Winter 2002, CS 790
Introduction to Pattern Recognition
Homework #3
Due date: 2/25

In recognition of the Wright State University policies of academic integrity, I certify that I have neither given nor received dishonest aid in this homework assignment.

Name: _____ Signature: _____

NOTES:

1. *Start each problem on a separate page*
2. *Provide hardcopies of all plots, MATLAB code and derivations.*
3. *Download the compressed file 'hw3.zip' from the course web page*
4. *Sign and return this page with your finished assignment.*

Problem 1 (33%)

Load the file 'hw3p1_data', which contains a database of 67 animals, where each animal is identified by its name and a feature vector containing sixteen descriptors –the file 'hw3p1_readme.txt' contains the definitions of these descriptors.

(i) Perform hierarchical clustering using different distance measures.

(ii) Generate dendrograms, labeling each leaf by the name of the animal. Can you identify any groupings of animals from the dendrograms?

(iii) Discuss your results.

HINT: use the MATLAB functions 'linkage' and 'dendrogram'.

Problem 2 (33%)

Load the dataset 'hw3p2_data', which contains a two-dimensional distribution of training examples.

(i) Generate a 2D scatter plot to visualize the structure of the data. How many clusters can be identified?

(ii) Create a MATLAB function that performs k-means clustering. The function should be passed a matrix of training examples and a pre-specified number of clusters (k), and should return the resulting centers of the k clusters, as well as a cluster assignment for each training example.

(iii) Generate a 2D scatter plot where each example is labeled by its corresponding cluster index. HINT: use the MATLAB function 'text'.

(iv) Perform k-means for different values of k (i.e., too small, appropriate, too large) and discuss your results.

Problem 3 (34%)

Load the file 'hw3p3_data', which contains a dataset of colors defined by their coordinates in RGB space.

(i) Generate a three-dimensional scatter plot of the training data. Can you identify any structure in this data?

(ii) Build a Kohonen Self-Organizing Map with a 10x10 grid topology to learn the 3D distribution of the data. As a result, each neuron will be defined by its coordinates in RGB space (a 3D vector) and its coordinates in the 10x10 mesh (a 2D vector). Select appropriate learning parameters.

(iii) Plot the trained SOM mesh in RGB space along with the training data. What is the relationship between the distribution of training examples and the SOM neurons?

(iv) As an alternative representation for the trained SOM, generate a 10x10 image where the color of each pixel is defined by the 3D RGB coordinates of the corresponding neuron in the SOM. Can you identify any structure across the mesh? Generate a color JPEG image and send it to me as an email attachment –you may use the MATLAB function call: print(gcf, '-djpeg75', 'som.jpg')

(v) What is the relationship between the two alternative representations in (iii) and (iv)?

(vi) Discuss your results

HINTS: You can write your own SOM code or use the MATLAB Neural Network Toolbox. If you decide to go for the second option, take a look at the MATLAB SOM demos (demosm1 and demosm2), as well as the functions 'newsom', 'plotsom' and 'train'.