Bài tập thực hành 3

Khảo sát bộ dữ liệu chữ số viết tay

Q1:

```
function TH3Q1(n)
    imgTrainAll = loadMNISTImages('train-images.idx3-ubyte');
    lblTrainAll = loadMNISTLabels('train-labels.idx1-ubyte');
    nTrainImages = size(imgTrainAll, 2);

    if n>0 && n<=nTrainImages
        figure;
        imgN = imgTrainAll(:,n);
        img2DN = reshape(imgN, 28, 28);
        strLabelImage = num2str(lblTrainAll(n));
        imshow(img2DN);
        title(strLabelImage);
    end
end
```

n	1	500	5000	10000	59000
Kết quả	5	8	2	7	4

Q2:

```
function TH3Q2(n)
    imgTestAll = loadMNISTImages('t10k-images.idx3-ubyte');
    lblTestAll = loadMNISTLabels('t10k-labels.idx1-ubyte');

    nTestImages = size(imgTestAll, 2);

    if n>0 && n<=nTestImages
        figure;
        imgN = imgTestAll(:,n);
        img2DN = reshape(imgN, 28, 28);
        strLabelImage = num2str(lblTestAll(n));
        imshow(img2DN);
        title(strLabelImage);
    end
end
```

n	1	500	5000	9000
Kết quả	7	6	0	0

Q3:

```
function numMat = TH3Q3()
    lblTrainAll = loadMNISTLabels('train-labels.idx1-ubyte');

    numMat = unique(lblTrainAll);

    for i=1:size(numMat,1)
        q = find(lblTrainAll == numMat(i));
```

```
        numMat(i,2) = size(q,1);
    end
    csvwrite('ThongKeTrain.csv',numMat');
end
```

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

Q4:

```
function numMat = TH3Q4()
    lblTestAll = loadMNISTLabels('t10k-labels.idx1-ubyte');

    numMat = unique(lblTestAll);

    for i=1:size(numMat,1)
        q = find(lblTestAll == numMat(i));
        numMat(i,2) = size(q,1);
    end
    csvwrite('ThongKeTest.csv',numMat');
end
```

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

Q5:

```
function TH3Q5(n)
    imgTrainAll = loadMNISTImages('train-images.idx3-ubyte');
    lblTrainAll = loadMNISTLabels('train-labels.idx1-ubyte');
    imgTestAll = loadMNISTImages('t10k-images.idx3-ubyte');
    lblTestAll = loadMNISTLabels('t10k-labels.idx1-ubyte');

    Mdl = fitcknn(imgTrainAll',lblTrainAll);

    nTestImages = size(imgTestAll, 2);

    if n>0 && n<=nTestImages
        imgTest = imgTestAll(:,n);
        lblPredictTest = predict(Mdl,imgTest');
        fprintf('Ket qua du doan anh test co thu tu n=%d la:
%d\n',n,lblPredictTest);
    end
end
```

n	5	500	900
Kết quả	4	6	8

Q6:

```
function TH3Q6(n)
    imgTrainAll = loadMNISTImages('train-images.idx3-ubyte');
    lblTrainAll = loadMNISTLabels('train-labels.idx1-ubyte');
    imgTestAll = loadMNISTImages('t10k-images.idx3-ubyte');
    lblTestAll = loadMNISTLabels('t10k-labels.idx1-ubyte');

    Mdl = fitcknn(imgTrainAll',lblTrainAll);

    nTestImages = size(imgTestAll, 2);

    if n>0 && n<=nTestImages
        imgTest = imgTestAll(:,n);
        lblPredictTest = predict(Mdl,imgTest');

        lblImageTest = lblTestAll(n);

        figure;
        img2D = reshape(imgTest, 28, 28);
        imshow(img2D);
        strLabelImage = 'Ban dau ';
        strLabelImage = [strLabelImage, num2str(lblImageTest), '.'];
        strLabelImage = [strLabelImage, ' Du doan: '];
        strLabelImage = [strLabelImage, num2str(lblPredictTest), '.'];

        if(lblPredictTest == lblImageTest)
            strLabelImage = [strLabelImage, ' Ket qua dung. '];
        else
            strLabelImage = [strLabelImage, ' Ket qua sai. '];
        end

        title(strLabelImage);
    end
end
```

Q7:

```
function TH3Q7(n)
    imgTrainAll = loadMNISTImages('train-images.idx3-ubyte');
    lblTrainAll = loadMNISTLabels('train-labels.idx1-ubyte');
    imgTestAll = loadMNISTImages('t10k-images.idx3-ubyte');
    lblTestAll = loadMNISTLabels('t10k-labels.idx1-ubyte');

    Mdl = fitcknn(imgTrainAll',lblTrainAll);

    nTestImages = size(imgTestAll, 2);

    if n>=0 && n<10
        count = 0;
        for t=1:nTestImages
            if(lblTestAll(t) == n)
                imgTest = imgTestAll(:,t);
                lblPredictTest = predict(Mdl,imgTest');
                if(lblPredictTest ~= n)
                    count = count+1;
                end
            end
```

```
        end
        fprintf('So luong du doan sai so %d: %d\n',n,count);
    end
end
```

Số lượng dự đoán sai:

0	1	2	3	4	5	6	7	8	9
7	6	40	40	38	32	14	36	54	42

Q7*:

```
function TH3Q72(n)
    imgTrainAll = loadMNISTImages('train-images.idx3-ubyte');
    lblTrainAll = loadMNISTLabels('train-labels.idx1-ubyte');
    imgTestAll = loadMNISTImages('t10k-images.idx3-ubyte');
    lblTestAll = loadMNISTLabels('t10k-labels.idx1-ubyte');

    Mdl = fitcknn(imgTrainAll',lblTrainAll);

    numMat = unique(lblTestAll);
    confMat = [numMat zeros(size(numMat,1),1)];

    if n>=0 && n<10
        q = find(lblTestAll == n);
        nTestImages = size(q,1);
        for t=1:nTestImages
            imgTest = imgTestAll(:,q(t));
            lblPredictTest = predict(Mdl,imgTest');

            if(lblPredictTest == n)
                rowID = find(numMat == n);
            else
                rowID = find(numMat == lblPredictTest);
            end
            confMat(rowID,2)= confMat(rowID,2) + 1;
        end
        fprintf('Confusion matrix du doan label co nhan la %d:\n',n);
        disp(confMat');
    end
end
```

Bảng confusion matrix:

Cột Nhãn	Cột dự đoán và các kết quả tương ứng									
	0	1	2	3	4	5	6	7	8	9
0	973	1	1	0	0	1	3	1	0	0
1	0	1129	3	0	1	1	1	0	0	0
2	7	6	992	5	1	0	2	16	3	0
3	0	1	2	970	1	19	0	7	7	3
4	0	7	0	0	944	0	3	5	1	22
5	1	1	0	12	2	860	5	1	6	4
6	4	2	0	0	3	5	944	0	0	0
7	0	14	6	2	4	0	0	992	0	10
8	6	1	3	14	5	13	3	4	920	5
9	2	5	1	6	10	5	1	11	1	967