the accuracy would not be as high as expectation because of over-fitting.
  Solution: do early stopping to avoid over-fitting.

II. Briefly describe the structure of your binary and multi-class classifiers.

- Binary-class classifier
  1. Use min-max scaling to normalize X
  2. Use batch gradient descent to train the model
  3. Use 9 hidden-layer (below is the detail)

```
layers_dims = [30, 11, 31, 23, 23, 25, 10, 5, 7, 3, 1]
activation_fn = ["relu", "relu", "relu", "relu", "relu", "relu", "relu", "relu", "relu", "sigmoid"]
```

- Multi-class classifier
  1. For X, do flatten and each is divided by 255 to normalize; for y, do one-hot encoding changed into 10 classes.
  2. Use mini-batch gradient descent to train the model
     - batch_size = 64
  3. Use 4 hidden-layer (below is the detail)

```
layers_dims = [784, 196, 49, 16, 33, 10]
activation_fn = ["relu", "relu", "relu", "relu", "softmax"]
```

III. Describe effort you put to improve your model (e.g., hyperparameter finetuning).

To improve my model, I put effort on the following several things:

First, since I tried for the number of nodes for each hidden-layer in descending order, the performance is not quite good. As a result, for some of the hidden-layers, I increase or remain the same number of nodes.

Second, while tracking the cost, I finetune the learning_rate to avoid only finding the local minimum.

Third, I add a helper function to track both the cost and the accuracy to finetune the num_iterations for early stopping.

Forth, when checking the overall performance of my model in the advanced part implementation, I use cross-validation (separate dataset into 10 partitions) to do model evaluation.