ECE/OPTI 533 Digital Image Processing **Spring Semester, 2003**
Homework 8
Due in ECE main office (Curt Booth) before 3PM, April 22, 2003

1.a. Write a basic K-means clustering program, with no merging or splitting, using the nearest-Euclidean distance $\sqrt{D1^2 + D2^2 + D3^2 + D4^2}$. The program should produce a cluster map the same size as the input image, with the DN assignments 1=cluster 1, 2=cluster 2, etc.

Apply the program to the 4-D data "band1", "band2", "band3" and "band4" on the class website. These are each of the size:

512byte header + 1 sample x 150 lines x 32bits/pixel

in the same format as the other class images.

Find the resulting cluster means for K=3 and 10 iterations. (60%)

b. Change the seed mean values, and repeat part (a) for the same number of iterations. (10%)

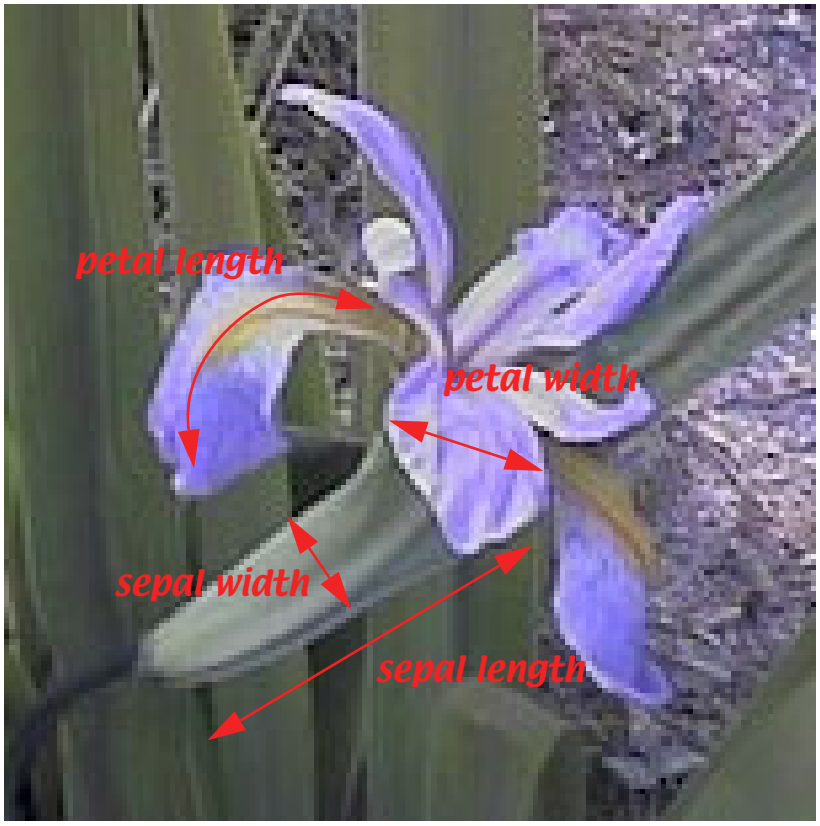
c. Plot the *mean cluster migration distance* and *number of pixels that change cluster assignment* versus iteration number for part (a). (20%)

2. Modify the program to use the "city-block" distance, i.e. |D1|+|D2|+|D3|+|D4| and find the cluster means again using the same seed mean values as in part (a). (10%)

These data are known as Fisher's Iris data (1936). They are measurements of dimensions of flowers of the iris family and are famous as a test set for statistical classification, clustering, etc (try a Google search on "Fisher's Iris").

Fisher, R.A. (1936), "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics , 7, 179 -188.

The measured variables are:



There are 3 species represented in the data, *I. Setosa*, *I. Versicolor*, and *I. Verginica*. The following table shows the data sorted by sepal width (the same as your data):

Fisher's Iris data

HW8 band	1	2	3	4
species	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
I. Verginica	5	2	3.5	1
I. Setosa	6	2.2	5	1.5
I. Verginica	6.2	2.2	4.5	1.5
I. Verginica	6	2.2	5	1
I. Verginica	5	2.3	3.3	1
I. Verginica	5.5	2.3	4	1.3
I. Verginica	6.3	2.3	4.4	1.3
I. Versicolor	4.5	2.3	1.3	0.3
I. Verginica	4.9	2.4	3.3	1
I. Verginica	5.5	2.4	3.7	1
I. Verginica	5.5	2.4	3.8	1.1
I. Setosa	4.9	2.5	4.5	1.7

I. Setosa	6.3	2.5	5	1.9
I. Setosa	5.7	2.5	5	2
I. Setosa	6.7	2.5	5.8	1.8
I. Verginica	5.1	2.5	3	1.1
I. Verginica	5.6	2.5	3.9	1.1
I. Verginica	5.5	2.5	4	1.3
I. Verginica	6.3	2.5	4.9	1.5
I. Setosa	6.1	2.6	5.6	1.4
I. Setosa	7.7	2.6	6.9	2.3
I. Verginica	5.7	2.6	3.5	1
I. Verginica	5.8	2.6	4	1.2
I. Verginica	5.7	2.6	4.2	1.3
I. Verginica	5.5	2.6	4.4	1.2
I. Setosa	6.3	2.7	4.9	1.8
I. Setosa	5.8	2.7	5.1	1.9
I. Setosa	5.8	2.7	5.1	1.9
I. Setosa	6.4	2.7	5.3	1.9
I. Verginica	5.8	2.7	3.9	1.2
I. Verginica	5.2	2.7	3.9	1.4
I. Verginica	5.8	2.7	4.1	1
I. Verginica	5.6	2.7	4.2	1.3
I. Verginica	6	2.7	5.1	1.6
I. Setosa	6.2	2.8	4.8	1.8
I. Setosa	5.6	2.8	4.9	2
I. Setosa	6.3	2.8	5.1	1.5
I. Setosa	5.8	2.8	5.1	2.4
I. Setosa	6.4	2.8	5.6	2.1
I. Setosa	6.4	2.8	5.6	2.2
I. Setosa	7.4	2.8	6.1	1.9
I. Setosa	7.7	2.8	6.7	2
I. Verginica	6.1	2.8	4	1.3
I. Verginica	5.7	2.8	4.1	1.3
I. Verginica	5.7	2.8	4.5	1.3
I. Verginica	6.5	2.8	4.6	1.5
I. Verginica	6.1	2.8	4.7	1.2
I. Verginica	6.8	2.8	4.8	1.4
I. Setosa	6.3	2.9	5.6	1.8
I. Setosa	7.3	2.9	6.3	1.8
I. Verginica	5.6	2.9	3.6	1.3
I. Verginica	6.4	2.9	4.3	1.3
I. Verginica	6	2.9	4.5	1.5
I. Verginica	6.1	2.9	4.7	1.4
I. Verginica	6.2	2.9	5.4	1.3
I. Verginica	6.6	2.9	5.6	1.3
I. Versicolor	4.4	2.9	1.4	0.2
I. Setosa	6	3	4.8	1.8

I. Setosa	6.1	3	4.9	1.8
I. Setosa	5.9	3	5.1	1.8
I. Setosa	6.7	3	5.2	1.3
I. Setosa	6.5	3	5.2	2
I. Setosa	6.5	3	5.5	1.8
I. Setosa	6.8	3	5.5	2.1
I. Setosa	7.2	3	5.8	1.6
I. Setosa	6.5	3	5.8	2.2
I. Setosa	7.1	3	5.9	2.1
I. Setosa	7.7	3	6.1	2.3
I. Setosa	7.6	3	6.6	2.1
I. Verginica	5.6	3	4.1	1.3
I. Verginica	5.7	3	4.2	1.2
I. Verginica	5.9	3	4.2	1.5
I. Verginica	6.6	3	4.4	1.4
I. Verginica	5.4	3	4.5	1.5
I. Verginica	5.6	3	4.5	1.5
I. Verginica	6.1	3	4.6	1.4
I. Verginica	6.7	3	5	1.7
I. Versicolor	4.3	3	1.1	0.1
I. Versicolor	4.4	3	1.3	0.2
I. Versicolor	4.8	3	1.4	0.1
I. Versicolor	4.9	3	1.4	0.2
I. Versicolor	4.8	3	1.4	0.3
I. Versicolor	5	3	1.6	0.2
I. Setosa	6.9	3.1	5.1	2.3
I. Setosa	6.9	3.1	5.4	2.1
I. Setosa	6.4	3.1	5.5	1.8
I. Setosa	6.7	3.1	5.6	2.4
I. Verginica	6.7	3.1	4.4	1.4
I. Verginica	6.7	3.1	4.7	1.5
I. Verginica	6.9	3.1	4.9	1.5
I. Versicolor	4.9	3.1	1.5	0.1
I. Versicolor	4.9	3.1	1.5	0.2
I. Versicolor	4.6	3.1	1.5	0.2
I. Versicolor	4.8	3.1	1.6	0.2
I. Setosa	6.5	3.2	5.1	2
I. Setosa	6.4	3.2	5.3	2.3
I. Setosa	6.9	3.2	5.7	2.3
I. Setosa	6.8	3.2	5.9	2.3
I. Setosa	7.2	3.2	6	1.8
I. Verginica	6.4	3.2	4.5	1.5
I. Verginica	7	3.2	4.7	1.4
I. Verginica	5.9	3.2	4.8	1.8
I. Versicolor	5	3.2	1.2	0.2
I. Versicolor	4.4	3.2	1.3	0.2

I. Versicolor	4.7	3.2	1.3	0.2
I. Versicolor	4.6	3.2	1.4	0.2
I. Versicolor	4.7	3.2	1.6	0.2
I. Setosa	6.7	3.3	5.7	2.1
I. Setosa	6.7	3.3	5.7	2.5
I. Setosa	6.3	3.3	6	2.5
I. Verginica	6.3	3.3	4.7	1.6
I. Versicolor	5	3.3	1.4	0.2
I. Versicolor	5.1	3.3	1.7	0.5
I. Setosa	6.2	3.4	5.4	2.3
I. Setosa	6.3	3.4	5.6	2.4
I. Verginica	6	3.4	4.5	1.6
I. Versicolor	5.2	3.4	1.4	0.2
I. Versicolor	4.6	3.4	1.4	0.3
I. Versicolor	5.1	3.4	1.5	0.2
I. Versicolor	5.2	3.4	1.5	0.2
I. Versicolor	5.4	3.4	1.5	0.4
I. Versicolor	4.8	3.4	1.6	0.2
I. Versicolor	5	3.4	1.6	0.4
I. Versicolor	5.4	3.4	1.7	0.2
I. Versicolor	4.8	3.4	1.9	0.2
I. Versicolor	5.5	3.5	1.3	0.2
I. Versicolor	5	3.5	1.3	0.3
I. Versicolor	5.1	3.5	1.4	0.2
I. Versicolor	5.1	3.5	1.4	0.3
I. Versicolor	5.2	3.5	1.5	0.2
I. Versicolor	5	3.5	1.6	0.6
I. Setosa	7.2	3.6	6.1	2.5
I. Versicolor	4.6	3.6	1	0.2
I. Versicolor	4.9	3.6	1.4	0.1
I. Versicolor	5	3.6	1.4	0.2
I. Versicolor	5.4	3.7	1.5	0.2
I. Versicolor	5.3	3.7	1.5	0.2
I. Versicolor	5.1	3.7	1.5	0.4
I. Setosa	7.9	3.8	6.4	2
I. Setosa	7.7	3.8	6.7	2.2
I. Versicolor	5.1	3.8	1.5	0.3
I. Versicolor	5.1	3.8	1.6	0.2
I. Versicolor	5.7	3.8	1.7	0.3
I. Versicolor	5.1	3.8	1.9	0.4
I. Versicolor	5.4	3.9	1.3	0.4
I. Versicolor	5.4	3.9	1.7	0.4
I. Versicolor	5.8	4	1.2	0.2
I. Versicolor	5.2	4.1	1.5	0.1
I. Versicolor	5.5	4.2	1.4	0.2
I. Versicolor	5.7	4.4	1.5	0.4

The 3 classes have the following means in each dimension (compare to your clustering results):

species	band 1	band 2	band 3	band 4
I. Setosa	6.588	2.974	5.552	2.006
I. Verginica	5.936	2.764	4.322	1.326
I. Versicolor	5.01	3.428	1.462	0.246

The dataset was clustered using tclSADIE, with the following Session Log output:

```
CLUSTER

 Input image:                  Iris BSQ
 Number of iterations:         10
 Number of seed classes:       3
 Minimum class size:           1
 Class merging threshold:      0.0000e+00
 Subsample increment: bands:   1
                      lines:   1
                      pixels:  1
 Outlier threshold:            1.0000e+02
 ...............

 Mean vector migration averaged over all classes:
 Iteration     Migration     Number of classes after merging
     1       3.5591e-01                  3
     2       5.3170e-02                  3
     2       5.3170e-02                  3
     3       2.6835e-02                  3
     4       0.0000e+00                  3
     5       0.0000e+00                  3
     6       0.0000e+00                  3
     7       0.0000e+00                  3
     8       0.0000e+00                  3
     9       0.0000e+00                  3
    10       0.0000e+00                  3
     *       0.0000e+00                  3
     0              N/A                          N/A                          N/A                          N/A                          0    0.0
     1   5.0060e+00 +/-3.4895e-01   3.4280e+00 +/-3.7525e-01   1.4620e+00 +/-1.7192e-01   2.4600e-01 +/-1.0433e-01         50   33.3
     2   5.9048e+00 +/-4.5962e-01   2.7460e+00 +/-2.9213e-01   4.4127e+00 +/-5.2296e-01   1.4333e+00 +/-2.9277e-01         63   42.0
     3   6.8703e+00 +/-4.7810e-01   3.0865e+00 +/-2.7915e-01   5.7459e+00 +/-4.8802e-01   2.0892e+00 +/-2.5657e-01         37   24.7
```

The cluster means are:

cluster	band 1	band 2	band 3	band 4
1	5.006	3.428	1.462	0.246
2	5.905	2.746	4.413	1.433
3	6.87	3.087	5.746	2.089

which are quite close to the above class means. The clusters evidently have the following correspondances:

cluster 1 = *I. Versicolor*	50 data points
cluster 2 = *I. Verginica*	63 data points
cluster 3 = *I. Setosa*	37 data points

So, 13 data points are assigned to cluster 2 by mistake because of overlap with cluster 3. Also, note the mean migration converges as expected,

iteration	net mean migration
1	0.356
2	0.053
3	0.027

FYI, the flowers look like:

I. Setosa

I. Verginica

I. Versicolor

BLUEFLAG IRIS
Iris versicolor L.
IRIS FAMILY