**Q1 : Khảo sát tập Train với n = 1, 500, 5000, 10000, 59000**

function showrandomTrainImage\_v2(n)

imgTrainAll = loadMNISTImages('train-images.idx3-ubyte');

lblTrainAll = loadMNISTLabels('train-labels.idx1-ubyte')

nNumber = n

figure;

img = imgTrainAll(:,nNumber);

img2D = reshape(img,28,28);

strLabel = num2str(lblTrainAll(nNumber));

strLabel = [strLabel,'(',num2str(nNumber),')'] ;

imshow(img2D);

title(strLabel)

end

Kết Quả :

|  |  |
| --- | --- |
| 1 | 5 |
| 500 | 8 |
| 5000 | 2 |
| 10000 | 7 |
| 59000 | 4 |

**Q2 : Khảo sát tập Test với n = 1, 500, 5000, 9000**

function showrandomTestImage\_v2(n)

imgTestAll = loadMNISTImages('t10k-images.idx3-ubyte')

lblTestAll = loadMNISTLabels('t10k-labels.idx1-ubyte');

nNumber = n;

figure;

img = imgTestAll(:,nNumber);

img2D = reshape(img,28,28);

strLabel = num2str(lblTestAll(nNumber));

strLabel = [strLabel,'(',num2str(nNumber),')'] ;

imshow(img2D);

title(strLabel)

end

Kết Quả :

|  |  |
| --- | --- |
| 1 | 7 |
| 500 | 6 |
| 5000 | 0 |
| 9000 | 0 |

**Q3 : Thống kê số lượng các ảnh tương ứng với các label trong tập huấn luyện (train) của tập dữ liệu MNIST.**

function saveTrain()

i = 0 ;

digitArray = [];

while i < 10

s = 0 ;

digitIndex = [];

for j = 1:60000

if lblTrainAll(j) ==i

digitIndex = [digitIndex,j];

s = s + 1;

end

end

digitArray = [digitArray,s];

fprintf('%d',digitIndex);

strFileName = ['index',num2str(i),'.csv']

csvwrite(strFileName,digitIndex,0,0);

i = i + 1 ;

end

csvwrite('csvlistTrain.csv',digitArray,0,0);

end

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **5923** | **6742** | **5958** | **6131** | **5842** | **5421** | **5918** | **6265** | **5851** | **5949** |

**Q4 : Thống kê số lượng các ảnh tương ứng với các label test của tập dữ liệu MNIST.**

function saveTest()

i = 0 ;

digitArray = [];

while i < 10

s = 0 ;

digitIndex = [];

for j = 1:60000

if lblTestAll(j) ==i

digitIndex = [digitIndex,j];

s = s + 1;

end

end

digitArray = [digitArray,s];

strFileName = ['index',num2str(i),'.csv']

csvwrite(strFileName,digitIndex,0,0);

i = i + 1 ;

end

csvwrite('csvlistTest.csv',digitArray,0,0);

end

Kết quả :

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **980** | **1135** | **1032** | **1010** | **982** | **892** | **958** | **1028** | **974** | **1009** |

**Q5 + Q6 : Kết quả nhận dạng với n = 5, 500, 900 ,hay với n bất kỳ [1, 10000]**

function detectimage(n)

imgTestAll = loadMNISTImages('t10k-images.idx3-ubyte')

lblTestAll = loadMNISTLabels('t10k-labels.idx1-ubyte');

imgTrainAll = loadMNISTImages('train-images.idx3-ubyte');

lblTrainAll = loadMNISTLabels('train-labels.idx1-ubyte');

Mdl = fitcknn(imgTrainAll',lblTrainAll)

index = n;

imgTest = imgTestAll(:,index);

lbPredictTest = predict(Mdl,imgTest');

strLabel = lblTestAll(index);

if strLabel == lbPredictTest

fprintf('Label đúng');

else

fprintf('Label sai');

end

% show result

figure;

img = imgTestAll(:,index );

img2D = reshape(img,28,28);

strLabel = num2str(lblTestAll(index));

strLabel = [strLabel,'(',num2str(index),')'] ;

imshow(img2D);

title(strLabel);

end

|  |  |  |
| --- | --- | --- |
| **n** | **Prediction** | **Actual (test)** |
| **5** | **4** | **4** |
| **500** | **6** | **6** |
| **900** | **8** | **8** |

**Q7 : Đếm số lượng các ảnh có label là n (n là tham số) bị nhận dạng sai theo thuật toán knn**

function countingImage(n)

strName = ['index',num2str(n),'.csv']

dataIndex = csvread(strName);

countNotTrue = 0

countTrue = 0

countindex = size(dataIndex,2);

for c = 1:countindex

indext = dataIndex(c);

imgTest = imgTestAll(:,indext);

lbPredictTest = predict(Mdl,imgTest');

if 1 == lbPredictTest

countTrue = countTrue + 1 ;

else

countNotTrue = countNotTrue + 1 ;

end

end

end

**Kết quả với n = [0,9]**

|  |  |  |
| --- | --- | --- |
|  | **True** | **False** |
| **0** | **973** | **7** |
| **1** | **1129** | **6** |
| **2** | **992** | **40** |
| **3** | **970** | **40** |
| **4** | **944** | **38** |
| **5** | **860** | **32** |
| **6** | **944** | **14** |
| **7** | **992** | **36** |
| **8** | **920** | **54** |
| **9** | **967** | **42** |

**Q8 : Confusion matrix**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **0** |  | **1** | **1** |  |  | **1** | **3** | **1** |  |  |
| **1** |  |  | **3** |  | **1** | **1** | **1** |  |  |  |
| **2** | **7** | **6** |  | **5** | **1** |  | **2** | **16** | **3** |  |
| **3** |  | **1** | **2** |  | **1** | **19** |  | **7** | **7** | **3** |
| **4** |  | **7** |  |  |  |  | **3** | **5** | **1** | **22** |
| **5** | **1** | **1** |  | **12** | **2** |  | **5** | **1** | **6** | **4** |
| **6** | **4** | **2** |  |  | **3** | **5** |  |  |  |  |
| **7** |  | **14** | **6** | **2** | **4** |  |  |  |  | **10** |
| **8** | **6** | **1** | **3** | **14** | **5** | **13** | **3** | **4** |  | **5** |
| **9** | **2** | **5** | **1** | **6** | **10** | **5** | **1** | **11** | **1** |  |

**Q8\* :**

**Train kNN với k = 1 độ chính xác : 96.91%**

|  |  |  |
| --- | --- | --- |
|  | **True** | **False** |
| **0** | **973** | **7** |
| **1** | **1129** | **6** |
| **2** | **992** | **40** |
| **3** | **970** | **40** |
| **4** | **944** | **38** |
| **5** | **860** | **32** |
| **6** | **944** | **14** |
| **7** | **992** | **36** |
| **8** | **920** | **54** |
| **9** | **967** | **42** |

**Train KNN voi k = 1, Distance = cosine .**

Mdl = fitcknn(imgTrainAll',lblTrainAll,'Distance','cosine')

|  |  |  |
| --- | --- | --- |
|  | **True** | **False** |
| **0** | **978** | **2** |
| **1** | **1129** | **6** |
| **2** | **1003** | **29** |
| **3** | **977** | **33** |
| **4** | **940** | **42** |
| **5** | **852** | **40** |
| **6** | **946** | **12** |
| **7** | **995** | **33** |
| **8** | **935** | **39** |
| **9** | **968** | **41** |

**Độ chính xác 97,23%**

**Train với k = 3**

Mdl = fitcknn(imgTrainAll',lblTrainAll,'NumNeighbors',3)

|  |  |  |
| --- | --- | --- |
|  | **True** | **False** |
| **0** | **974** | **6** |
| **1** | **1133** | **2** |
| **2** | **993** | **39** |
| **3** | **974** | **36** |
| **4** | **951** | **31** |
| **5** | **863** | **29** |
| **6** | **944** | **14** |
| **7** | **992** | **36** |
| **8** | **914** | **60** |
| **9** | **968** | **41** |

**Train kNN với k = 3 độ chính xác : 97.06%**