CS161: Fundamentals of Artificial Intelligence
Final Exam
Spring 2022

Name	ID Number

.

[Q1. (30 pts)] True or false?

1. Convolutional Neural Networks (CNNs) are usually used for vision applications and Recurrent Neural Networks (RNNs) are usually used in language applications.

2. Depth first search is typically implemented using a queue while breadth first search is implemented using a stack.

3. One of the main advantages of hill climbing search is its space requirements.

4. Uniform-cost search with a cost function $g(n) = depth(n)$ is equivalent to breadth first search.

5. Depth first search is a complete search strategy.

6. Breadth first search is an optimal but incomplete search strategy.

7. The *Minimax* algorithm is usually implemented using depth-first search.

8. $\alpha\beta$-pruning can potentially double the depth we can search to (in the same amount of time).

9. If we know the exact depth of an optimal solution, then limited–depth search is better than depth-first search.

10. The minimax procedure is guaranteed to compute an optimal move against any player.

11. *Resolution* is guaranteed to detect a contradiction in any knowledge base if one exists.

12. $(X \vee Y) \vee (\neg X \vee Y \vee \neg Z)$ is a conjunctive normal form (X, Y and Z are variables).

13. $A \vee B \vee \neg C$ is a Horn clause (A, B and C are variables).

14. Supervised learning is used with labeled data and unsupervised learning is used with unlabeled data.

15. A sentence is valid iff its negation is consistent.

16. Any propositional logic sentence can be represented by a CNF.

17. If $\alpha \models \beta$, then $Pr(\alpha) < Pr(\beta)$.

18. $\forall x \exists y Likes(x, y)$ is equivalent to $\exists x \forall y Likes(x, y)$.

19. $\forall x Nice(x)$ is equivalent to $\neg \exists x \neg Nice(x)$.

20. Any sentence in first-order logic can be expressed without using the existential $\exists$ quantifier.

21. In Bayesian networks, MAP queries are a special type of MPE queries.

22. If X and Y are probabilistically independent, then they must continue to be independent given any variable Z.

23. $Pr(\alpha) = Pr(\alpha \wedge \beta) + Pr(\alpha \wedge \neg\beta)$ for any events α and β.

24. $Pr(\alpha|\beta)Pr(\beta) = Pr(\beta|\alpha)Pr(\alpha)$ for any events α and β.

25. Neural networks are universal function approximators.

26. One can count the models of an NNF circuit that is decomposable, deterministic and smooth in linear time.

27. Decision trees are typically used for unsupervised learning using unlabeled data.

28. When learning the parameters of a Bayesian network from complete data, the maximum-likelihood parameter estimates are unique.

29. When learning a Bayesian network structure, one seeks a network that maximizes the probability of data.

30. The naive Bayes structure assumes that attributes are pairwise independent.

[Q2. (20 pts)] True or false?

31. If the branching factor $b = 1$, then iterative deepening will expand $\Theta(d^2)$ nodes, where d is the depth of an optimal solution.

32. The inputs and outputs of a neural network are restricted to be in $[0, 1]$.

33. If $\Delta \models \alpha$ or $\Delta \models \neg\alpha$ for any α, then Δ must have a single model (satisfied by one world).

34. A unifier exists for $P(F(B), F(G(w)), w)$ and $P(F(y), F(y), B)$.

35. In neural networks, the ReLU activation function is $g(x) = 0$ if $x < 0$ and $g(x) = x$ if $x \geq 0$.

36. The entropy of distribution $Pr(X)$ is $\sum_x Pr(x) \log Pr(x)$.

37. If α can be derived from a knowledge base Δ using some inference rules, then α can be derived from Δ' using the same rules, where Δ' is a larger knowledge base that includes Δ.

38. In Figure 1, and assuming all nodes are binary, the CPT for node S has 8 parameters.

39. In Figure 1: W and X are d-separated by R.

40. In Figure 1: W and T are independent given Y and Z.

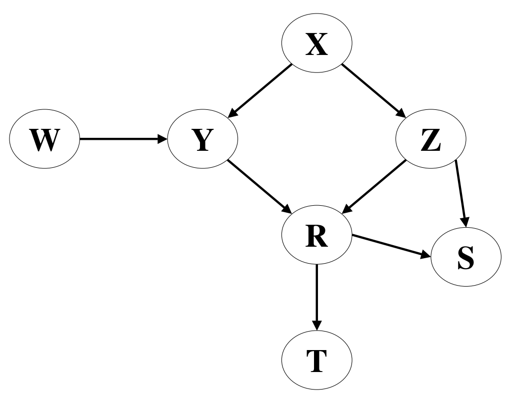

Figure 1: The structure of a Bayesian network.

[Q3. (20 pts)] Choose only one answer.

41. A sibling is another child of one's parent. Which of the following sentences reflects this fact:

    (a) $\forall x, y\ \ Sibling(x, y) \Leftrightarrow (x \neq y \wedge \forall p\ \ Parent(p, x) \wedge Parent(p, y))$.

    (b) $\exists x, y\ \ Sibling(x, y) \Leftrightarrow (x \neq y \wedge \exists p\ \ Parent(p, x) \wedge Parent(p, y))$.

    (c) $\forall x, y\ \ Sibling(x, y) \Leftrightarrow (x \neq y \wedge \exists p\ \ Parent(p, x) \wedge Parent(p, y))$.

    (d) $\forall x, y\ \ Sibling(x, y) \Leftrightarrow (\exists p\ \ Parent(p, x) \wedge Parent(p, y))$.

    (e) None of the above.

42. The sentence $\exists x Person(x) \wedge Nice(x) \wedge (\forall y (Person(y) \wedge Nice(y)) \Rightarrow x = y)$ says:

    (a) At least one person is nice.

    (b) There is exactly one person who is nice.

    (c) All persons are nice.

    (d) No person is nice.

    (e) None of the above.

43. Resolving $R(F(y)) \vee \neg G(y)$ with $G(A) \vee S(w)$ gives:

    (a) $S(A) \vee R(F(A))$.

    (b) $S(y) \vee R(F(A))$.

    (c) $S(w) \vee R(F(A))$.

    (d) $S(F(A)) \vee R(F(A))$.

    (e) None of the above.

44. The result of dropping quantifiers from $\forall x \exists y Likes(x, y)$ during the process of converting to CNF, gives:

    (a) $Likes(x, y)$.

    (b) $Likes(F(x), y)$.

    (c) $Likes(x, F(y))$.

    (d) $Likes(x, A)$.

    (e) None of the above.

45. The Markovian assumption for Bayesian networks says:

    (a) Every node is independent of its parents given its non-descendants.

    (b) Every node is independent of its descendants given its parents.

    (c) Every node is independent of its non-descendants given its parents.

    (d) Every node is independent of its parents given its descendants.

    (e) None of the above.

46. The EM algorithm is usually used for learning Bayesian networks when:

   (a) Structure is known and data is complete.
   (b) Structure is known and data is incomplete.
   (c) Structure is unknown and data is complete.
   (d) Structure is unknown and data is incomplete.
   (e) None of the above.

47. If a student scores an A+ on CS111 (X), then that student must be exceptional (E) and, hence, will most probably score an A on CS161 (Y). If we want to represent this scenario using a Bayesian network, which of the following causal structures should we use?

   (a) $X \leftarrow E$, $Y \leftarrow E$.
   (b) $X \leftarrow E$, $Y \rightarrow E$.
   (c) $X \rightarrow E$, $Y \leftarrow E$.
   (d) $X \rightarrow E$, $Y \rightarrow E$.
   (e) None of the above.

48. Consider the probability distribution in Table 1, where all missing worlds have probability 0. $Pr(I_1 = 0 | T = 1)$ is:

   (a) 1/4
   (b) 1/2
   (c) 2/3
   (d) 1/6
   (e) None of the above.

I_1	I_2	I_3	I_4	I_5	T	$Pr(.)$
1	0	0	0	0	1	1/12
0	0	0	0	0	0	1/12
0	0	0	0	1	1	3/12
0	0	0	1	0	1	3/12
0	1	0	0	0	1	2/12
0	0	1	0	0	1	2/12

Table 1: Probability distribution.

49. Evaluating the expression (CONS '(A B) (REST (CONS 'B '(C)))) gives:

   (a) '(A B (C)).
   (b) '((A B) C).
   (c) '(A B C).
   (d) '((A B) (B C)).
   (e) None of the above.

50. Consider the function:

```
(DEFUN FOO (L I)
      (COND ((NULL (REST L)) I)
            (T (+ 2 (FOO (REST L) (+ 1 I))))))
```

   The result of evaluating (FOO '(A B C) 2) is:

   (a) 5.
   (b) 6.
   (c) 7.
   (d) 8.
   (e) None of the above.

[Q4. (30 pts)] Choose only one answer.

51. When learning a random forest from data, the attributes of each decision tree contain:
    (a) All attributes in the dataset.
    (b) Only those attributes selected by the decision tree learning algorithm.
    (c) Only those attributes selected by the decision tree learning algorithm from among a randomly selected subset of attributes from the dataset.
    (d) A randomly selected subset of attributes from the dataset.

52. Consider the knowledge base $\Delta = \{X \Rightarrow Y, \neg Z \Rightarrow \neg Y, X \Rightarrow \neg Z\}$. Which of the following sentences is entailed by Δ:
    (a) $X \wedge Z$.
    (b) $X \vee Y$.
    (c) $\neg X$.
    (d) $\neg X \wedge Y$.
    (e) None of the above.

53. The following two sentences:

$$\exists t \, Time(t) \wedge (\forall x Person(x) \Rightarrow Fooled\_At(x, t))$$

$$\forall x Person(x) \Rightarrow (\exists t \, Time(t) \wedge Fooled\_At(x, t))$$

are:
    (a) Equivalent.
    (b) The first implies the second.
    (c) The second implies the first.
    (d) None of the above.

54. The following two sentences:

$$\forall x (\forall y Likes(x, y)) \Rightarrow Nice(x)$$

$$\forall x (\neg Nice(x)) \Rightarrow (\exists y Likes(x, y))$$

are:
    (a) Equivalent.
    (b) The first implies the second.
    (c) The second implies the first.
    (d) None of the above.

55. The CNF of $\neg(\forall x\ \exists y\ (P(x) \Rightarrow Q(x,y)))$ is:

(a) $P(F(A)) \vee \neg Q(F(A), y)$.
(b) $P(F(A)),\ \neg Q(F(A), y)$.
(c) $P(A) \vee \neg Q(A, y)$.
(d) $P(A),\ \neg Q(A, y)$.
(e) None of the above.

56. Consider a Bayesian network with structure $X \leftarrow Z \rightarrow Y$ (X and Y are children of Z). Then $Pr(x, y, z)$ is equal to:

(a) $Pr(x)Pr(y)Pr(z)$.
(b) $Pr(x|z)Pr(y|z)Pr(z)$.
(c) $Pr(x|z)Pr(y|z)$.
(d) $Pr(x)Pr(y)Pr(z|xy)$.
(e) None of the above.

57. Consider a Bayesian network $X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_n$ (a chain with n nodes). Assume that each variable X_i has only two values x_i and $\bar{x}_i$. Then $Pr(x_3 \mid \bar{x}_1)$ is equal to:

(a) $Pr(x_3)$.
(b) $Pr(x_3, \bar{x}_1)Pr(\bar{x}_1)$.
(c) $Pr(x_3|x_2, \bar{x}_1) + Pr(x_3 \mid \bar{x}_2, \bar{x}_1)$.
(d) $Pr(x_3|x_2)Pr(x_2 \mid \bar{x}_1) + Pr(x_3|\bar{x}_2)Pr(\bar{x}_2 \mid \bar{x}_1)$.
(e) None of the above.

58. The prior probability that a person has high cholesterol is 0.5. Maya took two high-cholesterol tests, and both tests came out positive. The tests false positive and false negative rates are 0.1 for both tests. The probability that Maya has high cholesterol is:

(a) 0.64
(b) 40/41
(c) 0.81
(d) 81/82
(e) None of the above.

59. We are learning a decision tree given n training instances, and we want to choose between two attributes A and B to test on next (according to the expected entropy criterion). Each attribute has two outcomes *true* and *false*. Moreover, $A = true$ in $n/2$ instances, and $B = false$ in also $n/2$ instances. There are $n/4$ positive instances given $A = true$ and $n/8$ positive instances given $A = false$. There are $n/3$ positive instances given $B = true$, and $n/4$ positive instances given $B = false$. We then have:

   (a) Attribute A is more informative than B and should be tested on first.
   (b) Attribute B is more informative than A and should be tested on first.
   (c) The two attributes are equally informative.
   (d) We don't have enough information to decide which attribute is more informative.

60. Consider the neural network below and assume that inputs are either 0 or 1. The output of the neural network is 1 iff:

   (a) both inputs are 1.
   (b) some input is 1.
   (c) inputs are both 0 or both 1.
   (d) one input is 1 and the other is 0.
   (e) None of the above.

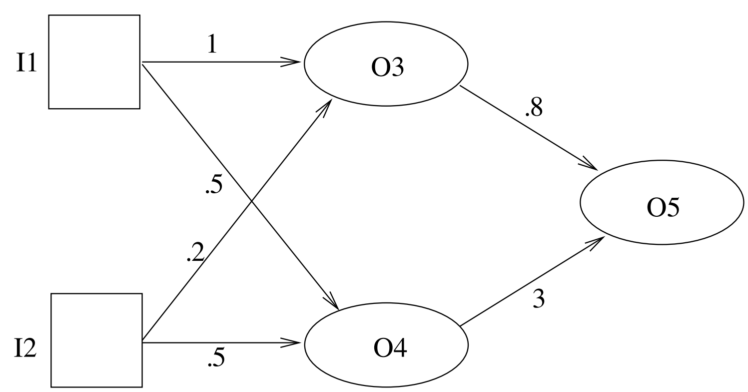

Figure 2: A neural network. All activation functions $g(x)$ are step functions, where $g(x) = 1$ for $x \geq 1$ and $g(x) = 0$ otherwise.