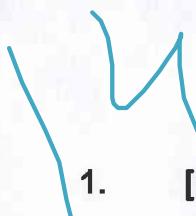


Total. Score: **92.0**

Q1. Score: **14.0**



1. [15%] True/False Questions

92D77825 67D0 4605-A970-531B925B8669

561-F14-Final

#93 2 of 12



Write T if the statement is always and unconditionally true, and write F otherwise.

- 1.1) T Perceptrons, arranged with no hidden layer, may express logical NOT.
- 1.2) F Probability theory and propositional logic make the same epistemological commitments.
- 1.3) T Bayesian Networks are a compact way in which to represent a joint probability distribution when some conditional independence assumptions hold.
- 1.4) T Bayes' rule contains both prior and posterior probabilities.
- 1.5) F BFS requires less memory than DFS.
- 1.6) F Neural networks can only be used to solve classification problems.
- 1.7) F SVM can classify data that is not linearly separable.
- 1.8) F A and B are independent if and only if  $P(A|B)=P(A)$
- 1.9) F Under the Markov assumption, current state can be determined by the next state.
- 1.10) F People usually use forward error propagation to train neural networks.
- 1.11) T An admissible heuristic function never overestimates the actual minimal cost of reaching the goal.
- 1.12) T When solving MDP, we usually cannot find the optimal policy in an accessible and stochastic environment with known transition model.
- 1.13) T We can use dynamic programming to solve MDPs.
- 1.14) T We usually solve CSPs by recursively backtracking.
- 1.15) F Understanding complicated visual scenes is a just trivial problem of AI.

Q2. Score: **10.0**

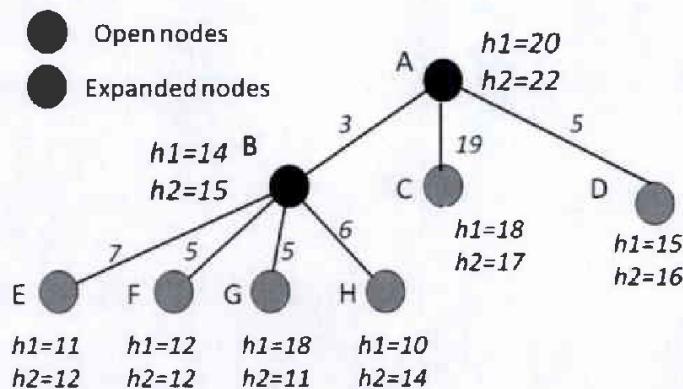
F6F5A850-2281-40B2-BF59-12432671F692

561-F14-Final

#93 3 of 12



## 2. [10%] Search



In the figure above, you are shown a partially expanded search tree with edge costs and two admissible heuristic estimates  $h_1$  and  $h_2$  for the distance between each node and a goal state that is not shown. Ties between nodes for all algorithms are broken by selecting the node that is first in alphabetic order.

2.1 [4%]: What is the next node that is expanded if only heuristic estimate  $h_2$  is used with:

1. [2%] Best First Search: G
2. [2%] A\* Search: G

2.2 [2%]: Suppose that  $h_2$  proves to be an inadmissible heuristic estimate for some reason. What is the potential impact upon the result obtained by the A\* algorithm that uses  $h_2$ ?

may be not optimal.

2.3 [2%]: From the information in the problem statement, can you deduce whether  $h_1$  or  $h_2$  is the better heuristic estimate for this search application? Why or why not?

No, some node  $h_1 > h_2$  and some  $h_1 < h_2$   
so can not deduce.

2.4 [2%]: Given that both  $h_1$  and  $h_2$  are admissible heuristics, can you construct a specific way to combine the heuristics  $h_1$  and  $h_2$  to yield the best heuristic possible, given the information above?

$$h' = \min(h_1 + h_2)$$

Q3. Score: **14.0**

99E5B31D-F531-47AB-A5E4-9A580A8A5BE2

561-F14-Final

#93 4 of 12



### 3. [15%] First Order Logic

Consider the following knowledge base:

1. If something is intelligent, it has common sense
2. Not everything that is intelligent understands emotion.
3. If something has common sense, it can play chess.
4. Something can understand emotion and play chess.
5. Siri can not play chess.

**3.1. [7.5%]** Convert the sentences in the above knowledge base into first order logic expressions, using the following predicates:

- $I(x)$ :  $x$  is intelligent
- $C(x)$ :  $x$  can play chess
- $S(x)$ :  $x$  has common sense
- $E(x)$ :  $x$  understands emotion

1.  $\neg I(x) \Rightarrow S(x)$
2.  $\exists x I(x) \wedge \neg E(x)$
3.  $\neg S(x) \Rightarrow C(x)$
4.  $\exists x E(x) \wedge C(x)$
5.  $\neg C(Siri)$

**3.2. [7.5%]** Use the above knowledge base and proof by contradiction with Resolution to prove that "Siri is not intelligent". You will lose points if you use anything else than resolution or if you do not clearly indicate all your substitutions.

Proof by contradiction.

introduce contradiction.

$$I(Siri) \quad I(x) \Rightarrow S(x) \text{ (1).}$$

$$\neg I(Siri) \quad \neg I(x) \Rightarrow \neg S(x) \text{ (2).}$$

$$S(Siri) \quad S(x) \Rightarrow C(x) \text{ (3).}$$

$$C(Siri) \quad C(x) \Rightarrow \neg C(x) \text{ (4).}$$

$$\emptyset$$

$$\therefore \neg I(Siri)$$

## Q4. Not graded

16D51BF6-DD46-4F44-9694-30AC661AA61F

561-F14-Final

#93 5 of 12

**4. [10%] Planning**

Consider the following STRIPS action schema for planning the end of semester:

**LookupExamDate(subject)**

Pre: Enrolled(subject)

Post: HaveExamDate(subject)

**BookFlight()**

Pre: HaveExamDate(Math), HaveExamDate(Physics)

Post: Have(Ticket)

**Study(subject)**

Pre:

Post: Prepared(subject)

**TakeExam(subject)**

Pre: At(School), HaveExamDate(subject), Prepared(subject)

Post: Passed(subject)

**Shop()**

Pre: At(School), Prepared(Math), Prepared(Physics)

Post: Have(Gift)

**Fly()**

Pre: At(School), Have(Ticket)

Post:  $\neg$ At(School),  $\neg$ Have(Ticket), At(Home)

Given the following:

Start: At(School), Enrolled(Math), Enrolled(Physics)

Final: At(Home), Have(Gift), Passed(Math), Passed(Physics)

Show the partial order plan for planning the end of semester. Your answer must be a complete plan and in the form of a graph showing the actions from the Start state to the Final state using the actions above. Your answer must clearly show the precedence relationships between actions and causal links as applicable, allowing the plan executor to have as much choice in the sequencing of actions without violating any causal links.

You should draw the partial order plan on the next page.

Q4. Score: 10.0

10

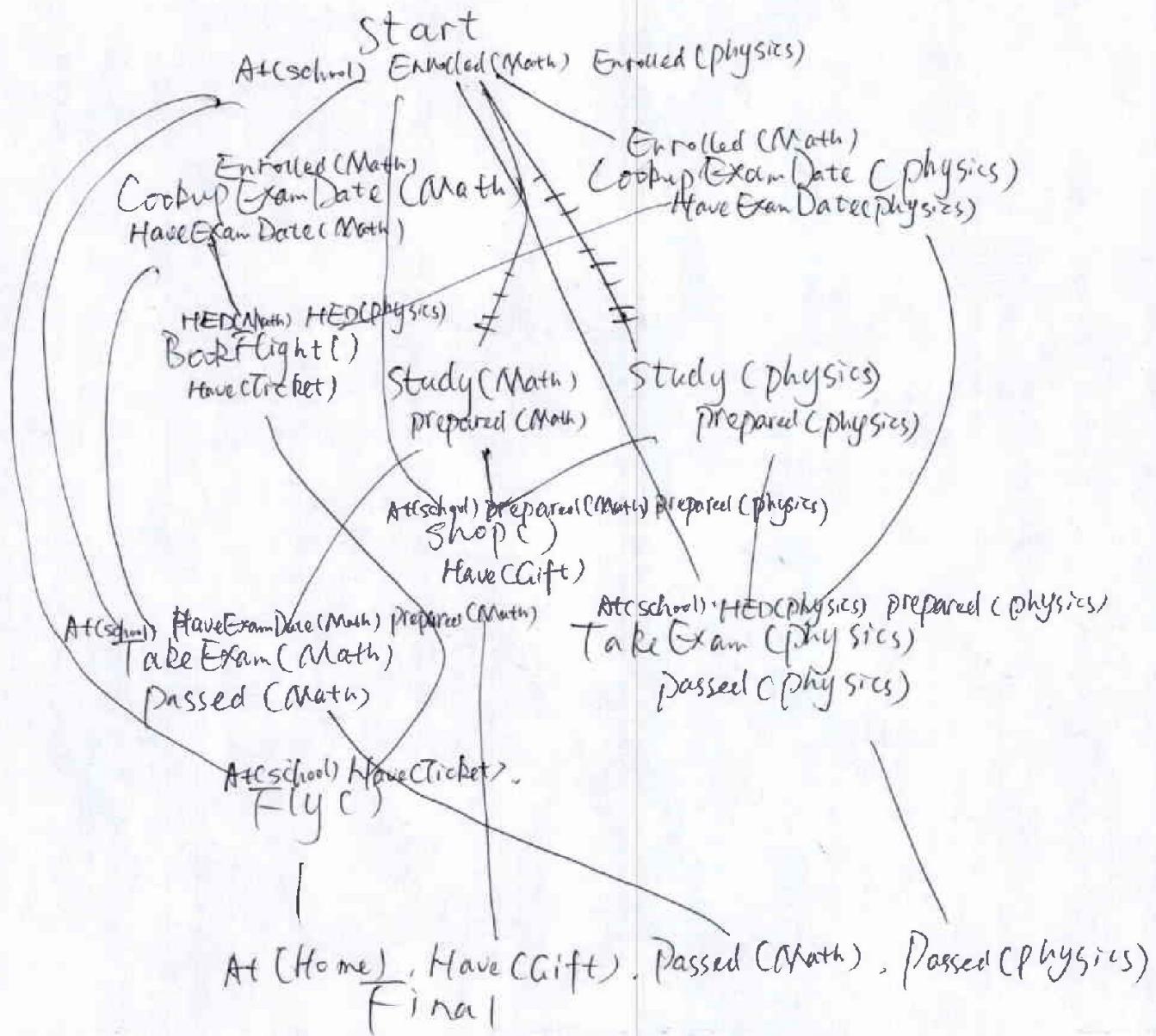
7A89B030-7A3E-4658-B894-242932FB4230

561-F14-Final

#93 6 of 12



Problem 4 (Planning) answer:



Q5. Score: 10.0

19B57C2D-4E20-4205-9887-3C32A6AD140A

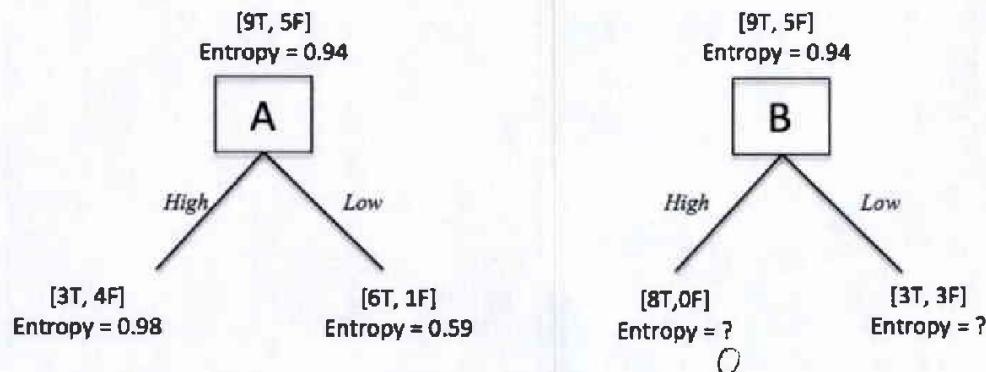
561-F14-Final

#93 7 of 12



## 5. [10%] Decision Tree Learning

Consider the following decision trees for attributes A and B. At each node of the tree,  $[mT, nF]$  specifies that there are  $m$  'True' and  $n$  'False' examples. Entropy values are in bits.



- 5.1 [2%] Fill in the missing entropy values, in bits, for attribute B.

$$\text{left: Entropy} = 0$$

$$\text{Right: Entropy} = -\frac{1}{2} \cdot \lg \frac{1}{2} - \frac{1}{2} \cdot \lg \frac{1}{2} = -\lg \frac{1}{2} = \lg 2 = 1$$

- 5.2 [7%] Calculate the information gain (or change in entropy), in bits, for each attribute.

Please show the steps clearly. You do not need to compute a final numerical result, just write down a formula that uses some of the numbers in the diagram above.

$$IG(A) = \text{Entropy}_A - \text{remainder}$$

$$= 0.94 - \left( \frac{7}{14} \cdot 0.98 + \frac{7}{14} \cdot 0.59 \right) = 0.94 - \frac{1}{2} \cdot 1.57 = 0.94 - 0.785 = 0.155$$

$$IG(B) = \text{Entropy}_B - \text{remainder}$$

$$= 0.94 - \left( \frac{8}{14} \cdot 0 + \frac{6}{14} \cdot 1 \right) = 0.94 - \frac{3}{7} \approx 0.94 - 0.427 = 0.51$$

- 5.3 [1%] Which one of the two attributes A and B is a better classifier according to the decision tree learning (DTL) algorithm?

B, because  $IG(B) > IG(A)$

Q6a. Score: 6.0

4789576C-B2CD-4D89-AE1F-4A6BAE949A2F

561-F14-Final

#93 8 of 12



## 6. [20%] Bayes Net

Suppose John can be very sick (S) that can be caused by two different reasons: getting the seasonal flu and catching a cold (F and C). John thinks that he sometimes catches a cold because he works so hard in his final project that he is exhausted (E) and feels sick. The Bayes' Net and corresponding conditional probability tables are shown below.

For each question, write down the symbolic answer using combinations of  $P(\dots)$  probabilities that are given to you below, e.g., " $P(+e) \times P(-f)$ ", where " $+e$ " means "E is True" and " $-f$ " means "F is False"). You do not need to calculate the numeric answer. You will lose marks if you do not fully develop your answer so that it only uses the probabilities given below.

$P(E)$	
$+e$	0.1
$-e$	0.9

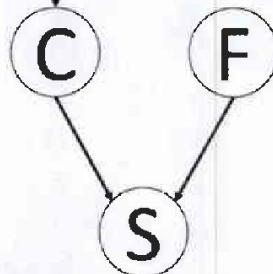


$P(F)$	
$+f$	0.4
$-f$	0.6

$P(C E)$		
$+e$	$+c$	1.0
$+e$	$-c$	0.0
$-e$	$+c$	0.1
$-e$	$-c$	0.9



$P(S C, F)$			
$+c$	$+f$	$+s$	1.0
$+c$	$+f$	$-s$	0.0
$+c$	$-f$	$+s$	0.9
$+c$	$-f$	$-s$	0.1
$-c$	$+f$	$+s$	0.8
$-c$	$+f$	$-s$	0.2
$-c$	$-f$	$+s$	0.1
$-c$	$-f$	$-s$	0.9



(1) [3%] Please write down the joint probability  $P(+s, +c, +f, +e)$ :

$$P(+s, +c, +f, +e) = p(c|es(+c, +f)) \cdot p(+f) \cdot p(+c|+e) \cdot p(+e)$$

↗

(2) [3%] What is the probability of catching a cold? (Hint: Do not refer to that John feels sick)

$$\begin{aligned}
 P(+c) &= \sum_E P(+c|E) \cdot P(E) \\
 &= P(+c|+e) \cdot \overset{P(e)}{+} P(+c|-e) \cdot P(-e) \\
 &= 1.0 \cdot 0.1 + 0.1 \cdot 0.9 = 0.19
 \end{aligned}$$

↗

Q6b. Score: 12.0

9B0C86B3-2AP6-4DA7-959B-1679D685D65A

561-F14-Final

#93 9 of 12



(3) [3%] What is the probability that John catches a cold given that he got the seasonal flu?

$$P(+c|+f) = \frac{P(c|s, f) \cdot P(s)}{P(s, f)} = \frac{P(c, s, f)}{\sum_{e} P(s, e, f) \cdot P(e)}$$

$C$  and  $F$  is independent

(4) [3%] What is the probability of getting exhausted given a cold,  $P(+e|+c)$ ?

$$P(+e|+c) = \frac{P(+c|+e) \cdot P(e)}{P(c)} = \frac{P(e|c)}{P(c)}$$

(5) [3%] What is the probability  $P(+e|+f)$ ?

$$P(+e|+f) = \frac{P(e|f)}{\sum_s P(s|f)} = \frac{P(e|f)}{P(f)} = \frac{P(e)}{P(f)}$$

$E$  and  $F$  is independent.

(6) [5%] What is the conditional probability  $P(+c|+s, +f)$ ?

$$P(+c|+s, +f) = \frac{P(c|s, f)}{P(s, f)} = \frac{P(c, s, f)}{P(s, f) \cdot P(f)}$$

$$= \frac{P(c|s, f)}{P(s|f) \cdot P(f)}$$

$$= \frac{P(+s|+c, +f) \cdot P(+c) \cdot P(+f)}{P(+s|+f) \cdot P(+f)}$$

$$= \frac{P(+s|+c, +f) \cdot P(+c)}{P(+s|+f) \cdot P(+f)} + P(+s|-c, +f) \cdot P(-c)$$

Q7. Score: 10.0

6146AB3C-B5BB-43F3-913A-3321DDB7CE11

561-F14-Final

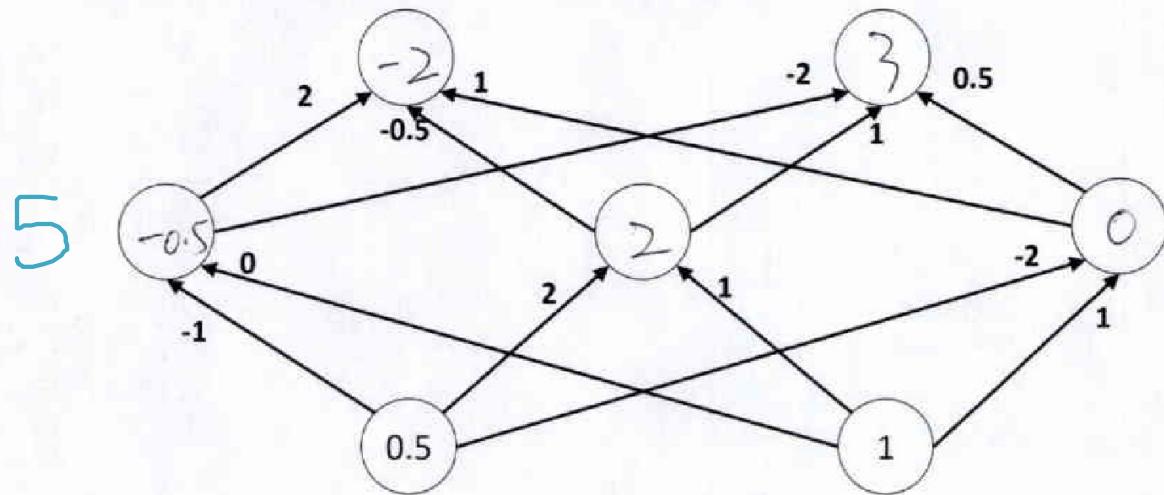
#93 10 of 12



## 7. [10%] Neural Networks

The following is a network of linear neurons, that is, neurons whose output is identical to their net input. The numbers in the circles indicate the output of a neuron, and the numbers at connections indicate the value of the corresponding weight.

7.1. [5%] Compute the output of the hidden-layer and the output-layer neurons for the given input  $(0.5, 1)$  and enter those values into the corresponding circles.



7.2. [3%] What is the output of the network for the input  $(1, 2)$ , i.e. the left input neuron having the value 1 and the right one having the value 2?

3

$$\begin{array}{l} 1 \times -1, \quad 1 \times 2 + 2 \times 1, \quad 1 \times -2 + 2 \times 1 \\ -1 \qquad \qquad \qquad 4 \qquad \qquad \qquad 0 \end{array} \text{ output } (-4, 6)$$

$$-1 \times 2 + 4 \times -0.5 \quad -1 \times -2 + 4 \times 1$$

7.3. [2%] To answer 7.2, do you have to do all the network computations once again? Explain why you do or do not have to do this.

2  
No, because the ~~network~~<sup>n neurons</sup> is linear,  
and  $(1, 2)$  is 2 times  $(0.5, 1)$  so  
the output will be 2 times old one  $(-2, 3)$   
 $= (-4, 6)$  too.

Q8. Score: **6.0**

BE560AAC-80CF-44D1-BBA8-FA4DC8DF730D

561-F14-Final

#93 11 of 12

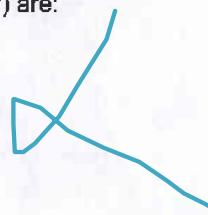


## 8. [10%] Applications

Circle the best choice for each question:

(a) [2%] Markov Decision Processes (MDP) are:

- a. A form of knowledge sharing
- b. Non-monotonic
- c. An ontology for multiple inheritance
- d. All of the above
- e. None of the above



(b) [2%] Occam's Razor:

- a. Passes the Turing Test
- b. Prefers smaller Decision Trees that generalize
- c. Causes overfitting in Neural Networks
- d. All of the above
- e. None of the above



(c) [2%] Machine learning is:

- a. Science
- b. Neat
- c. Rational
- d. All of the above
- e. None of the above



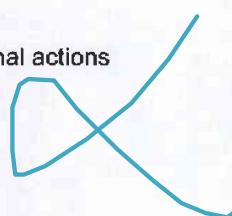
(d) [2%] For efficient real world applications, Knowledge Engineering is:

- a. An inexpensive process
- b. Not allowed to use entailment
- c. Must be both Neat and Scruffy
- d. All of the above
- e. None of the above



(e) [2%]: Bayesian Networks:

- a. Use probability to quantify ignorance
- b. Have a bias for the simplest hypothesis
- c. Calculate full joint probabilities for rational actions
- d. All of the above
- e. None of the above



Q11. Not graded

ACF37B16-DFED-4A18-90DE-A8D0E355735D

561-F14-Final

#93 12 of 12

