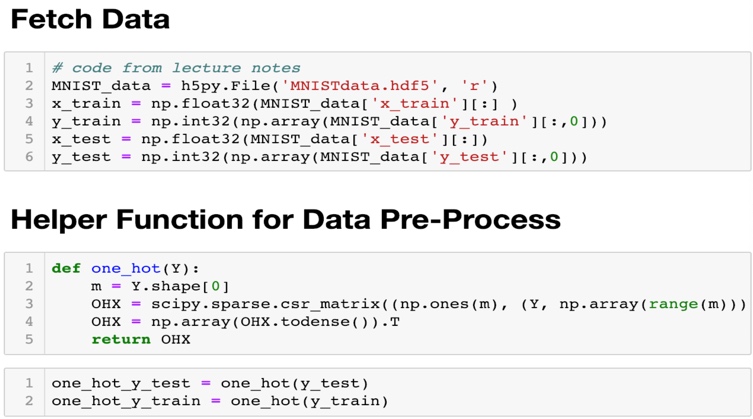
**CS398-Deep Learning**

**Homework 1**

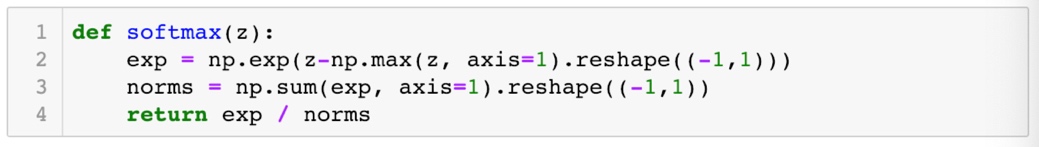
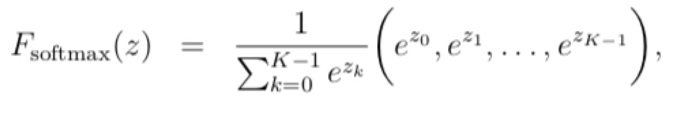
**Lingyi Xu (lingyix2)**

1. **Description of Implementation**

First, I load the MNIST data from the dataset and preprocess the labels to one-hot format.

Then I build my mini-batch softmax logistic regression model.

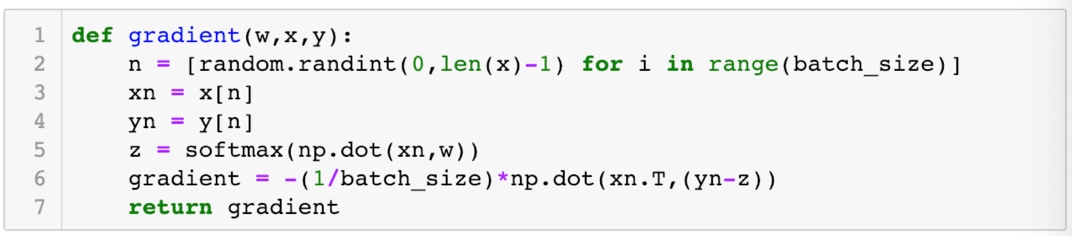
I define the softmax function I will use to classify the probabilities:



And the gradient I will use for update W:

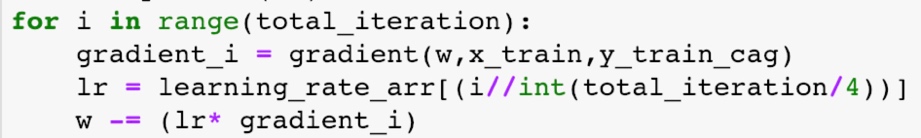
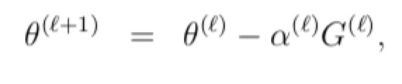
When training, I first initialize W, a zero matrix with shape (784,10). Then I use 2800 iterations to update the W by gradient decent.

For each iteration, I calculate the gradient of F(w) from the sampling:



After calculating the gradient of the function, we can update the parameter matrix w by

(where “theta” is “W” in my implementation)



For this particular dataset, I use the following parameters:

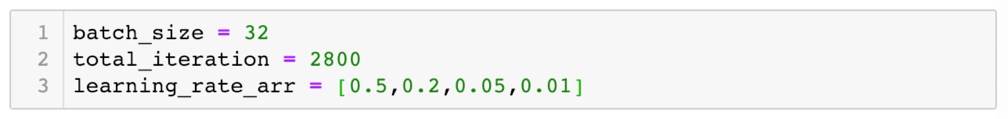
Iteration = 2800

Batch size = 32

Learning rate = 0.5 when 0<=itr<700;

0.2 when 700<=itr<1400;

0.05 when 1400<=itr<2100;

0.01 when 2100<=itr<2800

1. **Final Test Accuracy:**

91.9%

****