CS 440: Introduction to AI

Homework 1

Due: Thursday, September 9th

Your answers must be concise and clear. Explain sufficiently that we can easily determine what you understand. We will give more points for a brief interesting discussion with no answer than for a bluffing answer.

AI Models

1. Suppose that we expect to filter spam messages from a mailbox. We want to build a model that can classify a message into a spam or a ham (non-spam). Consider using the features below:

   (1) The number of words in the message.

   (2) If the sender has sent a spam message before, the message is likely to be a spam by 90% accuracy.

   (3) If a text contains a frequently used spam phrase, the message is a spam by 70% accuracy.

   (4) Whether the message contains a picture or not.

   (5) Whether the sender is in the receiver's contacts.

   (6) In the previous spams, the number of sexual words is 5.7 on average.

   Recall that a model is a stand-in, or an approximate, mathematically precise representation, for the real thing.

   (a) Consider models that include each feature, and say whether the resulting model is analytic or empirical and briefly explain why (for some, both answers might be acceptable if it is properly justified).

i. *(Example)* Feature (1)
Analytic. The number of words can be counted and given to the model without experiments.

ii. Feature (2)

iii. Feature (3)

iv. Feature (4)

v. Feature (5)

vi. Feature (6)

(b) For the spam classification task above, give another feature that is:

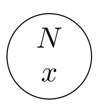
i. Purely analytic. (i.e., requires no experiments/observation)

ii. Purely empirical. Say what must be observed.

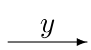
iii. Combination of analytic and empirical. Say what must be observed.

Search

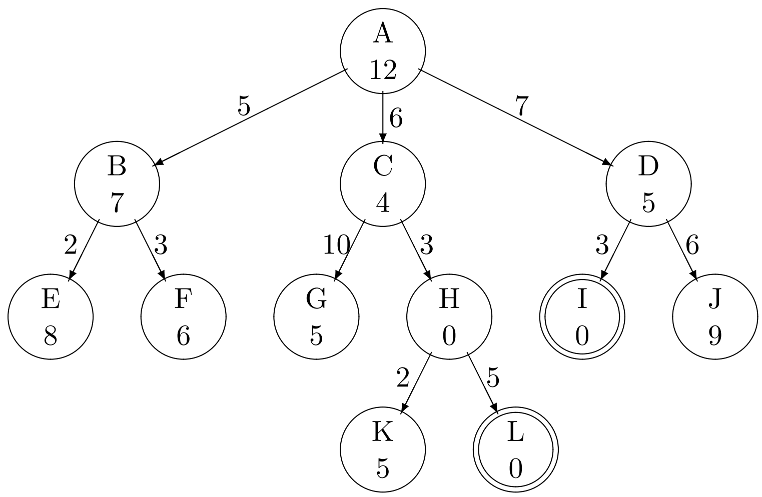
2. Consider the search tree below. The initial state is at the top, and the goal states are represented by the double circles. Note that the edges are directed.



N is the name of the node.
x is the heuristic function's estimate of the cost to the nearst goal, $h(N)$.



y is the actual cost of traversing an edge.



(a) For each of the search strategies listed below, (1) *form an ordered list of the states explored* and (2) *indicate which goal state is reached (if any).* Also (3) *specify the final contents of the queue at the time the search terminates.* When the states are expanded, new nodes are generated in ascending alphabetical order. Assume the sort is stable (i.e., ties are left as initially ordered).

i. Depth First Search

ii. Breadth First Search

iii. Uniform Cost

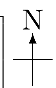
iv. Greedy / Best First

v. A^*

(b) In this example, is $h(N)$ admissible? Is it consistent? Briefly explain your answer.

3. Consider the following "grid world." Imagine a robot located at the start state trying to devise a route to reach the goal using a search algorithm. In each state, 4 operations are allowed: moving north, east, south, or west. Moving into a "Swamp" state costs 4 units, while moving into any other states costs 1 unit. Moving into the surrounding walls (e.g., moving west in the start state) is allowed and will result in the same state before the move, and the cost for the move depends on the state in which the move is attempted (4 for swamp, 1 otherwise).

Start	Swamp	Swamp
	Swamp	Goal



(a) i. Define a data structure that is adequate to model the state representation at the nodes of this problem.

ii. Define the operator "Move South" by specifying its preconditions and effects.

(b) Suppose that the robot is allowed to choose a random expansion order and a tie-breaker before starting exploration. Suppose also that repeating state (e.g., moving into the wall continuously) cannot be detected. For each of the search algorithms below, (1) What is the fewest possible number of nodes that must be visited? (2) Is the algorithm guaranteed to reach the goal state? (3) Would an optimal solution be found in this case?

i. Depth First Search

ii. Breadth First Search

iii. Uniform Cost

iv. A^* ($h(N)$ is Manhattan distance to Goal, e.g., $h(N)$ from Start is 3.)