**Dataset size:**

There are three datasets for this stage of the project: MNIST, ORL and CIFAR, which are all image datasets.

**MNIST (gray image)**

Training set size: 60,000, testing set size: 10,000, number of classes: 10. Each instance is a 28x28 gray image, and will have one single class label denoted by an integer from {0, 1, …, 9}.

**ORL (gray image)**

Training set size: 360, testing set size: 40, number of classes: 40. Each instance is a 28x28 gray image, and will have one single class label denoted by an integer from {1, 2, …, 39, 40}.

**CIFAR (color image)**

Training set size: 50,000, testing set size: 10,000, number of classes: 10. Each instance is a 32x32 color image, and will have one single label denoted by an integer from {0, 1, 2, …, 9}.

**Dataset organization:**

These dataset are organized as with a dictionary data structure in Python as follows:

{

‘train’: [

{‘image’: a matrix/tensor representing a image, ‘label’: an integer representing the label}

{‘image’: a matrix/tensor representing a image, ‘label’: an integer representing the label}

…

{‘image’: a matrix/tensor representing a image, ‘label’: an integer representing the label}

]

‘test’: [

{‘image’: a matrix/tensor representing a image, ‘label’: an integer representing the label}

{‘image’: a matrix/tensor representing a image, ‘label’: an integer representing the label}

…

{‘image’: a matrix/tensor representing a image, ‘label’: an integer representing the label}

]

}

**Dataset visualization:**

You can load and show the dataset with the following code (it requires the matplotlib toolkit installed in pycharm).

import pickle

import matplotlib.pyplot as plt

if 1:

f = open('MNIST', 'rb') # or change MNIST to other dataset names

data = pickle.load(f)

f.close()

print('training set size:', len(data['train']), 'testing set size:', len(data['test']))

for pair in data['train']:

#for pair in data['test']:

plt.imshow(pair['image'], cmap="Greys")

plt.show()

print(pair['label'])

**Task To Be Done:**

Please train a CNN for these three datasets, respectively, and evaluate its performance on the testing set. Please report your experimental results on all these three datasets.