# OASIS ML Group TRAINING 02

## ◆ Logistic Regression

You are given a dataset of **handwritten character digits (EMNIST.zip)** derived from the EMNIST dataset. This dataset contains **8 classes with 128 different images** in each class. Supervised learning is performed for training data. In this exercise, you need to implement a **logistic regression model for classification.**

※**Note**：You need to normalize the data samples before training and randomly select 32 images as test data for each class and the remaining images as training data.

### ▫ Classification Problem

#### ■ Analysis datasets：

EMNIST datasets contain eight classes：「a」,「b」….. ,「h」. Each class has 128 data, which means there're total 1024 data in the datasets, and the pixel of data is 28*28. You should use a 1-of-K binary coding scheme (one-hot encoding) for the target vector t. Implement the logistic regression model using **batch GD** *(batch gradient descent)*, **SGD** *(stochastic gradient descent)* and **mini-batch SGD** with **softmax** activation.

| Algorithms | Batch size | No. of iterations in each epoch |
|---|---|---|
| batch GD | $N$ | 1 |
| SGD | 1 | $N$ |
| mini-batch SGD | $B$ | $N/B$ |

$N$ is the number of training data. $B$ is the batch size.

Set the initial weight vector

$$\boldsymbol{w}_k = [w_{k1}, \dots\dots, \quad w_{kF}], 0 \le k \le K$$

to be a zero vector where F is the number of features and K is the number of classes.

$$\boldsymbol{t_0} = \boldsymbol{w_{00}} + \boldsymbol{w_{01}x_1} + \boldsymbol{w_{02}x_2} \dots\dots + \boldsymbol{w_{0F}x_F}$$
$$\boldsymbol{t_1} = \boldsymbol{w_{10}} + \boldsymbol{w_{11}x_1} + \boldsymbol{w_{12}x_2} \dots\dots + \boldsymbol{w_{1F}x_F}$$
$$\cdot$$
$$\cdot$$
$$\boldsymbol{t_K} = \boldsymbol{w_{K0}} + \boldsymbol{w_{K1}x_1} + \boldsymbol{w_{K2}x_2} \dots\dots + \boldsymbol{w_{KF}x_F}$$

#### ■ Loss function：

- Using Cross entropy function as loss function.
  Discuss what is cross entropy function?

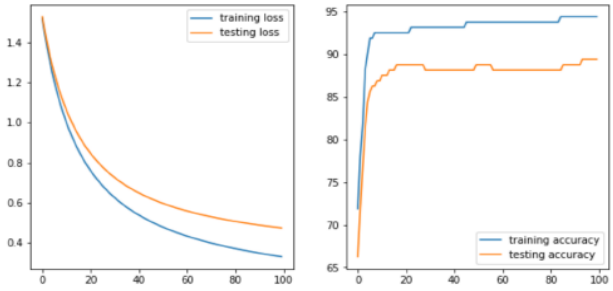
$$E(w) = \frac{-1}{M}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \log y_{nk}$$

Where N is the number of training data, K is the number of classes.

## ■ Initial parameter：

Epoch, learning rate：try by yourself

(a) Plot the learning curves of the loss function and the accuracy of classification versus the number of epochs until convergence for training data as well as test data, e.g.




(b) Show the final classification accuracy and loss value of training and testing data.

(c) Based on your observation about different algorithms *(batch GD, SGD and mini-batch SGD)*, please makes some discussions.

## ▫ Bonus problem

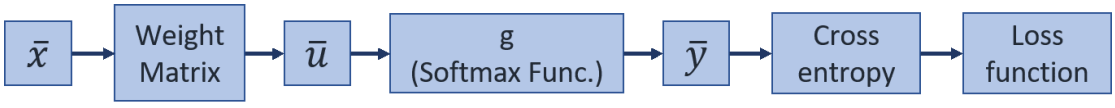



Figure shows the relationship of weight matrix, softmax and loss function.

- **Please derive：**

**When** $err = pred\_y - label\_y$,

$$\frac{\partial L}{\partial W_{ij}} = x_j \cdot err_i$$

**or**

$$\frac{\partial L}{\partial W_{ij}} = \frac{1}{M}\sum_{m=0}^{M-1} x_{j(m)} \cdot err_{i(m)}$$

- **Constraints**

  You can only use **numpy, pandas, matplotlib, PIL**, **cv2** and **Python standard libraries** in this lab, you can choose one of PIL and cv2 packages for image reading.