

Name: Fatema Hesham ID: 20190373

Mnist dataset:

The MNIST dataset is an acronym that stands for the Modified National Institute of Standards and Technology dataset. It is a dataset of 60,000 small square 28×28 pixel grayscale images of handwritten single digits between 0 and 9.

Features Extraction:

- 1- At first divide the image to 16 blocks each block has 7x7 pixels so I reshaped the image from 28x28 to 16x7x7
- 2- Then I have to calculate the centroid of each block using this formula for x and y

$$x = \frac{\sum f(x,y) * x}{\sum f(x,y)} , y = \frac{\sum f(x,y) * y}{\sum f(x,y)} .$$

KNN MODEL:

- For each testing point we need to calc. all distances that is between testing point and all training points.
- Choose the smallest K distances.
- Classify the testing point according to the major class.
- Build accuracy function and it detect if the predicted label for each point equal to its actual label or not.

- best accuracy at $k=1$ which is 91.2% training sample = 60000 and testing sample = 10000.

$k=1$, accuracy=91.20%

$k=3$, accuracy=89.60%

$k=5$, accuracy=89.70%

$k=7$, accuracy=89.20%

$k=9$, accuracy=88.50%

$k=11$, accuracy=88.70%

$k=13$, accuracy=88.30%

$k=15$, accuracy=88.40%

$k=17$, accuracy=88.00%

$k=19$, accuracy=87.90%

$k=21$, accuracy=87.80%

$k=23$, accuracy=87.70%

$k=25$, accuracy=87.30%

$k=27$, accuracy=86.90%

$k=29$, accuracy=86.80%

END