

Περνάμε την βιβλιοθήκη mnist στην octave

Δεδομένα για test

```
data1 = load('mnist_test.csv');
```

```
labels1 = data1(:,1);
```

```
images1 = data1(:,2:785);
```

Δεδομένα εκπαίδευσης

```
data2 = load('mnist_train.csv');
```

```
labels2 = data2(:,1);
```

```
images2 = data2(:,2:785);
```

Ερώτημα 1

```
% Task 1
```

```
j = 1;
```

```
for i=1:10000
```

```
    if labels1(i) == 0
```

```
        Ni(j,:) = images1(i,:);
```

```
        Lte(j,:) = labels1(i);
```

```
        j++;
```

```
    elseif labels1(i) == 1
```

```
        Ni(j,:) = images1(i,:);
```

```
        Lte(j,:) = labels1(i);
```

```
        j++;
```

```
    elseif labels1(i) == 2
```

```
        Ni(j,:) = images1(i,:);
```

```
        Lte(j,:) = labels1(i);
```

```

    j++;
elseif labels1(i) == 3
    Ni(j,:) = images1(i,:);
    Lte(j,:) = labels1(i);
    j++;
end
end
end

```

Τα δεδομένα για το test που ανοίκουν στις κλάσεις 0-3 είναι 4157

```

j = 1;

for i=1:60000
    if labels2(i) == 0
        Mi(j,:) = images2(i,:);
        Ltr(j,:) = labels2(i);
        j++;
    elseif labels2(i) == 1
        Mi(j,:) = images2(i,:);
        Ltr(j,:) = labels2(i);
        j++;
    elseif labels2(i) == 2
        Mi(j,:) = images2(i,:);
        Ltr(j,:) = labels2(i);
        j++;
    elseif labels2(i) == 3
        Mi(j,:) = images2(i,:);
        Ltr(j,:) = labels2(i);
    end
end

```

```
j++;
```

```
end
```

```
end
```

Τα δεδομένα εκπαίδευσης που ανοίκουν στις κλάσεις 0-3 είναι 24754

```
% Task 2
```

```
for i = 1:length(Ltr)
```

```
    im = reshape(Mi(i,:),28,28);
```

```
    x = mean(im,2);
```

```
    y = mean(im,1);
```

```
    M1(i,[1 2]) = [mean(x),mean(y)]; % M1 = M^
```

```
end
```

Ο M1 είναι ο πίνακας M^ και έχει μέγεθος 24754x2

```
hold on
```

```
for i = 1:length(Ltr)
```

```
    if Ltr(i) == 0
```

```
        scatter(M1(i,1),M1(i,2),[],[1 0 0], 'filled');
```

```
    elseif Ltr(i) == 1
```

```
        scatter(M1(i,1),M1(i,2),[],[0 1 0], 'filled');
```

```
    elseif Ltr(i) == 2
```

```
        scatter(M1(i,1),M1(i,2),[],[0 0 1], 'filled');
```

```
    elseif Ltr(i) == 3
```

```
        scatter(M1(i,1),M1(i,2),[],[1 1 0], 'filled');
```

```
    end
```

```
end
```

```
hold off
```

% Task 3

[c,d,idx] = Kmeans(M1,4,0); %c = centroid location, d = max centroid-point distance

O c έχει μέγεθος 4x2 και περιέχει τις συντεταγμένες των κέντρων τις κάθε ομάδας

O d έχει μέγεθος 4x1 και περιέχει την μέγιστη απόσταση σημείου απο το κέντρο

O idx έχει μέγεθος 24754x1 και περιέχει την ετικέτα της ομάδας που ανοίκει στο σημείο i, i=1,...,24574

hold on

for i = 1:length(idx)

if idx(i) == 1

scatter(M1(i,1),M1(i,2),[],[1 0 0],'filled');

elseif idx(i) == 2

scatter(M1(i,1),M1(i,2),[],[0 1 0],'filled');

elseif idx(i) == 3

scatter(M1(i,1),M1(i,2),[],[0 0 1],'filled');

elseif idx(i) == 4

scatter(M1(i,1),M1(i,2),[],[1 1 0],'filled');

end

end

hold off

[Acc,rand_index,match] = AccMeasure(Ltr,idx);

Το Acc είναι το ποσοστό της καθαρότητας (με τα train data αργούσε πολύ, με τα test ήταν κοντά στο 42%, το ίδιο ισχύει και για τις υπόλοιπες υλοποιήσεις του AccMeasure)

Το match είναι ένας πίνακας που δείχνει κατά πόσο ταιριάζει το αποτέλεσμα με το πρωτότυπο

% Task 4

% V = 2

[mappedM,mapping] = pca(Mi,2);

Ο πίνακας mappedM είναι το αποτέλεσμα του αλγορίθμου pca

M2 = mappedM; % M2 = M~

hold on

for i = 1:length(Ltr)

if Ltr(i) == 0

scatter(M2(i,1),M2(i,2),[],[1 0 0], 'filled');

elseif Ltr(i) == 1

scatter(M2(i,1),M2(i,2),[],[0 1 0], 'filled');

elseif Ltr(i) == 2

scatter(M2(i,1),M2(i,2),[],[0 0 1], 'filled');

elseif Ltr(i) == 3

scatter(M2(i,1),M2(i,2),[],[1 1 0], 'filled');

end

end

hold off

[c,d,idx] = Kmeans(M2,4,0);

hold on

for i = 1:length(idx)

```

if idx(i) == 1
    scatter(M2(i,1),M2(i,2),[],[1 0 0], 'filled');
elseif idx(i) == 2
    scatter(M2(i,1),M2(i,2),[],[0 1 0], 'filled');
elseif idx(i) == 3
    scatter(M2(i,1),M2(i,2),[],[0 0 1], 'filled');
elseif idx(i) == 4
    scatter(M2(i,1),M2(i,2),[],[1 1 0], 'filled');
end
end
hold off

```

```

[Acc,rand_index,match] = AccMeasure(Ltr,idx);

```

Με τα test data ήταν κοντά στο 36-40%

```

V(1,:) = [2,Acc];

```

```

% V = 25

```

```

[mappedM,mapping] = pca(Mi,25);

```

```

M2 = mappedM;

```

```

[c,d,idx] = Kmeans(M2,4,0);

```

```

[Acc,rand_index,match] = AccMeasure(Ltr,idx);

```

Με τα test data ήταν περίπου 66%

```
V(2,:) = [25,Acc];
```

```
% V = 50
```

```
[mappedM,mapping] = pca(Mi,50);
```

```
M2 = mappedM;
```

```
[c,d,idx] = Kmeans(M2,4,0);
```

```
[Acc,rand_index,match] = AccMeasure(Ltr,idx);
```

Με τα test data ήταν κοντά στο 50%

```
V(3,:) = [50,Acc];
```

```
% V = 100
```

```
[mappedM,mapping] = pca(Mi,100);
```

```
M2 = mappedM;
```

```
[c,d,idx] = Kmeans(M2,4,0);
```

```
[Acc,rand_index,match] = AccMeasure(Ltr,idx);
```

Με τα test data ήταν κοντά στο 30%

```
V(4,:) = [100,Acc];
```

```
Vmax = V(V(:,2) == max(V(:,2)),:);
```

Vmax = 25 66%

% Task 5

```
[mappedM,mapping] = pca(Mi,Vmax(1));
```

```
M2 = mappedM;
```

```
[mappedM,mapping] = pca(Ni,Vmax(1));
```

```
N2 = mappedM; % N2 = N~
```

```
Labels = gnb(M2,Ltr,N2);
```

Ο πίνακας Labels είναι 4157x1 και είναι το αποτέλεσμα του gnb

```
conf = Lte == Labels;
```

Ο πίνακας conf είναι 4157x1 και είναι ο έλεγχος ορθής ταξινόμησης

```
accuracy = (sum(conf)/length(conf))*100;
```

Το accuracy είναι το ποσοστό ορθής ταξινόμησης και είναι 90,859%