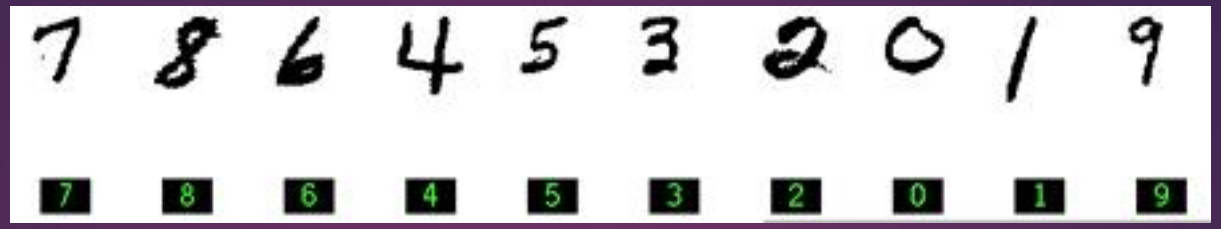

Hand Drawn Digits Recognition

CSCE521-AUTOMATED REASONING

Submitted to:
Dr.Loganantharaj

Prepared by:
Ramesh S(rxs6616)

Outline

- Problem Statement
- Data Source
- Approach
- Evaluation Results
- Conclusion

Problem Statement

Recognize the hand written digits [0-9] which are represented as gray-scaled images of 28*28 resolution present in the test data set using the train data.

Data Source

- This problem contains two Datasets namely Train Data and Test Data.
- Both the data sets are provided by Dr.Logan.

- Training Data Set
  - It contains the pixel data of gray-scaled images of resolution 28*28 i.e., a total of 784 pixels for each image(digit).
  - The training data has total of 785 columns, in which the first column is “label” i.e., the digit drawn by the user and the remaining 784 columns [1-785] contains pixel values of the associated image.

Data Source (contd..)

- Each pixel has a single pixel-value [0-225] associated with it, indicating the lightness or darkness of that pixel.
- Total of 42,000 images.

- Test Data Set
  - It is same as Train Data except it doesn't contain label.
  - Total of 28,000 images.

Approach

- Used Weka-api to train the chosen classifier and classify the test data instances.
- Performed 10-fold cross validation on the train data to evaluate the classifier.
- Developed a Stand-alone java application.
- Source Code:
  - https://github.com/rameshsunkara/HandWrittenRecognizer

Evaluation Results

- Support Vector Machine
  - Correctly Classified Instances 38850 [92.5 %]
  - Incorrectly Classified Instances 3150 [7.5 %]

- Naïve Bayes
  - Correctly Classified Instances 29162 [69.4 %]
  - Incorrectly Classified Instances 12838 [30.5 %]

- J48 (Tree)
  - Correctly Classified Instances 36704 [87.39 %]
  - Incorrectly Classified Instances 5296 [12.61 %]

→ Total Number of Instances 42000

THANK YOU