

## CS 511A Fall 2004, HW5

**IMPORTANT:** You must follow the collaboration policy (check the course webpage if you forgot)!

**Due date:** Dec. 13, 1:00pm, Jolley 506.

### **Problem 1. (10 pts)**

You are asked to apply decision tree method to learn a prediction (classification) model based on a set of training points. Assume that you have the following training samples: (5, 2, +), (4, 1, +), (5, 3, -), (6, 4, -), (2, 0, +), (1, 1, +), (4, 4, -), where the data points are in the form of (x, y, +/-), x and y are two features and + and - are the two class labels. Which feature should you use for the root node of the tree? Describe your decision with some computation? You have to work out all the calculation (information gain or remainder) you need.

### **Problem 2. (15 pts)**

Problem 20.4 in the textbook

### **Problem 3. (10 pts)**

We are going to cluster the following eight points (with (x, y) representing location) into three clusters.

A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2(4,9)

The distance function is Euclidean distance. Suppose initially we assign A1, B1 and C1 as the center of each cluster, respectively. Use the k-means algorithm to show the three cluster centers and corresponding clusters after two round execution of the algorithm.