Machine Learning Lab 4

Instance based learning: KNN

In this lab, you will use instance based learning technique (KNN) to solve the problem of handwritten digit recognition. As for the dataset, we still use the subset of MNIST which we used in our previous labs. Note that using KNN, we can directly do multi-class classification instead of binary classification as using perceptrons. You should finish the following tasks.

1. Implement a function called `myknn` that takes four input arguments. One is the input sample to be classified. The second and third arguments are the feature vector (`tr_feats`) and the labels (`tr_label`). The last argument is the hyper parameter *k*. The output is the label predicted for the input sample:

   ```
   function output_label = myknn(sample, tr_feats, tr_label, k)
   ```

2. Implement cross validation on the training data for selecting the best *k* for the problem. You should use 5 fold cross validation and error rate to evaluate the models and try possible *k*s from 1 to 9. Plot a figure to show the error rates vs *k* (x-axis is k and y-axis is error rate).

3. After selecting the model with the best *k*, you will classify the test images and calculate F1 scores for each digit. Also, compute the confusion matrix to check which pair(s) of digits are most likely to be confused with each other by your KNN classifier.