CORNELL TECH

CS5785 Applied Machine Learning - Fall 2017

Homework 1

Team:	Navdeep Singh
	Vicente Rotman
Professor:	Sergie Belongie
Date:	18-10-2015

1 DigitRecognizer

After downloaded the data from Digit Recognizer competition on Kaggle. Start working on the python code in the file named **digit_recognizer.py**. The function called **main** open the file, it could be the test or train data and store all the data in 2 python lists. One called **labels_test_data** that has the labels of each digit and another called **test_data** that stores the digits in the same order as the labels.

1.1 b) Write a function to display an MNIST digit. Display one of each digit

For this part there are 2 functions involved. The first one called **getOneSampleOfEachDigit(labels, data)** which returns a list with 10 digits, one of each option from 0 - 9. Then the other one is called **displayDigit(image)** this one receive the digit and plots it. Below there is the sample plot of each digit.
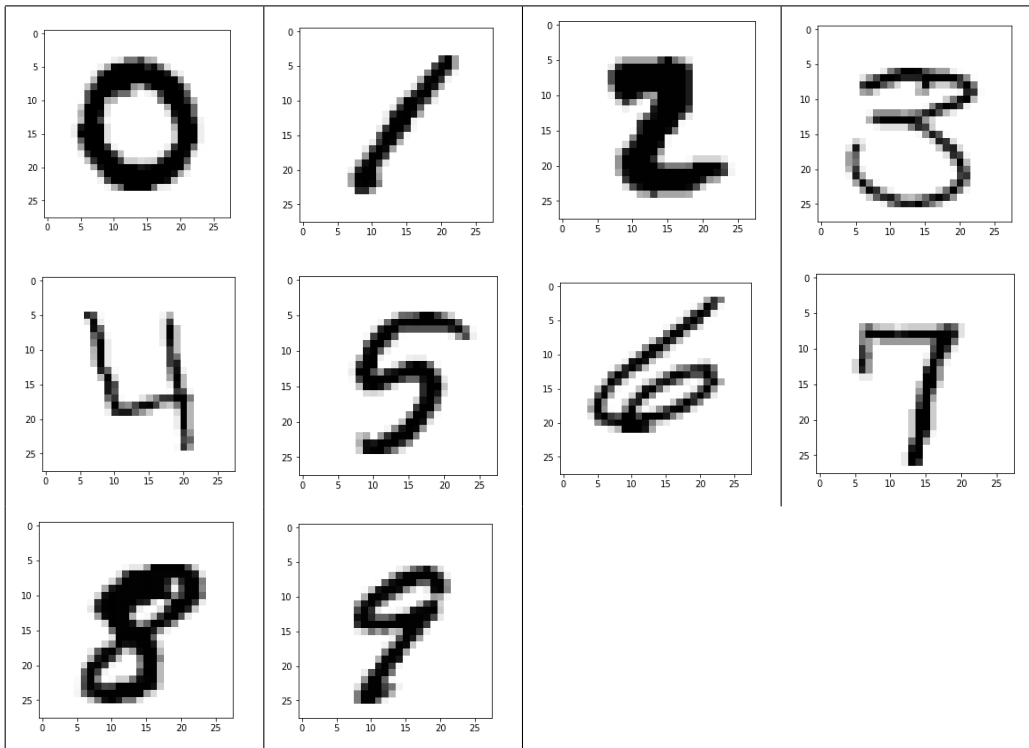
1.2 c) Examine the prior probability of the classes in the training data. Is it uniform across the digits? Display a normalized histogram of digit counts. Is it even?

To make the histogram, the only thing to do is to run a function called **displayHistogram(labels)** Looking at the histogram below we can see that the distribution between the digits in enough uniform. We can say that is even.
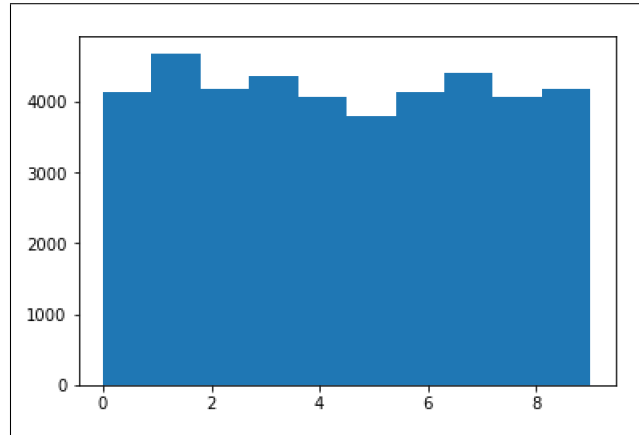

Figure 1: Histogram of digit count

1.3 d) Pick one example of each digit from your training data. Then, for each sample digit, compute and show the best match (nearest neighbor) between your chosen sample and the rest of the training data. Use L2 distance between the two images' pixel values as the metric. This probably won't be perfect, so add an asterisk next to the erroneous examples (if any).

3

2 Titulo 2