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 2016
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 uninformed search over a tree, where the root is at level 0, each internal node has k children, and all leaves are at level D. There is a single goal state somewhere at level d, where $0 \le d \le D$. When is DFS definitely going to goal-check more nodes than BFS?

   (A) always (B) never (C) when $k > d$ (D) when $k > D$ (E) none of the above

   Due to the ambiguity of natural language, we will accept both B and E: it depends on where the goal is on level d

2. Recall in uniform-cost search, each node has a path-cost from the initial node (sum of edge costs along the path), and the search expands the least path-cost node first. Consider a search graph with $n > 1$ nodes (n is an even number): 1, 2, $\ldots$, n. For all $1 \le i < j \le n$ there is a directed edge from node i to node j with an edge cost j. The initial node is 1, and the goal node is n. How many goal-checks with uniform-cost search perform?

   (A) 2 (B) $n/2$ (C) n (D) $n(n+1)/2$ (E) none of the above

   C: all nodes will be expanded

3. X, Y are both Boolean random variables taking value 0 or 1. If $P(X = 1) = 1/2$, $P(Y = 1) = 1/3$, $P(X = 1 \mid Y = 0) = 1/4$, what is $P(X = 1 \mid Y = 1)$?

   (A) 1/6 (B) 1/4 (C) 3/4 (D) 1
   (E) undetermined because $P(X = 1 \mid Y = 0)$ and $P(X = 1 \mid Y = 1)$ do not sum to one.

   D:

```
    X=0 X=1
Y=0 1/2 1/6| 2/3
Y=1  0  1/3| 1/3
-----------
    0.5 0.5

p(x|y) = p(x,y)/p(y) = p(y|x)p(x)/p(y) = (1-p(!y|x))p(x)/p(y) = (1-p(x|y!)p(y!)/p(
p(x=1|y=1)
=p(y=1|x=1)p(x=1)/p(y=1)
```

```
=(1-p(y=0|x=1))p(x=1)/p(y=1)
=(1-p(x=1|y=0)p(y=0)/p(x=1) )p(x=1)/p(y=1)
=(1-1/4 * 2/3 *2) * 1/2 * 3
=(1-1/3)*3/2
= 2/3 * 3/2 = 1
```

4. Mary rolled a loaded six-sided die many times and recorded the number of outcomes: 1: zero times, 2: one time, 3: two times, 4: three times, 5: four times, 6: five times. Mary estimates the probability of each face with add-1 smoothing. What is her estimated probability for face 3?

   (A) 2/15 (B) 1/5 (C) 3/16 (D) 1/7 (E) none of the above

   D. $(1+2)/(6+1+2+3+4+5) = 1/7$

5. Consider a state space where each state is represented by a triple of integers (x_1, x_2, x_3), where $x_1, x_2, x_3 \in \{0, 1, 2, \ldots, 10\}$. The neighbors of a state (x_1, x_2, x_3) are constructed as follows. Pick an i from 1,2,3 then change x_i by one (add or subtract, as long as the resulting number is within the range). For example, $(x_1, x_2 - 1, x_3)$ is a neighbor if $x_2 - 1 \geq 0$. Therefore, most states will have 6 neighbors. The score of a state is defined as $f(x_1, x_2, x_3) = x_1 - 2x_2 - 4x_3$. Starting from the state $(5, 5, 5)$, perform simple hill climbing to maximize the score. Which state will hill climbing end up with?

   (A) (5,5,5) (B) (10,0,0) (C) (5,4,5) (D) (5,0,5) (E) none of the above

   B. This is the global maximum, and the 3D grid allows any starting point to greedily reach it

6. Consider a board with 4 positions in a row: $o\sqcup\sqcup G$. There is a piece 'o' at position 1, and the goal is at position 4 (indicated by 'G'). Two players play in turns: A goes first, B goes second. Each player must move the piece one position to the left or right on the board. The player who moves the piece to 'G' wins. Who will win, and after how many total moves from both players?

   (A) A, 3 (B) A, 5 (C) B, 4 (D) B, 7 (E) none of the above

   E. Infinite game

7. Perform iterated elimination of strictly dominated strategies. Player A's strategies are the rows. The two numbers are (A,B)'s payoffs, respectively. Recall each player wants to maximize their own payoff.

0,2	3,1	2,3
1,4	2,1	4,1
2,1	4,4	3,2

   What is left over?

   (A) (2,3) (B) (3,1) (C) (4,4) (D) multiple cells (E) none of the above

   D: after removing the first row, we don't have strict domination.

8. Consider Euclidean distance in $\mathbb{R}$. There are three clusters: $A = (0, 2, 6)$, $B = (3, 9)$, $C = (11)$. Which two clusters will single linkage and complete linkage merge next, respectively?

   (A) AB and AB (B) AB and AC (C) AB and BC (D) AC and AC (E) none of the above

   C.
   single: AB=1, AC=5, BC=2, so merge AB
   complete: AB=9, AC=11, BC=8, so merge BC

9. Perform k-means clustering on six points $x_i = i$ for $i = 1 \ldots 6$ with two clusters. Initially the cluster centers are at $c_1 = 1, c_2 = 6$. Run k-means until convergence. What is the reduction in distortion?

   (A) 2 (B) 4 (C) 6 (D) 8 (E) none of the above

   C. The clustering stays the same but the centers move. (1 2 3) (4 5 6)
   (0+1+4)*2=10
   (1+0+1)*2=4

10. List English letters from A to Z. Define the distance between two letters in the natural way, that is, d(A,A)=0, d(A,B)=1, d(A,C)=2 and so on. Each letter has a label: vowel or consonant. For our purpose, A,E,I,O,U are vowels and others are consonants. This is your training data. Now classify each letter using kNN for odd $k = 1, 3, 5, 7, \ldots$. What is the smallest k where all letters are classified the same?

    (A) 1 (B) 3 (C) 5 (D) 25 (E) none of the above

    B: -BCD-FGH-JKLMN-PQRST-VWXYZ

11. Consider a training set with n items. Each item is represented by a d-dimensional feature vector $(x_1, \ldots, x_d)$, and a class label y. However, each feature dimension is continuous (i.e. it is a real number). To build a decision tree, one may ask questions of the form "Is $x_1 \geq \theta$?" where θ is a threshold value. Ideally, how many different θ values should we consider for the first dimension x_1?

    (A) Infinite, because x_1 is continuous
    (B) As many as floating point numbers representable by a computer
    (C) nd, all possible numbers in the training set
    (D) One, at the middle of x_1's range
    (E) none of the above

    E: there are at most n distinct values that x_1 can take in the training set. These divide the infinite θs into at most $n+1$ equivalence classes (each training item will answer all questions in an equivalence class in the same way). So we only needed to consider at most $n+1$ values.

12. Consider three Boolean variables where $R = (P \wedge \neg Q) \vee (\neg P \wedge Q)$. Suppose P has probability 1/2 to be true, and independently Q also has probability 1/2 to be true. What is the mutual information $I(P; R)$?

(A) 0 (B) 1/4 (C) 1/2 (D) 1 (E) none of the above

A: R = P XOR Q

13. For simple linear regression with training data $(x_1 = 5, y_1 = -3), (x_2 = 1, y_2 = -3)$, what is the coefficient of determination?

(A) 0 (B) 1 (C) -3 (D) undefined (E) none of the above

D: the least squares fit is a constant function y=-3 itself, so we get a 0/0.

14. Learning the w's in which of the following functions can be posed as linear regression?

(A) $f(x) = w_0 + w_1 x + w_2 x^2 + w_3 x^3$
(B) $f(x) = w_0 + w_1 \sin(x) + w_2 \cos(x)$
(C) $f(x) = w_0 x^{w_1} (1-x)^{w_2}$ for $x \in (0, 1)$
(D) all of A,B,C
(E) none of A,B,C

D. For (c) take the log on both sides so it's linear in w_1, w_2. Also we can do a change of variable on $\log(w_0)$

15. Let $f(\sum_{i=1}^d w_i x_i) = 1/(1 + e^{-\sum_{i=1}^d w_i x_i})$ be a sigmoid perceptron with inputs $x_1 = \ldots = x_d = 0$ and weights $w_1 = \ldots = w_d = 1$. There is no constant bias input of one. If the desired output is $y = 1/4$, and the sigmoid perceptron update rule has a learning rate of $\alpha = 1$, what will happen after one step of update?

(A) $x_1, \ldots, x_d$ will decrease
(B) $x_1, \ldots, x_d$ will increase
(C) $w_1, \ldots, w_d$ will decrease
(D) $w_1, \ldots, w_d$ will increase
(E) none of the above

E: w stays the same because the gradient is zero due to x's being zero

16. What is the minimum number of linear threshold units needed to implement the NOR function? Recall P NOR $Q = \neg(P \vee Q)$ for Boolean inputs P, Q.

(A) 1 (B) 2 (C) 3 (D) 4 (E) none of the above

A: the decision boundary is linear in the (P,Q) space

17. You have a joint probability table over k random variables $X_1, \ldots, X_k$, where each variable takes m possible values: 1, ..., m. To compute $P(X_1 = m)$, how many cells in the table do you need to access?

(A) mk (B) m^k (C) k^m (D) $\frac{m!}{k!(k-m)!}$ (E) none of the above

E: m^{k-1}

18. Consider a classification problem with 10 classes $y \in \{1, 2, \ldots, 10\}$, and two binary features $x_1, x_2 \in \{0, 1\}$. Suppose $p(Y = y) = 1/10$, $p(x_1 = 1 \mid Y = y) = y/10$, $p(x_2 = 1 \mid Y = y) = y/540$. Which class will naive Bayes classifier produce on a test item with $(x_1 = 0, x_2 = 1)$?

(A) 1 (B) 3 (C) 5 (D) 8 (E) 10

C

$$\begin{aligned}
&\text{argmax}_y P(y \mid x_1 = 0, x_2 = 1) && (1)\\
&= \text{argmax}_y P(x_1 = 0, x_2 = 1 \mid y)p(y) && (2)\\
&= \text{argmax}_y P(x_1 = 0 \mid y)P(x_2 = 1 \mid y)p(y) && (3)\\
&= \text{argmax}_y (1 - P(x_1 = 1 \mid y))P(x_2 = 1 \mid y)p(y) && (4)\\
&= \text{argmax}_y (1 - y/10)y/540 && (5)\\
&= \text{argmax}_y (10 - y)y && (6)
\end{aligned}$$

Now you could enumerate y to find the maximum. Or pretend y is continuous, taking derivative and set to zero: 10-2y=0 leads to y=5.

19. With three Boolean variables P, Q, R, under how many interpretations is $(P \vee Q \vee R) \Rightarrow (P \Rightarrow P)$ true?

(A) 0 (B) 3 (C) 7 (D) 8 (E) none of the above

D. Note $P \Rightarrow P$ is valid.

20. $\neg(\alpha \vee \beta)$ entails which sentence?

(A) $\beta \Rightarrow \neg\alpha$ (B) $\alpha \Rightarrow \beta$ (C) $\alpha \Rightarrow \neg\beta$ (D) all of A,B,C (E) none of A,B,C

D. The interpretation is when α, β are both false, then all implications are true.