Written Exam
DM533 - Introduction to Artificial Intelligence

Department of Mathematics and Computer Science

University of Southern Denmark, Odense

Monday, January 4, 2010

The exam is closed book. Only two double-side sheets of notes and a pocket calculator are allowed. Laptop, books and further notes are not allowed.

The exam consists of seven questions.

The weight of each exercise in the evaluation is given in percentage. A minimum of 50% is required to pass the exam. The remaining 50% of weight will be used to determine the grade. The duration of the exam is 3 hours.

---

1. **(10%) Search**

   Suppose you have two containers, initially empty. One has a capacity of exactly 3 liters; the other has a capacity of 5 liters. You can pour water from one container to another, empty a container, or fill a container at any time. Your problem is to place exactly 4 liters of water in the 5-liter container. Describe how this problem could be framed as a search problem defining the related components. Solve the problem within this framework and report the search performed.

2. **(10%) Search**

   Consider the tree of Figure 1. The leaves are roots of remaining subtrees not represented in the figure. We can examine this tree with different algorithms. The values on the edges of the tree represent the step cost, while the h values on the nodes, next to the identifying letter, are the heuristic values for completing a solution from that node. Simulate the search for the following strategies:

   - uniform cost search,
   - best-first search,
   - greedy construction.

   Please, write your answer in a similar form as follows:

Step	Visited	Open list
1	A	B I
2	B	C F B I
3	C	D E F B I

   At each step indicate the node visited and the queue of expanded nodes consequent to the visit (i.e., the open list). Order the queue such that the first node is the next

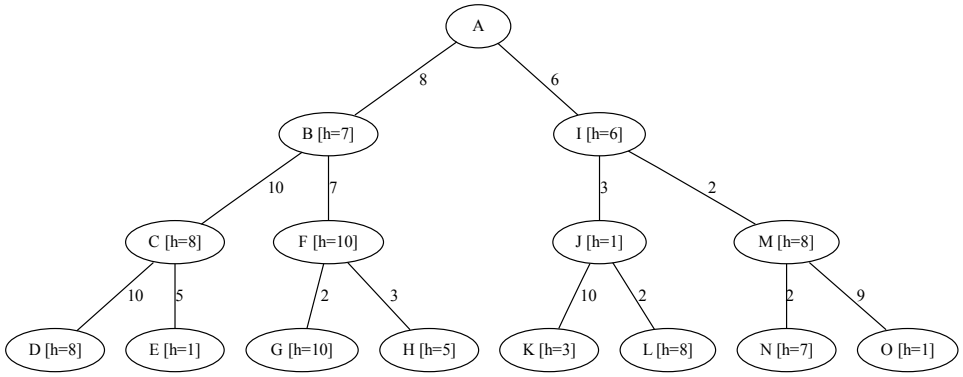

Figure 1: Values on the edges represent step costs and values on the nodes the heuristically computed value of completing the solution from that node.

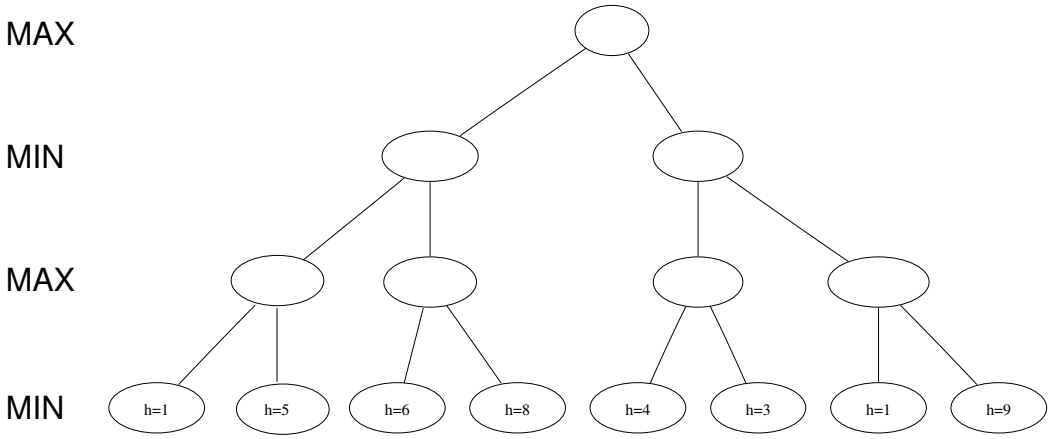

Figure 2: A search tree for α-β pruning. At the root is MAX to play.

node to be visited. Assume that the children of a node are expanded in alphabetical order when no other order is specified by the search. In addition, comment what would make the best-first an A* search.

3. **(10%) Games**

   Perform a left-to-right alpha-beta pruning of the tree in Figure 2. Indicate, in the figure, the direction you are taking (use arrows), the alpha and beta values at the nodes ($[\alpha, \beta]$), and exactly where pruning takes place (use a bar transversal to the pruned edge). You are playing the top node and want to win. Heuristic values are at leaf nodes.

4. **(30%) First Order Logic**

   In this problem, we formalize the permissions of files and directories in a file system. Here are the intended interpretations of the predicates we use:

   - $Owns(x, y)$: person x owns object y
   - $In(x, y)$: object x (which could be a file or a directory) is in directory y
   - $File(x)$: x is a file

- $Ne(x)$: x is a non-empty directory
- J: constant standing for Joachim

(a) Write each of the following sentences in first-order logic, using the predicates above.

i. A person owns a file if the same person owns the directory in which the file is.

ii. A directory is non-empty if and only if there is a file in the directory.

iii. There is a non-empty directory that is owned by Joachim

iv. There is a file that is owned by Joachim.

(b) How can you show that a set of sentences entails another sentence?

(c) How can you show that a set of sentences does not entail another sentence?

(d) Do the first three sentences entail the fourth? Show your answer using one of the two methods you just described.

5. **(20%) Bayesian Networks and Learning**

Consider a Naive Bayes problem with three features, x_1, x_2, x_3. Imagine that we have seen a total of 16 training examples, 8 positive (with y = 1) and 8 negative (with y = 0). Here is a table with some of the counts:

	y = 0	y = 1
$x_1 = 1$	7	7
$x_2 = 1$	1	1
$x_3 = 1$	3	5

What are the values estimated from the data for the following parameters:

- $\Pr(x_1 = 1|y = 0) = \theta_{10}$,
- $\Pr(x_2 = 1|y = 1) = \theta_{21}$,
- $\Pr(x_3 = 0|y = 0) = 1 - \theta_{30}$?

(You need to show the full derivation, answers by intuition without analytical justification do not count.)

6. **(10%) Learning**

In a classification task you are given the following data points: Negative: (-1, -1) (2, 1) (2, -1); Positive: (-2, 1) (-1, 1) (1,-1). The points are depicted in Figure 3.

Decision Tree

(a) Construct a decision tree using the recursive bi-partitioning algorithm based on information gain described in class. (Discretize the continue scale considering for f_1 only the values $\{-1.5, 0, 1.5\}$ and for f_2 only 0.) Represent graphically the tree constructed and draw the decision boundaries in the Figure 3.

(b) Explain how you chose the top-level attribute in the tree. Table 1 might be useful.

(c) Use the tree to predict the outcome for the new point (1,1).

Nearest Neighbor

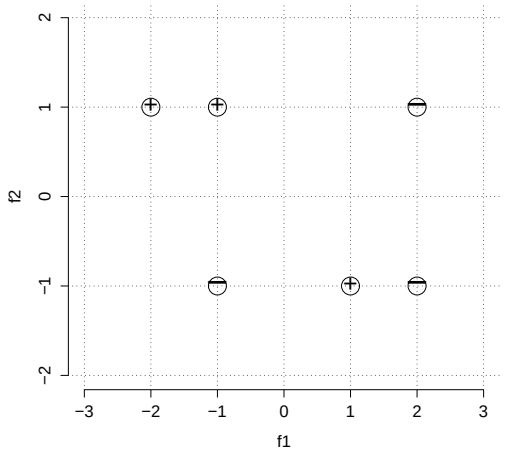

Figure 3: The data points for classification.

x	y	$-(x/y)\cdot\log(x/y)$	x	y	$-(x/y)\cdot\log(x/y)$
1	2	0.50	1	5	0.46
1	3	0.53	2	5	0.53
2	3	0.39	3	5	0.44
1	4	0.50	4	5	0.26
3	4	0.31			

Table 1: Numerical values for the computation of information gains.

(d) Draw the decision boundaries for 1-Nearest Neighbors on the Figure 3. Make it accurate enough so that it is possible to tell whether the integer-valued coordinate points in the diagram are on the boundary or, if not, which region they are in.

(e) What class does 1-NN predict for the new point: (1, 1) Explain why.

(f) What class does 3-NN predict for the new point: (1, 0) Explain why.

(g) In general, how would you select between two alternative values of k for use in k-nearest neighbors?

Perceptron

(h) Imagine to apply the perceptron learning algorithm to the points in Figure 3. Describe qualitatively what the result would be.

7. **(10%) Neural Networks**

Express the output of a neural network with one single hidden layer as a function of the input parameters when the activation function at the units is a linear function. Assume the same linear function at each unit. Would it be possible to simplify the network to a one layer perceptron?