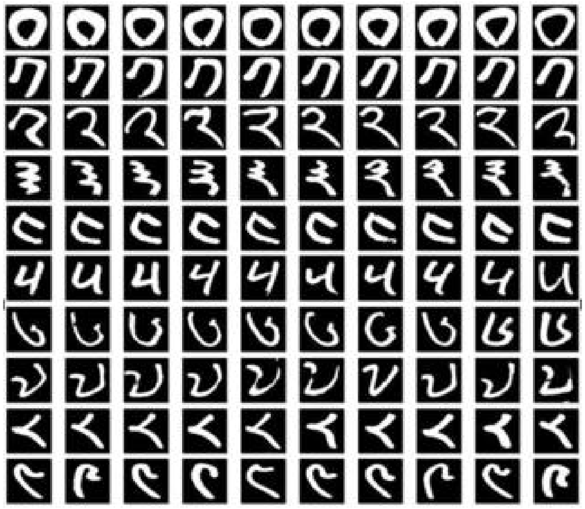
Deep Learning (Homework 1)

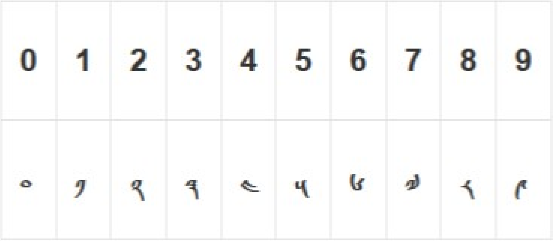
Due date : 4/17/2020

- For problem 1, any tools with automatic differentiation are forbidden, such as Tensorflow, Pytorch, Keras, etc. You should implement backpropagation algorithm by yourself.
- For problem 2, high-level API are forbidden, such as Keras, slim, TFLearn, etc. You should implement the forward computation by yourself.
(Only this problem you can use **PyTorch** or **Tensorflow**).
- Homework submission – Please zip each of your source code and report into a single compressed file and name the file using this format : HW1_StudentID_StudentName.zip (rar, 7z, tar.gz, . . . etc are *not* acceptable)

1 Deep Neural Network for Classification

In this exercise, please implement a Deep Neural Network (DNN) model by yourself to recognize Tibetan handwriting numerals. The following table shows the corresponding Arabic numerals with respect to Tibetan numerals. The samples of this dataset can be found in the following figure. The details of this dataset can be referred to: https://github.com/bat67/TibetanMNIST. Please use train.npz as training data and test.npz as test data. **(Download page)**





0	1	2	3	4	5	6	7	8	9
༠	༡	༢	༣	༤	༥	༦	༧	༨	༩

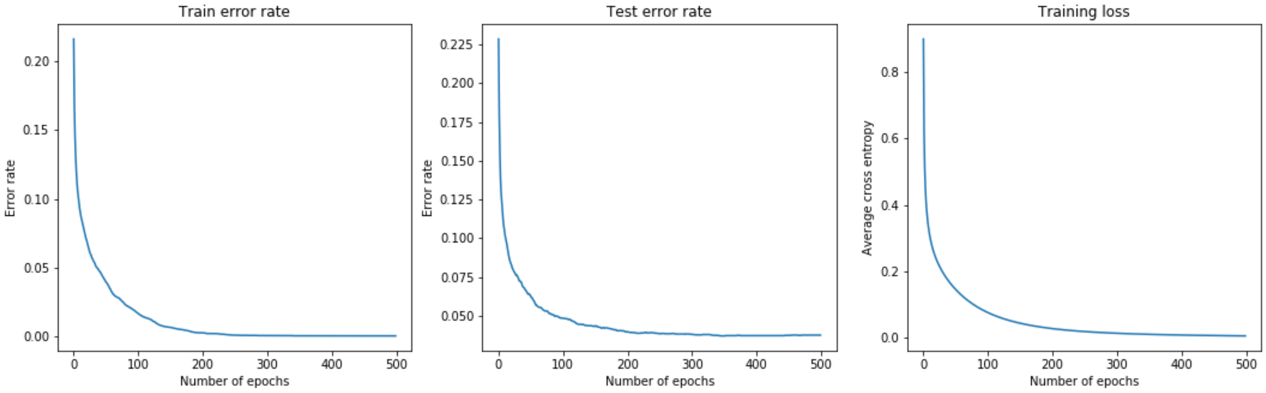
i. Please construct a DNN for classification. For N samples and K categories, the cross-entropy objective function is expressed by

$$E(\mathbf{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \log y_k(\mathbf{x}_n, \mathbf{w})$$

Please minimize the objective function $E(\mathrm{w})$ by running the error backpropagation algorithm using the Stochastic Gradient Descent (SGD)

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E(\mathbf{w}^{(\tau)})$$

You should decide the following hyperparameters: number of hidden layers, number of hidden units, learning rate, number of iterations and mini-batch size. You have to show your (a) learning curve, (b) training error rate and (c) test error rate in the report. You should design the network architecture by yourself.

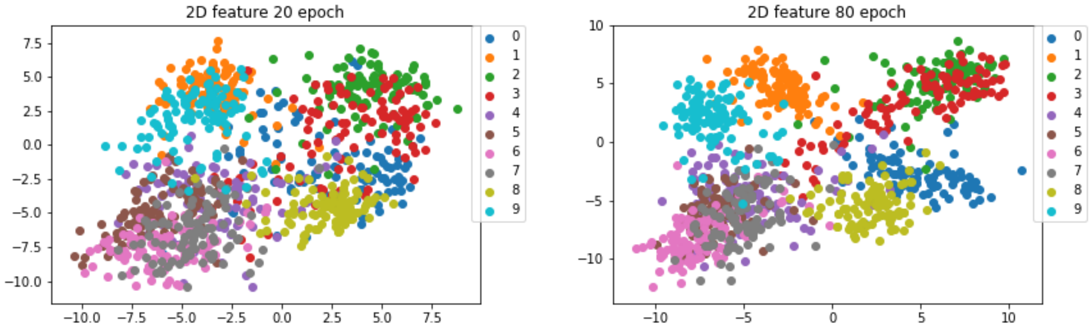


ii. Please perform zero and random initializations for the model weights and compare the corresponding error rates. Are there any difference between two initializations? Please discuss in the report.

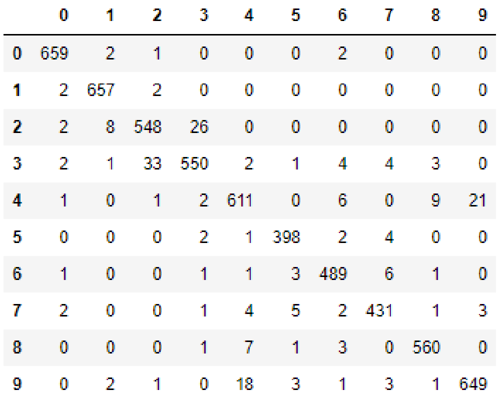
iii. Design your network architecture with the layer of 2 nodes before the output layer.

(1) Plot the distributions of latent features at different training stages. For example, you may show the results when running at 20th and 80th learning epochs.

(2) Please discuss the evolution of latent features at different training stage.



iv. Please list your confusion matrix and discuss about your results.



	0	1	2	3	4	5	6	7	8	9
0	659	2	1	0	0	0	2	0	0	0
1	2	657	2	0	0	0	0	0	0	0
2	2	8	548	26	0	0	0	0	0	0
3	2	1	33	550	2	1	4	4	3	0
4	1	0	1	2	611	0	6	0	9	21
5	0	0	0	2	1	398	2	4	0	0
6	1	0	0	1	1	3	489	6	1	0
7	2	0	0	1	4	5	2	431	1	3
8	0	0	0	1	7	1	3	0	560	0
9	0	2	1	0	18	3	1	3	1	649

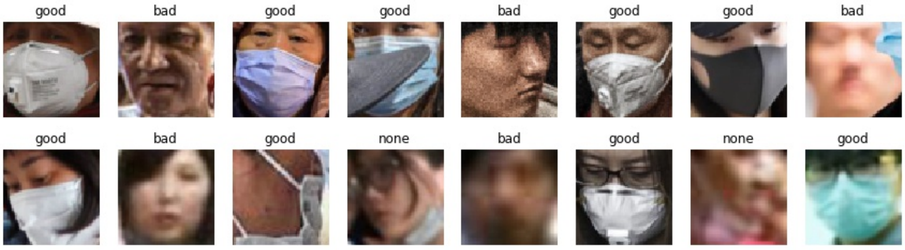
2 Convolutional Neural Network for Image Recognition

In this exercise, you will construct a Convolutional Neural Network (CNN) for image recognition by using **Medical Masks Dataset**. The original dataset comes from **Eden Social Welfare Foundation** which contains the pictures of people wearing medical masks along with the labels containing their descriptions. There are 682 images with over 3000 faces wearing masks and around 700 masks worn either wrongly or not worn at all.

The files train.csv and test.csv contain the corresponding images and their **filename**, **width**, **height**, **label** and **bounding box**. The details of this dataset are shown below:

- number of total pictures : 682
- number of labels by category (train/test):
  - ***good*** (wearing mask) : 3129 (2846/283)
  - ***none*** (wrongly wearing mask) : 126 (104/22)
  - ***bad*** (no wearing mask) : 667 (578/89)
- label format:

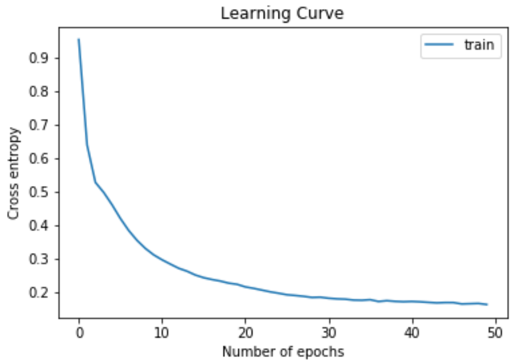
filename	width	height	label	xmin	ymin	xmax	ymax
c1_1844849.jpg	1500	999	good	1246	127	1312	227
c1_1844849.jpg	1500	999	good	1415	144	1486	232
c1_1844849.jpg	1500	999	none	745	889	862	999
stsciRq.png	828	1717	good	249	614	535	914
stsciRq.png	828	1717	good	350	1415	503	1571

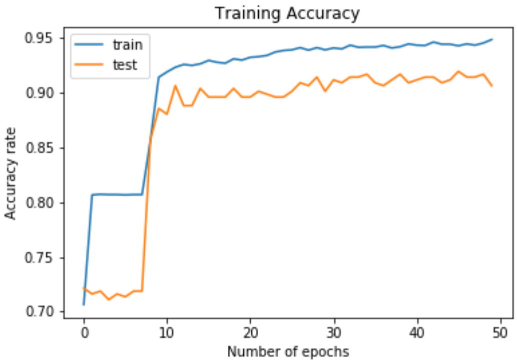


Here is the **Download page**. You should preprocess the images such as cropping through the bounding box or resizing by yourself before implementation.

i. Please describe in details how to preprocess images because of the different resolution images and various bounding boxes region in Medical Masks dataset and explain why. You have to submit your preprocessing code.

ii. Please implement a CNN for image recognition by using Medical Masks dataset. You need to design the network architecture, describe your network architecture and analyze the effect of different settings including stride size and filter size. Plot the learning curve and the accuracy rate of training and test data.



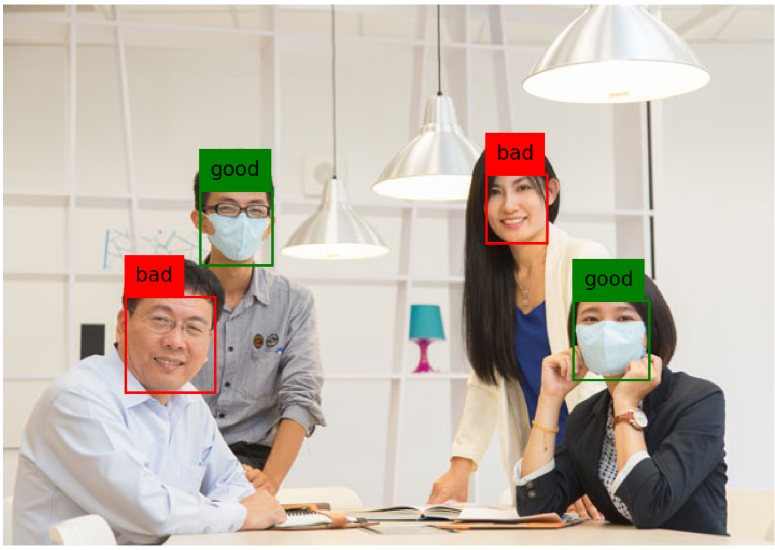


iii. Show some examples of classification result, list your accuracy of each classes for both training and test data, and answer the following questions:

(1) Which class has the worst classification result and why?

(2) How to solve this problem? (explain and do some experiment to compare the result)

(3) Do some discussion about your results.



Class	Train Acc	Test Acc
good	95.2%	94.5%
none	15.3%	14.1%
bad	93.5%	91.8%