CS 640 – 2025 Fall – Sample Midterm

Name:

BU ID:

This midterm is closed book and device free. Please do not write your answers on the backs of each page. Only the front sides will be scanned for grading.

Tip: You do not need to write background material in your answers if it does not contribute to answering the question.

1. What does it mean when a model has a true positive rate lower than the false positive rate? Give an example improvement to such a model that illustrates the issue.
2. Suresh and Guttag’s paper mentioned in class listed 7 kinds of biases in machine learning; historical bias, representation bias, measurement bias, learning bias, aggregation bias, evaluation bias, and deployment bias. Give a concrete example of one of those biases.
3. Given the following Prolog rules, explain how to answer the question

?- ancestor(mary, john).  
ancestor(X, Z) :- ancestor(X, Y), ancestor(Y, Z).  
ancestor(X, Y) :- parent(X, Y).  
parent(X, Y) :- father(X, Y).  
parent(X, Y) :- mother(X, Y).  
father(john, jack).  
father(jack, mary).  
mother(mary, jim).  
father(jim, john).

1. Briefly describe how bidirectional search works given a start node and a single goal state .
2. Consider the following heuristic time estimate for finding the fastest driving directions. =straight line distance from to the destination divided by 65 miles per hour. Is this an admissible heuristic? You may assume that all the relevant speed limits range from 20 to 65 miles per hour.
3. Explain what the Markov property means for the rewards of a Markov decision process.
4. In both value and policy iteration, a common subexpression is

What is the meaning of this subexpression?

1. In Q-learning, the learning rate alpha is used to bias the weighted average towards more recent data. Why would we want to down-weight the older samples when estimating the value?
2. While covering the backward algorithm for HMM evaluation, we discussed the function  
     
   Describe the meaning of in words.
3. The Viterbi algorithm identifies a sequence of states that maximizes the probability of the given observations . Now consider a variation where we only care about matching every other entry of . That is, we only care about observations where is even). Explain a modification to the Viterbi algorithm to maximize the probability of just those observations.