Homework 4

CS 591Q/791V - Pattern Recognition
Instructor: Dr. Arun Ross
Due Date: April 30, 2009

Note: You are permitted to discuss the following questions with others in the class. However, you *must* write up your *own* solutions to these questions. Any indication to the contrary will be considered an act of academic dishonesty. Code developed as part of this assignment should be placed in a zip file and sent to arun.ross at mail.wvu.edu with the subject line "CS 591Q/791V : Homework 4". Also, include a hard-copy of your code when you submit the homework.

1. Consider a set of two-dimensional points pertaining to two classes that are linearly separable. The data can be accessed here. Note that the class labels (+1 or -1) are indicated at the end of every pattern.

   (a) [10 points] Write a program to compute the linear decision boundary for this two-class problem by implementing the fixed-increment single-sample perceptron learning algorithm.

   (b) [10 points] Report the linear decision boundary computed by the learning algorithm and the number of iterations taken by the algorithm to converge to this boundary when the weight vector is initialized as follows: (i) $(0, 0, 0)^t$, (ii) $(1, 1, 1)^t$, (iii) $(5, 5, 5)^t$, (iv) $(-30, 1, 1)^t$, and (v) $(10, 10, 10)^t$. In each case plot the two dimensional-points, the initial decision boundary and the final decision boundary.

   (c) [5 points] Discuss the impact of the inital weight vector on the number of iterations required by the algorithm to converge. Will the algorithm converge to a solution if the classes are not linearly separable? Why or why not?

2. [15 points] The iris (flower) dataset consists of 150 4-dimensional patterns belonging to three classes (setosa, versicolor, and virginica). There are 50 patterns per class. The 4 features correspond to (a) sepal length in cm, (b) sepal width in cm, (c) petal length in cm, and (d) petal width in cm. The data can be accessed here. Note that the class labels are indicated at the end of every pattern.

   Write a program to project these four-dimensional points onto a two dimensional plane using Fisher's linear discriminant analysis. Report the two projection vectors estimated by the technique. Plot the entire dataset in two dimensions using these projection vectors.