CPSC 422,
Homework 3

General Instructions:

- Working in Teams: You may work with at most one other person. That person must also be enrolled in CPSC 422 this term. If you are working with a partner, the two of you must submit only one assignment, listing both of your names. Keep in mind that when you work in pairs, each of you must understand the work that you submit.

- Note: You are permitted to derive your solutions by implementing the respective algorithms, even when it is not explicitly required in the question. You are NOT permitted to take existing implementations of the algorithms and use those for your solutions. Such existing algorithms include those available from friends, CD, or off the web. You are also NOT permitted to provide your implementation to anyone else outside your team.

- Important: It is best if you type your assignment. Handwritten work will be marked only if clearly legible, and will still have to be submitted via handin.

- Make sure to include your name and student ID (and the names and student IDs of the two team members if you are working with someone else) and the assignment number, at the top of the assignment. Ensure that each question and sub-question is appropriately labeled and clearly identifiable.

1 Question 1 [40 points]

In this question, we will study decision trees for the mail reading task discussed in class. We will use the dataset in the table below. As a stopping criterion for decision tree construction,

Example	User Action	Author	Thread	Length	Where read
e1	skips	known	new	long	home
e2	reads	unknown	new	short	work
e3	skips	unknown	follow_up	long	work
e4	skips	known	follow_up	long	home
e5	reads	known	new	short	home
e6	skips	known	follow_up	long	work
e7	skips	unknown	follow_up	short	work
e8	reads	unknown	new	short	work
e9	skips	known	follow_up	long	home
e10	skips	known	new	long	work
e11	skips	unknown	follow_up	short	home
e12	skips	known	new	long	work
e13	reads	known	follow_up	short	home
e14	reads	known	new	short	work
e15	reads	known	new	short	home
e16	reads	known	follow_up	short	work
e17	reads	known	new	short	home
e18	reads	unknown	new	short	work

we will require that all of the examples in a leaf have the same classification.

a [**5 points**] What is the tree constructed by algorithm DECISION-TREE-LEARNING (Figure 18.5 of Russell & Norvig) when attributes are selected to maximize information gain? The AIspace decision tree applet at
http://aispace.org/dTree/
can solve this question for you. (You can load the data set from the table above using
File→ Load Sample Dataset
and selecting the Mail Reading example.)

b [**10 points**] Suppose you change the attribute selection criterion to always select the first element of the list of features. What tree is found when the features are in the order [Author, Thread, Length, Where read]? Does this tree represent a different function than that found with the maximum information gain split? Explain.

c [**10 points**] What tree is found when the features are in the order [Where read, Thread, Length, Author]? Does this tree represent a different function than that found in question a) or the one in question b)? Explain.

d [**5 points**] Which of the 3 trees from questions a), b), and c) would you consider best? Why?

e [**10 points**] Is there a tree that correctly classifies the training examples, but represents a different function than those found by the preceding algorithms? If so, give it. If not, explain why.

2 Question 2 [40 points]

In this exercise on neural networks, we will trace the behaviour of the backpropagation algorithm using the applet at

http://www.aispace.org/neural/
(At that page, there are also two 5-minute video tutorials available that explain how to use the applet, which you can watch if youre having trouble with the applet.) Start the applet, and load the provided data set by clicking
File → Load Graph and Data From URL
and selecting the file
http://www.cs.ubc.ca/∼conati/422/422-2010World/Homework/Hmw3/single_datapoint_w1.xml
Alternatively, you can download the data from the course webpage and load it from the file. For simplicity, the training set in this sample problem consists of a single data point, with inputs x = [1, 0, 1, 0] and output y=0; also, all network weights are initialized with 1.

a **[15 points]** Write down and instantiate the equations for one step of backpropagation, for the two weights leading into the output node y from units Hidden 1 and Hidden 2. You can check your computation in the applet by going into *Solve* mode in the applet and clicking *Step* once.

b **[15 points]** Now load the provided data set
http://www.cs.ubc.ca/∼conati/422/422-2010World/Homework/Hmw3/two_datapoints_w0.xml, which has two training examples, the one from question a) and another one with inputs x = [0, 1, 0, 0] and output y=1, and initializes the weights at zero. Go into solve mode and perform a number of steps; why does nothing happen? (Note that one step in the Applet updates the weights based on all training examples at once, using the sum of the weight gradients due to each example; this is quite standard, but differs from the loop over training examples in the textbook.) [Hint: it can help to write down the equations from b) again for this example.] Which general problem of neural network training does this simple case exemplify? What is a possible solution to the problem?

c **[10 points]** Compute the training error (squared error as defined in the lecture) with this initial set of weights. (Write down and instantiate the formula.) . You can check your result by looking at the error value at the top of the applets main window.

3 Question 3 [10 points]

[Note that this question is worth marks, so dont forget to do it.]

a **[3 points]** For each question in this assignment, say how long you spent on it.

b **[3 points]** Rate each question in this assignment by how much you learnt doing it, on a scale from 1 (very little) to 5 (a whole lot)

c **[4 points]** For each question in this assignment, what did you learn?