54345382 ZHAI GUANXUN
DEFAULT PROJECT：DIGIT CLASSIFICATION

November 6, 2015

1. Project Background and Description

Work out 10 binary classifiers for digits from 0-9 with 2000 handwriting digits pictures as training data. And then test how the classifier works on distinguish new hand written digits.

2. Feature vector

*The feature vector here is a vector with 784 greyscale values which represents the corresponding pixel of a 28 * 28 digit picture.*

3. Algorithm

SVM, Naive Bayesian, Logistic Regression.

These algorithm will return a binary result along with the confidence for evaluation.

4. Dataset

MNIST database of hand written digits in matlab version. Totally there are 4000 pictures and half of them are used for training, the other half are for testing.

5. Evaluation

1) *Use cross-validation between the 3 different algorithms: SVM, Naive Bayesian, Logistic Regression.*

2) *Use 1-NN (nearest neighbour) method to evaluate the test result.*

6. Pre-process

Use PCA method to omit the pixels that are not close enough to the digit patterns.