CS540 ANSWER SHEET

Name ____ Email ____________________

1.	2.
3.	4.
5.	6.
7.	8.
9.	10.
11.	12.
13.	14.
15.	16.
17.	18.
19.	20.

Final Examination CS540-1: Introduction to Artificial Intelligence Fall 2014
20 questions, 5 points each

INSTRUCTIONS: Choose ONE answer per question. WRITE YOUR ANSWERS ON THE ANSWER SHEET. WE WILL NOT GRADE ANSWERS ON OTHER PAGES. ON THE ANSWER SHEET WRITE DOWN THE ANSWERS ONLY – DO NOT INCLUDE INTERMEDIATE STEPS OR DERIVATIONS. BE SURE TO INCLUDE YOUR NAME AND EMAIL ON THE ANSWER SHEET, TOO.

1. Consider Iterative Deepening Search on a tree, where the nodes are denoted by letters. Without knowing the tree, which of the following **cannot** be the sequence of nodes expanded by iterative deepening search?

   (A) SABSACD... (B) SABSABCD... (C) SASAB... (D) SABCD...
   (E) they all can

   E: different tree structures

2. Consider a 3-puzzle where, like in the usual 8-puzzle game, a tile can only move to an adjacent empty space. Given the initial state

1	2
	3

   which of the following state **cannot** be reached?

   (A)

3	1
	2

   (B)

	3
2	1

   (C)

1	3
	2

   (D)

	2
1	3

   (E) they all can

   C: odd vs. even permutation (or just enumerate)

3. To specify the joint probability distribution in any directed graphical model on n binary variables, how many conditional probability tables are needed?

   (A) n (B) 2^n (C) n^2 (D) 2 (E) none of the above

   A: this is the number of tables, not the entries

4. Given the following game matrix, suppose A knows that B will use the mixed strategy (1/3, 2/3, 0) on B-I, B-II, B-III. What is the expected payoff for A if A plays optimally?

	B-I	B-II	B-III
A-I	0	-1	1
A-II	1	0	-1
A-III	-1	1	0

   (A) 2/3 (B) -2/3 (C) 1/3 (D) -1/3 (E) none of the above

   C: convex combinations of AII and AIII are optimal in this case

5. Identify the pure strategy Nash equilibrium in the following zero-sum game. A is the max player and B is the min player.

	B-I	B-II	B-III
A-I	-5	0	6
A-II	-3	-6	-2
A-III	10	2	8

(A) A-I, B-III (B) A-III, B-I (C) A-II, B-III (D) A-III, B-II
(E) none of the above

D: by local maximum definition

6. We define a new Boolean logic connective ♡ as follows:

P	Q	P ♡ Q
F	F	T
F	T	F
T	F	F
T	T	F

Which of the following is equivalent to P ∨ Q?

(A) (P ♡ Q) ♡ (P ♡ Q) (B) (P ♡ P) ♡ (Q ♡ Q)

(C) (P ♡ P) ♡ Q (D) (Q ♡ Q) ♡ P (E) none of the above

A: NOR implements OR

7. Characterize $P \Rightarrow ((Q \vee R) \Rightarrow P)$.

(A) Unsatisfiable (B) Tautology (C) Satisfiable but not tautology

(D) Not a propositional logic sentence (E) None of the above

B: e.g. convert to CNF

8. Convert to Conjunctive Normal Form: $(P \Rightarrow (Q \Leftrightarrow R))$

(A) $(\neg P \vee \neg Q \vee \neg R) \wedge (P \vee \neg Q \vee \neg R)$
(B) $(\neg P \vee Q \vee R) \wedge (\neg P \vee \neg Q \vee \neg R)$
(C) $(\neg P \vee \neg Q \vee R) \wedge (\neg P \vee Q \vee \neg R)$
(D) $(P \vee Q \vee R) \wedge (\neg P \vee \neg Q \vee \neg R)$
(E) none of the above

C: by definition

9. Consider the following directed graphical model over binary variables:

$A \rightarrow B \quad C$

Note C is disconnected. Given the CPTs:

$P(A = T) = 0.2$, $P(B = T \mid A = T) = 0.1$, $P(B = T \mid A = F) = 0.5$, $P(C = T) = 0.4$, compute $P(B = F \mid C = T)$.

(A) 0.6 (B) 0.232 (C) 0.58 (D) 0.24 (E) none of the above

C: C and B are independent. But you can also just derive it from the definition of directed graphical models.

10. Assume the prior probability of having a female child is the same as having a male child, both are 0.5. The Smith family has two kids. One day you saw one of the Smith children, and she is a girl. The Wood family has two kids, too, and you heard that at least one of them is a girl. What is the chance that the Smith family has a boy? What is the chance that the Wood family has a boy?

    (A) 1/2, 1/2 (B) 1/2, 2/3 (C) 2/3, 1/2 (D) 2/3, 2/3
    (E) none of the above

    B: the Wood observation only eliminates (boy,boy)

11. Which one is the translation of "John has exactly one brother"?

    (A) $\exists x, y\ brother(John, x) \land brother(John, y) \land x = y$
    (B) $\exists x\ brother(John, x) \Rightarrow \forall y (brother(John, y) \land x = y)$
    (C) $\exists x\ brother(John, x) \Rightarrow \forall y (brother(John, y) \Rightarrow x = y)$
    (D) $\forall x\ brother(John, x) \Rightarrow \exists y (brother(John, y) \land x = y)$
    (E) none of the above

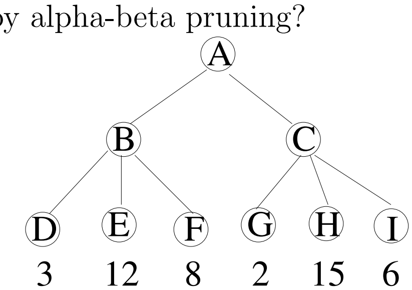
    E: $\exists x\ brother(John, x) \land \forall y (brother(John, y) \Rightarrow x = y)$

12. Let $v(x)$ mean x is a vegetarian, $m(y)$ for meat, and $e(x, y)$ for x eats y. The following sentences are all equivalent to each other **except**:

    (A) $\forall x\ v(x) \Leftrightarrow (\forall y\ e(x, y) \Rightarrow \neg m(y))$
    (B) $\forall x\ v(x) \Leftrightarrow (\neg(\exists y\ m(y) \land e(x, y)))$
    (C) $\forall x\ (\exists y\ m(y) \land e(x, y)) \Leftrightarrow \neg v(x)$
    (D) $\forall x\ (\neg(\forall y\ m(y) \Rightarrow \neg e(x, y))) \Leftrightarrow \neg v(x)$
    (E) No exception, they are all equivalent

    E

13. Which nodes will be pruned by alpha-beta pruning?



    (A) I (B) HI (C) GHI (D) CHI (E) none of the above

    B: after seeing 2, pruning kicks in

14. In simulated annealing one accepts a transition $s \to t$ with probability $\exp\left(-\frac{f(s)-f(t)}{T}\right)$ if $f(t) \leq f(s)$, where T is the temperature parameter. Assume that two states t, r are both worse than s, and the transition probability $s \to t$ at temperature T_1 equals the

transition probability $s \to r$ at a cooler temperature $T_2 = \frac{1}{2}T_1$. What is the relation between $f(s), f(t), f(r)$?

(A) $f(r) = \log(\exp(f(s)) + \exp(f(t)))$
(B) $f(r) = \frac{f(s)+f(t)}{2}$
(C) $f(r) = \sqrt{f(s)f(t)}$
(D) $f(r) = \frac{f(s)f(t)}{f(s)+f(t)}$
(E) none of the above

B: simple math

15. Consider A* search on the following grid, with initial state A and goal state G, and one can move left, right, up, or down one step at a time (no wrapping around). The cost g is the number of moves taken, and the heuristic h is the Manhattan distance to G.

A	B	C
D	E	F
G	H	I

At the moment that A* declares success, which states remain in OPEN?

(A) ABDEG (B) ABDE (C) ABE (D) BE
(E) none of the above

D: run the algorithm

16. The sigmoid function in a neural network is defined as

$$g(x) = \frac{e^x}{1+e^x}.$$

There is another commonly used activation function called the hyperbolic tangent function, which is defined as

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

How are these two functions related?

(A) $tanh(x) = g(x) - 1$ (B) $tanh(x) = 2g(x) - 1$
(C) $tanh(x) = g(2x) - 1$ (D) $tanh(x) = 2g(2x) - 1$ (E) none of the above

D: simple math

17. Consider the linear SVM problem without slack variables or kernels: this is known as the hard margin SVM. If you give it a linearly separable training data set where $x_1, \ldots, x_n \in \mathbb{R}^d$ and $y_1 \ldots y_n \in \{-1, 1\}$, it will learn a hyperplane in $\mathbb{R}^d$. Tom did something to your data set, and hard margin SVM no longer works on the modified data set. What might have Tom done?

(A) $x_i \leftarrow x_i + c$ for a fixed $c \in \mathbb{R}^d$ and $i = 1 \dots n$
(B) $x_i \leftarrow a x_i$ for a fixed $a \in \mathbb{R}$ and $i = 1 \dots n$
(C) rotated the data set in R^d around the origin
(D) swapped the 1st and 2nd coordinates of each point $x_{i1} \rightleftharpoons x_{i2}$ for $i = 1 \dots n$

B: only when $a = 0$

18. A traffic light repeats the following cycle: green 8 seconds, yellow 4 seconds, and red 4 seconds. To a car that arrives at the light at a random time, what is the entropy of the light signal?

(A) 1 bit (B) 2/3 bits (C) 3/2 bits (D) 2 bits (E) none of the above

C: p=(1/2, 1/4, 1/4)

19. For kNN on a fixed training set with n items, if one increasing k from 1 gradually to n, which of the following description is **impossible about training set accuracy**?

(A) Attain maximum at $k = 1$
(B) Attain maximum at $k > 1$ but not $k = 1$
(C) Constant for all k

B: by definition 100 accuracy at k=1. (C if all training items have the same label)

20. Consider three 2D points $a = (0, 0), b = (0, 1), c = (1, 0)$. Run k-means with two clusters. Let the initial cluster centers be $(-1, 0), (0, 2)$. What clusters will k-means learn?

(A) $\{a\}, \{b, c\}$ (B) $\{a, b\}, \{c\}$ (C) $\{a, c\}, \{b\}$ (D) none of the above

C