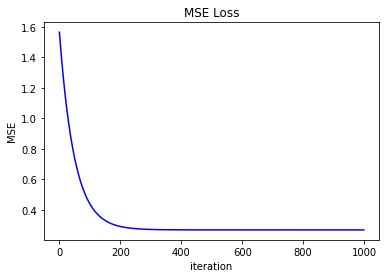
0819823 陳子祈 Image recognition HW1

Part 1

1. MSE loss



2. Final testing data MSE = 0.068705834617267

3. weight = 0.8179331278291383

intercept = 0.7845389075277833

4.

(1) **Gradient Descent**: Send all the data into the network at one time. Although lot of data makes the gradient easy to find the general direction, but when the amount of data is large, each round of training will take a lot of time

(2) **Mini-Batch Gradient Descent**: Each time you input m data to train the model, the gradient of this method is easier to find the roughly correct direction than the gradient of SGD, and because the data size of each round of training is m, it is easier for the computer to process this vectorized data

(3) **Stochastic Gradient Descent**: Input one data at a time to train the model. Although each round of training in this method is fast, it is easy for the gradient to find the correct direction, and the chaotic travel direction may not necessarily help it find the lowest point faster.

Part2

(next page)