Machine Learning Lab 2

Throughout the lab sessions, we will continuously do experiments on the MNIST digits dataset. The ultimate goal is to build a handwritten digit recognition system using Python.

Play with MNIST

- 1. For those students who did not finish Lab 1, you should finish Lab 1 first before doing the tasks in this lab.
- 2. To save some time, we should just use the first 1000 images for this lab. To further speed up, you can save the first 1000 data in both X and y using savez from numpy after downloading the data using fetch_openml and load the saved file in the subsequent running using load. For example,

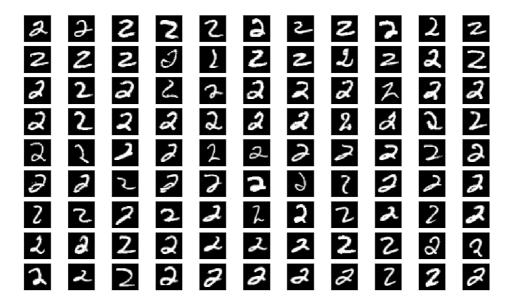
```
np.savez('lk.npz', X_small, y_small)
np.load('lk.npz)
```

Note X_small and y_small should contain the first 1000 data samples.

3. First, count the number of images for each class. Then we should visualize images within the same class, for example, first 99 images of digit 2. You can use the following code to find the index of all images in a specific class:

```
index=np.nonzero(y_small=='2')[0]
```

Then you can plot some of these images using index. Your figure should be similar to the following:



Play around with difference classes.

4. It is sometimes necessary to normalize the data to make them more comparable. The aim of normalization is to make the Euclidean norm

http://en.wikipedia.org/wiki/Euclidean_norm#Euclidean_norm of each digit equal to 1. You can judge by yourself in the following task whether this normalization step is necessary or not.