EXAM 1

Midterm 3 Exam

CSCI 561 Fall 2018: Artificial Intelligence

Please print neatly.

First Name:

Last Name:

Student ID:

USC email:

@usc.edu

Instructions:

1. Date: **11/26/2018 from 8:00 pm to 9:50 pm**
2. Maximum credits/points/percentage for this midterm: 100
3. The percentages for each question are indicated in square brackets [ ].
4. **No books** (or any other material) are allowed.
5. Write down your name, student ID and USC email address on both test and answer booklets.
6. Your exam will be scanned and uploaded online.
7. Write within the spaces provided for your answers in the test answer booklet.
8. **Do NOT write on the 2D barcode.**
9. **The back of the exam pages will NOT be graded. You may use it for scratch.**
10. No questions during the exam. If something is unclear to you, write that in your exam.
11. Be brief: a few words are often enough if they are precise and use the correct vocabulary studied in class.
12. When finished before 9:40 pm, raise completed exam sheets until approached by a proctor.
13. At 9:50 pm, put down your pencil. Stay seated and pass your exam to the nearest aisle or **receive 10% exam penalty.**
14. **Adhere to the Academic Integrity code.**

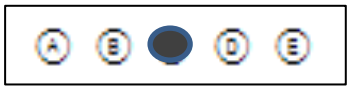
Problems	**100 Percent total**
PART A Short Answer	
AI Applications	4
PART B Multiple Choice, True/False	
General AI Knowledge	8
Bayesian Network	8
Probability	16
Decision Tree	16
MDP and Reinforcement Learning	16
Neural Network	16
Decision Network	16

PART A – Short Answer

AI Applications [4%]

During week 13 there were three presentations about AI applications: Chris Stewart from Google, Ben Marcus from AirMap, and Prof. Wei-min Shen. Choose one of these presentations and briefly describe how the application applies AI to reasoning with uncertainty. (One or two sentences max in the space provide on the answer sheet)

PART B – Multiple Choice & True/False



In this section choose the best option **A**, **B**, **C**, **D**, or **E** and fill in the appropriate circle in the answer sheet. Correct answers receive the % shown in the []. Answers left blank receive 0% and **wrong answers receive -1%**. Fill in at most one circle per problem.

General AI Knowledge: True/False [8%] For each of the statements below, answer **TRUE** if the statement is **always true**, or **FALSE** otherwise.

1. [1%] **A. TRUE B. FALSE** Cross validation helps to avoid overfitting.

2. [1%] **A. TRUE B. FALSE** Unsupervised learning is learning from unlabeled data to maximize a cumulative reward.

3. [1%] **A. TRUE B. FALSE** Support vectors are data points that lie closest to the hyperplane.

4. [1%] **A. TRUE B. FALSE** A decision tree cannot represent all Boolean function.

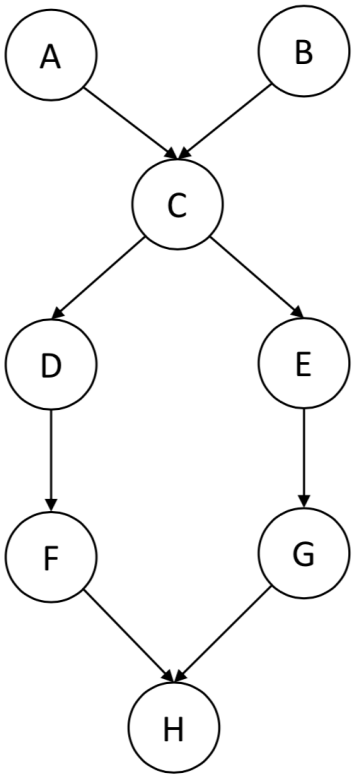
5. [1%] **A. TRUE B. FALSE** Q-learning creates an explicit transition model.

6. [1%] **A. TRUE B. FALSE** A node *X* is conditionally independent of all other nodes in a Bayesian Network given its Markov Blanket.

7. [1%] **A. TRUE B. FALSE** A policy is a mapping from states to actions that tells the agent what action to take in any possible environment state.

8. [1%] **A. TRUE B. FALSE** A Bayesian Network modeling *n* random variables that are all mutually independent of each other would have directed arcs between every pair of variables in the network.

Bayesian Network [8%]



9. [4%] For the Bayesian network given above which statement[s] is true:
   A. A and B are independent given C.
   B. A and B are independent.
   C. D is conditionally independent of G given C.
   D. All of the above.
   E. None of the above.

10. [4%] Which statement[s] is true in any Bayesian Network:
   A. A node is conditionally independent of all other nodes given its Markov blanket.
   B. A node is independent of its non-descendants given its parents.
   C. At least one node in the network must have no incoming arcs.
   D. All of the above.
   E. None of the above.

Probability [16%]

For questions 11 & 12 consider the following joint probability table for the boolean random variables **Q** and **S**:

Q	**S**	
T	T	0.1
T	F	0.2
F	T	0.3
F	F	0.4

11. [4%] Which probability value is the largest?
    A. P(Q)
    B. P(¬ S)
    C. P(Q ∨ S)
    D. P(Q ⇒ S)
    E. P(Q | S)

12. [4%] Which of the following statement[s] is true:
    A. P(Q ⇒ S) > P(S | Q)
    B. P(Q ∨ S) + P(Q , S) = 1
    C. P(¬Q) = 1- P(¬ S)
    D. All of the above.
    E. None of the above.

13. [4%] If random variable *X* is conditionally independent of *Y* given *Z* then
    A. P(X,Y|Z)=P(X|Z)P(Y|Z)
    B. P(X|Y,Z)=P(X|Z)
    C. P(Y|X,Z)=P(Y|Z)
    D. All of the above.
    E. None of the above

14. [4%] Assume that the boolean random variables TEST1 and TEST2 are conditionally independent given FLU and the following probabilities: P(FLU) = 0.1, P(TEST1 | FLU) = 0.9, P(TEST1 | ¬ FLU) = 0.2, P(TEST2 | FLU) = 0.8, P(TEST2 | ¬ FLU) = 0.1

What is the probability of FLU is true given TEST1 and TEST2 are both true:
    A. .25
    B. .79
    C. .0002
    D. All of the above.
    E. None of the above.

Decision Tree [16%] NASA wants to train a robot to recognize two species Martians (M) and Humans (H) based on the following characteristics: Green ∈ *N*,*Y*, *Legs* ∈ 2,3, *Height* ∈ *Short*, *Tall*, Smelly ∈ *N*,*Y*. The available training data is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

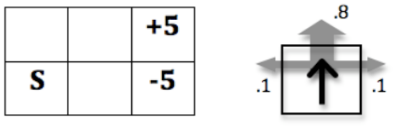
15. [4%] Using the ID3 algorithm to learn a decision tree, the root node chosen would be:
    A. Green
    B. Legs
    C. Both Height and Smelly
    D. All of the above.
    E. None of the above.

16. [4%] Once the learned concept for Martian is an ID3 decision tree, it can be converted into a set of conjunctive rules. For that tree which of following rules would be included:
    A. if (Green=Y and Legs=2 and Height=T and Smelly=N) then M;
    B. if (Green=N and Smelly=Y) then M;
    C. if (Legs=2) then M;
    D. All of the above.
    E. None of the above.

17. [4%] How is Occam's Razor employed in decision tree learning?
    A. ID3 is a greedy approach to construct the shortest tree.
    B. Occam's Razor does not apply to decision tree learning.
    C. Occam's Razor prunes nodes from the tree.
    D. All of the above.
    E. None of the above.

18. [4%] In decision tree learning, the main goal is to:
    A. Minimize the entropy of the training examples associated with each node.
    B. Maximize the entropy of the training examples associated with each node.
    C. Find the best tree structure, such that each training example is a separate tree node.
    D. All of the above.
    E. None of the above.

MDPs and Reinforcement Learning [16%] Given a Gridworld MDP as shown below, the states are grid squares, identified by row then column number. The agent always starts in the state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. The transition function is such that the intended agent movement (North, South, West, or East) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.



(a) (b)

(a) Gridworld MDP. (b) Transition function.

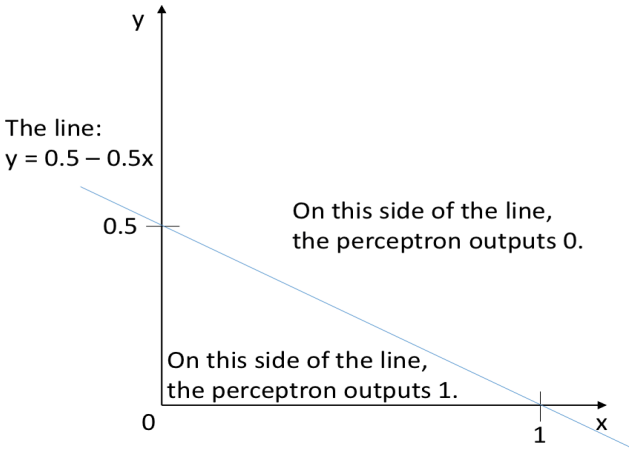
19. [4%] If the agent is deterministic, the optimal policy for this grid is:
    A. State (1,1) North, State (1,2) North, State (2,1) East, State (2,2) East
    B. State (2,1) East, State (2,2) East, State (1,1) North, State (1,2) North
    C. State (1,2) North, State (2,2) East, State (2,1) East, State (1,1) North
    D. All of the above.
    E. None of the above.

20. [4%] Now suppose the agent is stochastic and knows the transition probabilities. Give the second round of value iteration updates for each state, with a discount factor of 0.9 (Assume initial V values are 0 and compute V2 for each non-terminal state).
    A. State (1,1) is 0, State (1,2) is 0, State (2,1) is 0, State (2,2) is 3.924
    B. State (1,1) is 0, State (1,2) is 2.142, State (2,1) is 2.592, State (2,2) 3.924
    C. State (1,1) is 0, State (1,2) is 2.82528, State (2,1) is 2.82528, State (2,2) 3.924
    D. All of the above.
    E. None of the above.

21. [4%] Suppose the stochastic agent does not know the transition probabilities. What does it need in order to learn the optimal policy?
    A. Probability values
    B. Optimal Q values
    C. Training examples of the agent's movement in the grid with rewards received
    D. All of the above.
    E. None of the above.

22. [4%] Reinforcement learning can have problems with scaling because of:
    A. Sparse training data
    B. Calculations taking too much time
    C. It is hard to program
    D. All of the above
    E. None of the above

Neural Networks [16%] Design a perceptron with inputs *x* and *y* and a threshold function as activation function whose output for a given input pair (*x*, *y*) is given by:



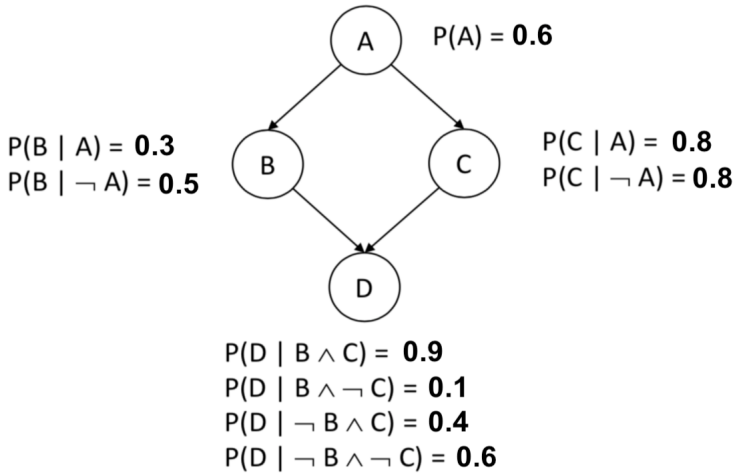
23. [4%] The correct weights **w1** for input *x* and **w2** for input *y* and threshold **t** of the perceptron are:
   A. **w1**= -0.5, **w2**= 1, if **t**=> 0.1 outputs 1 else 0
   B. **w1**= 0.5, **w2**= 0.5, if **t**=>1 outputs 1 else 0
   C. **w1**= 0.5, **w2**= 1, if **t** <= .5 outputs 1 else 0
   D. All of the above.
   E. None of the above.

24. [4%] Given input *x* and *y* many different perceptrons can be created whose output is the line above, by making changes to:
   A. Biases
   B. Activation function
   C. Probability distribution
   D. All of the above
   E. None of the above

25. [4%] How many weights does a 2-layer feed-forward neural network with 5 input units, 3 hidden units and 2 output units contain, including the biases (dummy input weights)?
   A. 42
   B. 24
   C. 22
   D. 9
   E. None of the above

26. [4%] A perceptron is guaranteed to learn a given linearly separable function:
   A. When the activation function is a step function.
   B. Within a finite number of training steps.
   C. If all the weights are positive.
   D. All of the above.
   E. None of the above.

Decision Network [16%] Consider the following Bayesian network. A, B, C, and D are boolean random variables.



27. [4%] Given that A is true and B is false, what is the probability of D being true?
    A. .004
    B. .0034
    C. .36
    D. All of the above.
    E. None of the above.

28. [4%] Converting A into a decision node in the above network and given the following utility function: if D is true then utility is 10 and if D is false utility is 100, which decision for A will give the higher expected utility?
    A. A is true
    B. A is false
    C. Utility is equal in both decisions of A.
    D. All of the above.
    E. None of the above.

29. [4%] In the decision network described in question 28 which of the following events will increase the expected utility the most?
    A. B is true
    B. B is false
    C. C is false
    D. All of the above.
    E. None of the above.

30. [4%] Variable elimination is an algorithm for:
    A. Approximate probabilistic inference in decision networks.
    B. Exact probabilistic inference.
    C. Computing optimal policies.
    D. Reinforcement learning in unknown environments.
    E. None of the above.

EXAM 2

Midterm 3 Exam

CSCI 561 Fall 2018: Artificial Intelligence

Please print neatly.

First Name:

Last Name:

Student ID:

USC email: @usc.edu

Instructions:

14. Date: **11/26/2018 from 8:00 pm to 9:50 pm**
15. Maximum credits/points/percentage for this midterm: 100
16. The percentages for each question are indicated in square brackets [ ].
17. **No books** (or any other material) are allowed.
18. Write down your name, student ID and USC email address on both test and answer booklets.
19. Your exam will be scanned and uploaded online.
20. Write within the spaces provided for your answers in the test answer booklet.
21. **Do NOT write on the 2D barcode.**
22. **The back of the exam pages will NOT be graded. You may use it for scratch.**
23. No questions during the exam. If something is unclear to you, write that in your exam.
24. Be brief: a few words are often enough if they are precise and use the correct vocabulary studied in class.
25. When finished before 9:40 pm, raise completed exam sheets until approached by a proctor.
26. At 9:50 pm, put down your pencil. Stay seated and pass your exam to the nearest aisle or **receive 10% exam penalty.**
14. **Adhere to the Academic Integrity code.**

Problems	**100 Percent total**
PART A Short Answer	
Probabilistic Reasoning	4
PART B Multiple Choice, True/False	
Bayesian Network	8
Probability	16
Decision Network	16
Decision Tree	16
MDP and Reinforcement Learning	16
Neural Network	16
General AI Knowledge	8

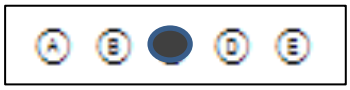
PART A – Short Answer

Probabilistic Reasoning [4%] Assume that the boolean random variables **J** and **K** are conditionally independent given **L** and with the following probabilities:

P(L) = 0.1, P(J | L) = 0.9, P(J | ¬ L) = 0.2, P(K | L) = 0.8, P(K | ¬ L) = 0.1

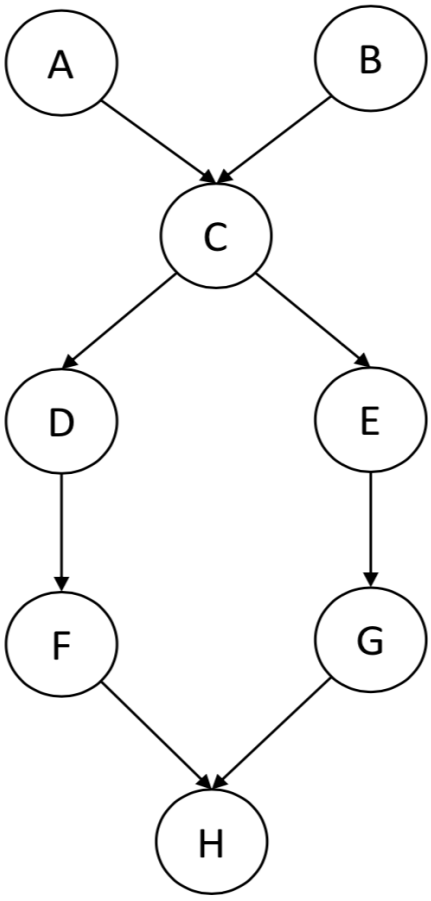
Draw the Bayesian Network in space provide in the answer sheet.

PART B – Multiple Choice & True/False



In this section choose the best option **A**, **B**, **C**, **D**, or **E** and fill in the appropriate circle in the answer sheet. Correct answers receive the % shown in the []. Answers left blank receive 0% and **wrong answers receive -1%**. Fill in at most one circle per problem.

Bayesian Network [8%]



1. [4%] For the Bayesian network given which statement[s] is true:
   A. A and B are independent given C.
   B. A and B are independent.
   C. D is conditionally independent of G given C.
   D. All of the above.
   E. None of the above.

2. [4%] Which statement[s] is true in any Bayesian Network:
   A. A node is conditionally independent of all other nodes given its Markov blanket.
   B. A node is independent of its non-descendants given its parents.
   C. At least one node in the network must have no incoming arcs.
   D. All of the above.
   E. None of the above.

Probability [16%]

For questions 3 & 4 consider the following joint probability table for the boolean random variables **Q** and **S**:

Q	**S**	
T	T	0.1
T	F	0.2
F	T	0.3
F	F	0.4

3. [4%] Which probability value is the largest?
   A. P(Q)
   B. P(¬ S)
   C. P(Q ∨ S)
   D. P(Q ⇒ S)
   E. P(Q | S)

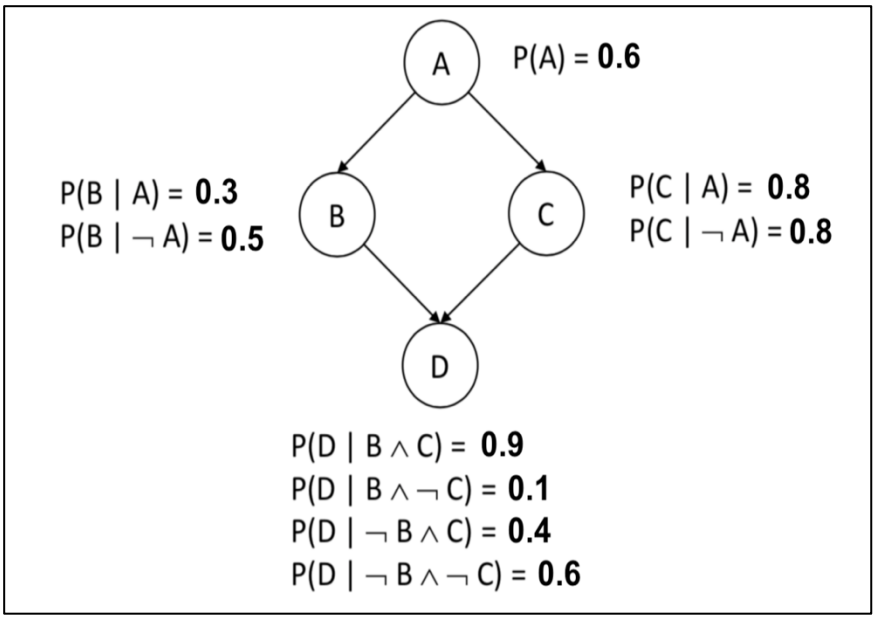
4. [4%] Which of the following statement[s] is true:
   A. P(Q ⇒ S) > P(S | Q)
   B. P(Q ∨ S) + P(Q , S) = 1
   C. P(¬Q) = 1- P(¬ S)
   D. All of the above.
   E. None of the above.

5. [4%] If random variable *X* is conditionally independent of *Y* given *Z* then
   A. P(X,Y|Z)=P(X|Z)P(Y|Z)
   B. P(X|Y,Z)=P(X|Z)
   C. P(Y|X,Z)=P(Y|Z)
   D. All of the above.
   E. None of the above

6. [4%] Assume that the boolean random variables TEST1 and TEST2 are conditionally independent given FLU and the following probabilities: P(FLU) = 0.1, P(TEST1 | FLU) = 0.9, P(TEST1 | ¬ FLU) = 0.2, P(TEST2 | FLU) = 0.8, P(TEST2 | ¬ FLU) = 0.1

What is the probability of FLU is true given TEST1 and TEST2 are both true:
   A. .25
   B. .79
   C. .0002
   D. All of the above.
   E. None of the above.

Decision Network [16%] Consider the following Bayesian network. A, B, C, and D are boolean random variables.



7. [4%] Given that A is true and B is false, what is the probability of D being true?
   A. .004
   B. .0034
   C. .36
   D. All of the above.
   E. None of the above.

8. [4%] Converting A into a decision node in the above network and given the following utility function: if D is true then utility is 10 and if D is false utility is 100, which decision for A will give the higher expected utility?
   A. A is true
   B. A is false
   C. Utility is equal in both decisions of A.
   D. All of the above.
   E. None of the above.

9. [4%] In the decision network described in question 28 which of the following events will increase the expected utility the most?
   A. B is true
   B. B is false
   C. C is false
   D. All of the above.
   E. None of the above.

10. [4%] Variable elimination is an algorithm for:
   A. Approximate probabilistic inference in decision networks.
   B. Exact probabilistic inference.
   C. Computing optimal policies.
   D. Reinforcement learning in unknown environments.
   E. None of the above.

Decision Tree [16%] NASA wants to train a robot to recognize two species Martians (M) and Humans (H) based on the following characteristics: Green ∈ *N*,*Y*, *Legs* ∈ 2,3, *Height* ∈ *Short*, *Tall*, Smelly ∈ *N*,*Y*. The available training data is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

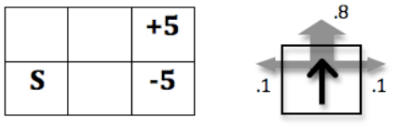
11. [4%] Using the ID3 algorithm to learn a decision tree, the root node chosen would be:
    A. Green
    B. Legs
    C. Both Height and Smelly
    D. All of the above.
    E. None of the above.

12. [4%] Once the learned concept for Martian is an ID3 decision tree, it can be converted into a set of conjunctive rules. For that tree which of following rules would be included:
    A. if (Green=Y and Legs=2 and Height=T and Smelly=N) then M;
    B. if (Green=N and Smelly=Y) then M;
    C. if (Legs=2) then M;
    D. All of the above.
    E. None of the above.

13. [4%] How is Occam's Razor employed in decision tree learning?
    A. ID3 is a greedy approach to construct the shortest tree.
    B. Occam's Razor does not apply to decision tree learning.
    C. Occam's Razor prunes nodes from the tree.
    D. All of the above.
    E. None of the above.

14. [4%] In decision tree learning, the main goal is to:
    A. Minimize the entropy of the training examples associated with each node.
    B. Maximize the entropy of the training examples associated with each node.
    C. Find the best tree structure, such that each training example is a separate tree node.
    D. All of the above.
    E. None of the above.

MDPs and Reinforcement Learning [16%] Given a Gridworld MDP as shown below, the states are grid squares, identified by row then column number. The agent always starts in the state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. The transition function is such that the intended agent movement (North, South, West, or East) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.



(a) (b)

(a) Gridworld MDP. (b) Transition function.

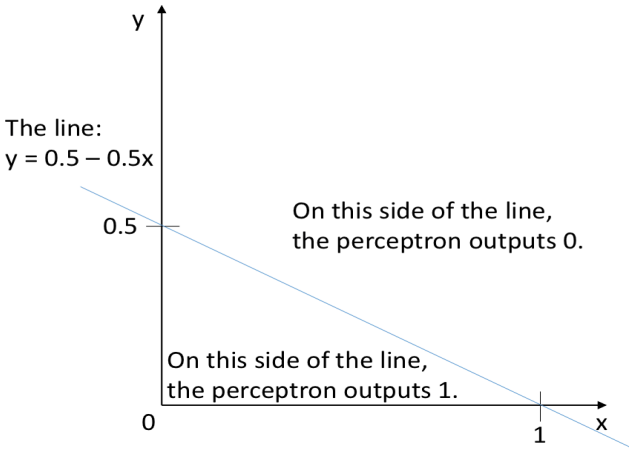
15. [4%] If the agent is deterministic, the optimal policy for this grid is:
    A. State (1,1) North, State (1,2) North, State (2,1) East, State (2,2) East
    B. State (2,1) East, State (2,2) East, State (1,1) North, State (1,2) North
    C. State (1,2) North, State (2,2) East, State (2,1) East, State (1,1) North
    D. All of the above.
    E. None of the above.

16. [4%] Now suppose the agent is stochastic and knows the transition probabilities. Give the second round of value iteration updates for each state, with a discount factor of 0.9 (Assume initial V values are 0 and compute V2 for each non-terminal state).
    A. State (1,1) is 0, State (1,2) is 0, State (2,1) is 0, State (2,2) is 3.924
    B. State (1,1) is 0, State (1,2) is 2.142, State (2,1) is 2.592, State (2,2) 3.924
    C. State (1,1) is 0, State (1,2) is 2.82528, State (2,1) is 2.82528, State (2,2) 3.924
    D. All of the above.
    E. None of the above.

17. [4%] Suppose the stochastic agent does not know the transition probabilities. What does it need in order to learn the optimal policy?
    A. Probability values
    B. Optimal Q values
    C. Training examples of the agent's movement in the grid with rewards received
    D. All of the above.
    E. None of the above.

18. [4%] Reinforcement learning can have problems with scaling because of:
    A. Sparse training data
    B. Calculations taking too much time
    C. It is hard to program
    D. All of the above
    E. None of the above

Neural Networks [16%] Design a perceptron with inputs *x* and *y* and a threshold function as activation function whose output for a given input pair (*x*, *y*) is given by:



19. [4%] The correct weights **w1** for input *x* and **w2** for input *y* and threshold **t** of the perceptron are:
   A. **w1**= -0.5, **w2**= 1, if **t**=> 0.1 outputs 1 else 0
   B. **w1**= 0.5, **w2**= 0.5, if **t**=>1 outputs 1 else 0
   C. **w1**= 0.5, **w2**= 1, if **t** <= .5 outputs 1 else 0
   D. All of the above.
   E. None of the above.

20. [4%] Given input *x* and *y* many different perceptrons can be created whose output is the line above, by making changes to:
   A. Biases
   B. Activation function
   C. Probability distribution
   D. All of the above
   E. None of the above

21. [4%] How many weights does a 2-layer feed-forward neural network with 5 input units, 3 hidden units and 2 output units contain, including the biases (dummy input weights)?
   A. 42
   B. 24
   C. 22
   D. 9
   E. None of the above

22. [4%] A perceptron is guaranteed to learn a given linearly separable function:
   A. When the activation function is a step function.
   B. Within a finite number of training steps.
   C. If all the weights are positive.
   D. All of the above.
   E. None of the above.

General AI Knowledge: True/False [8%] For each of the statements below, answer **TRUE** if the statement is **always true**, or **FALSE** otherwise.

23. [1%] **A. TRUE** **B. FALSE** Cross validation helps to avoid overfitting.

24. [1%] **A. TRUE** **B. FALSE** Unsupervised learning is learning from unlabeled data to maximize a cumulative reward.

25. [1%] **A. TRUE** **B. FALSE** Support vectors are data points that lie closest to the hyperplane.

26. [1%] **A. TRUE** **B. FALSE** A decision tree cannot represent all Boolean function.

27. [1%] **A. TRUE** **B. FALSE** Q-learning creates an explicit transition model.

28. [1%] **A. TRUE** **B. FALSE** A node X is conditionally independent of all other nodes in a Bayesian Network given its Markov Blanket.

29. [1%] **A. TRUE** **B. FALSE** A policy is a mapping from states to actions that tells the agent what action to take in any possible environment state.

30. [1%] **A. TRUE** **B. FALSE** A Bayesian Network modeling n random variables that are all mutually independent of each other would have directed arcs between every pair of variables in the network.

EXAM 3

Midterm 3 Exam

CSCI 561 Fall 2018: Artificial Intelligence

Please print neatly.

First Name:

Last Name:

Student ID:

USC email:

								@usc.edu

Instructions:

1. Date: **11/26/2018 from 8:00 pm to 9:50 pm**
2. Maximum credits/points/percentage for this midterm: 100
3. The percentages for each question are indicated in square brackets [ ].
4. **No books** (or any other material) are allowed.
5. Write down your name, student ID and USC email address on both test and answer booklets.
6. Your exam will be scanned and uploaded online.
7. Write within the spaces provided for your answers in the test answer booklet.
8. **Do NOT write on the 2D barcode.**
9. **The back of the exam pages will NOT be graded. You may use it for scratch.**
10. No questions during the exam. If something is unclear to you, write that in your exam.
11. Be brief: a few words are often enough if they are precise and use the correct vocabulary studied in class.
12. When finished before 9:40 pm, raise completed exam sheets until approached by a proctor.
13. At 9:50 pm, put down your pencil. Stay seated and pass your exam to the nearest aisle or **receive 10% exam penalty.**
14. **Adhere to the Academic Integrity code.**

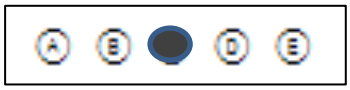
Problems	**100 Percent total**
PART A Short Answer	
AI Applications	4
PART B Multiple Choice, True/False	
General AI Knowledge	8
Bayesian Network	8
Probability	16
Decision Tree	16
MDP and Reinforcement Learning	16
Neural Network	16
Decision Network	16

PART A – Short Answer

AI Applications [4%]

During week 13 there were three presentations about AI applications: Chris Stewart from Google, Ben Marcus from AirMap, and Prof. Wei-min Shen. Choose one of these presentations and briefly describe how the application applies AI to reasoning with uncertainty. (One or two sentences max in the space provide on the answer sheet)

PART B – Multiple Choice & True/False



In this section choose the best option **A**, **B**, **C**, **D**, or **E** and fill in the appropriate circle in the answer sheet. Correct answers receive the % shown in the []. Answers left blank receive 0% and **wrong answers receive -1%**. Fill in at most one circle per problem.

General AI Knowledge: True/False [8%] For each of the statements below, answer **TRUE** if the statement is **always true**, or **FALSE** otherwise.

1. [1%] **A. TRUE B. FALSE** Cross validation helps to avoid overfitting.

2. [1%] **A. TRUE B. FALSE** Unsupervised learning is learning from unlabeled data to maximize a cumulative reward.

3. [1%] **A. TRUE B. FALSE** Support vectors are data points that lie closest to the hyperplane.

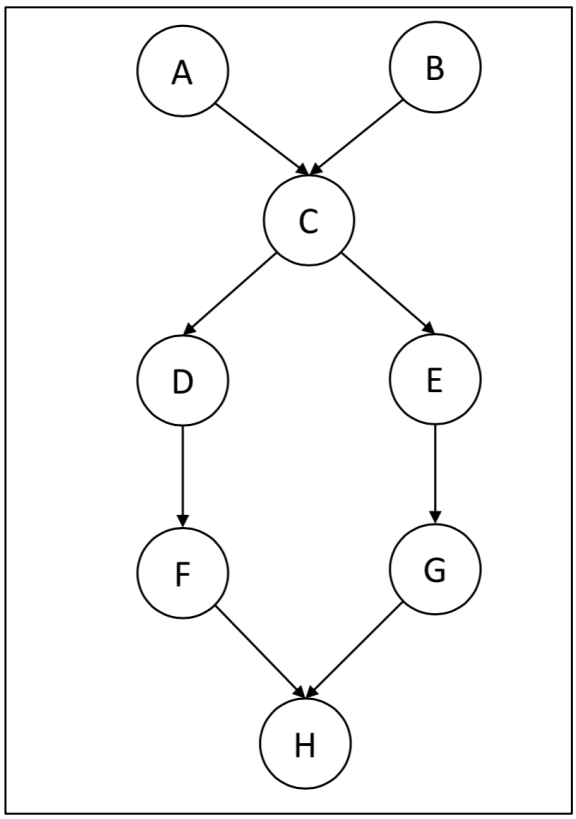
4. [1%] **A. TRUE B. FALSE** A decision tree cannot represent all Boolean function.

5. [1%] **A. TRUE B. FALSE** Q-learning creates an explicit transition model.

6. [1%] **A. TRUE B. FALSE** A node *X* is conditionally independent of all other nodes in a Bayesian Network given its Markov Blanket.

7. [1%] **A. TRUE B. FALSE** A policy is a mapping from states to actions that tells the agent what action to take in any possible environment state.

8. [1%] **A. TRUE B. FALSE** A Bayesian Network modeling *n* random variables that are all mutually independent of each other would have directed arcs between every pair of variables in the network.



Bayesian Network [8%]

9. [4%] For the Bayesian network given above which statement[s] is true:
   A. A and B are independent given C.
   B. A and B are independent.
   C. D is conditionally independent of G given C.
   D. All of the above.
   E. None of the above.

10. [4%] Which statement[s] is true in any Bayesian Network:
   A. A node is conditionally independent of all other nodes given its Markov blanket.
   B. A node is independent of its non-descendants given its parents.
   C. At least one node in the network must have no incoming arcs.
   D. All of the above.
   E. None of the above.

Probability [16%]

For questions 11 & 12 consider the following joint probability table for the boolean random variables **Q** and **S**:

Q	**S**	
T	T	0.1
T	F	0.2
F	T	0.3
F	F	0.4

11. [4%] Which probability value is the largest?
    A. P(Q)
    B. P(¬ S)
    C. P(Q ∨ S)
    D. P(Q ⇒ S)
    E. P(Q | S)

12. [4%] Which of the following statement[s] is true:
    A. P(Q ⇒ S) > P(S | Q)
    B. P(Q ∨ S) + P(Q , S) = 1
    C. P(¬Q) = 1- P(¬ S)
    D. All of the above.
    E. None of the above.

13. [4%] If random variable *X* is conditionally independent of *Y* given *Z* then
    A. P(X,Y|Z)=P(X|Z)P(Y|Z)
    B. P(X|Y,Z)=P(X|Z)
    C. P(Y|X,Z)=P(Y|Z)
    D. All of the above.
    E. None of the above

14. [4%] Assume that the boolean random variables TEST1 and TEST2 are conditionally independent given FLU and the following probabilities: P(FLU) = 0.1, P(TEST1 | FLU) = 0.9, P(TEST1 | ¬ FLU) = 0.2, P(TEST2 | FLU) = 0.8, P(TEST2 | ¬ FLU) = 0.1

What is the probability of FLU is true given TEST1 and TEST2 are both true:
    A. .25
    B. .79
    C. .0002
    D. All of the above.
    E. None of the above.

Decision Tree [16%] NASA wants to train a robot to recognize two species Martians (M) and Humans (H) based on the following characteristics: Green ∈ *N*,*Y*, *Legs* ∈ 2,3, *Height* ∈ *Short*, *Tall*, Smelly ∈ *N*,*Y*. The available training data is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

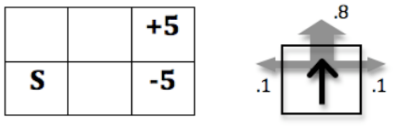
15. [4%] Using the ID3 algorithm to learn a decision tree, the root node chosen would be:
    A. Green
    B. Legs
    C. Both Height and Smelly
    D. All of the above.
    E. None of the above.

16. [4%] Once the learned concept for Martian is an ID3 decision tree, it can be converted into a set of conjunctive rules. For that tree which of following rules would be included:
    A. if (Green=Y and Legs=2 and Height=T and Smelly=N) then M;
    B. if (Green=N and Smelly=Y) then M;
    C. if (Legs=2) then M;
    D. All of the above.
    E. None of the above.

17. [4%] How is Occam's Razor employed in decision tree learning?
    A. ID3 is a greedy approach to construct the shortest tree.
    B. Occam's Razor does not apply to decision tree learning.
    C. Occam's Razor prunes nodes from the tree.
    D. All of the above.
    E. None of the above.

18. [4%] In decision tree learning, the main goal is to:
    A. Minimize the entropy of the training examples associated with each node.
    B. Maximize the entropy of the training examples associated with each node.
    C. Find the best tree structure, such that each training example is a separate tree node.
    D. All of the above.
    E. None of the above.

MDPs and Reinforcement Learning [16%] Given a Gridworld MDP as shown below, the states are grid squares, identified by row then column number. The agent always starts in the state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. The transition function is such that the intended agent movement (North, South, West, or East) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.



(a) (b)

(a) Gridworld MDP. (b) Transition function.

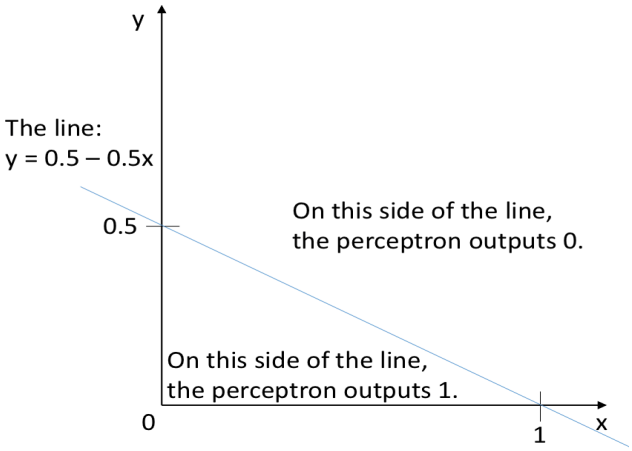
19. [4%] If the agent is deterministic, the optimal policy for this grid is:
    A. State (1,1) North, State (1,2) North, State (2,1) East, State (2,2) East
    B. State (2,1) East, State (2,2) East, State (1,1) North, State (1,2) North
    C. State (1,2) North, State (2,2) East, State (2,1) East, State (1,1) North
    D. All of the above.
    E. None of the above.

20. [4%] Now suppose the agent is stochastic and knows the transition probabilities. Give the second round of value iteration updates for each state, with a discount factor of 0.9 (Assume initial V values are 0 and compute V2 for each non-terminal state).
    A. State (1,1) is 0, State (1,2) is 0, State (2,1) is 0, State (2,2) is 3.924
    B. State (1,1) is 0, State (1,2) is 2.142, State (2,1) is 2.592, State (2,2) 3.924
    C. State (1,1) is 0, State (1,2) is 2.82528, State (2,1) is 2.82528, State (2,2) 3.924
    D. All of the above.
    E. None of the above.

21. [4%] Suppose the stochastic agent does not know the transition probabilities. What does it need in order to learn the optimal policy?
    A. Probability values
    B. Optimal Q values
    C. Training examples of the agent's movement in the grid with rewards received
    D. All of the above.
    E. None of the above.

22. [4%] Reinforcement learning can have problems with scaling because of:
    A. Sparse training data
    B. Calculations taking too much time
    C. It is hard to program
    D. All of the above
    E. None of the above

Neural Networks [16%] Design a perceptron with inputs *x* and *y* and a threshold function as activation function whose output for a given input pair (*x*, *y*) is given by:



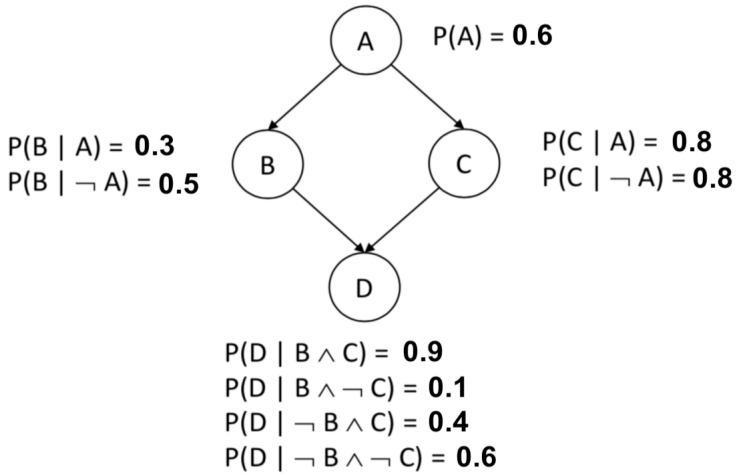
23. [4%] The correct weights **w1** for input *x* and **w2** for input *y* and threshold **t** of the perceptron are:
   A. **w1**= -0.5, **w2**= 1, if **t**=> 0.1 outputs 1 else 0
   B. **w1**= 0.5, **w2**= 0.5, if **t**=>1 outputs 1 else 0
   C. **w1**= 0.5, **w2**= 1, if **t** <= .5 outputs 1 else 0
   D. All of the above.
   E. None of the above.

24. [4%] Given input *x* and *y* many different perceptrons can be created whose output is the line above, by making changes to:
   A. Biases
   B. Activation function
   C. Probability distribution
   D. All of the above
   E. None of the above

25. [4%] How many weights does a 2-layer feed-forward neural network with 5 input units, 3 hidden units and 2 output units contain, including the biases (dummy input weights)?
   A. 42
   B. 24
   C. 22
   D. 9
   E. None of the above

26. [4%] A perceptron is guaranteed to learn a given linearly separable function:
   A. When the activation function is a step function.
   B. Within a finite number of training steps.
   C. If all the weights are positive.
   D. All of the above.
   E. None of the above.

Decision Network [16%] Consider the following Bayesian network. A, B, C, and D are boolean random variables.



27. [4%] Given that A is true and B is false, what is the probability of D being true?
    A. .004
    B. .0034
    C. .36
    D. All of the above.
    E. None of the above.

28. [4%] Converting A into a decision node in the above network and given the following utility function: if D is true then utility is 10 and if D is false utility is 100, which decision for A will give the higher expected utility?
    A. A is true
    B. A is false
    C. Utility is equal in both decisions of A.
    D. All of the above.
    E. None of the above.

29. [4%] In the decision network described in question 28 which of the following events will increase the expected utility the most?
    A. B is true
    B. B is false
    C. C is false
    D. All of the above.
    E. None of the above.

30. [4%] Variable elimination is an algorithm for:
    A. Approximate probabilistic inference in decision networks.
    B. Exact probabilistic inference.
    C. Computing optimal policies.
    D. Reinforcement learning in unknown environments.
    E. None of the above.

EXAM 4

Midterm 3 Exam

CSCI 561 Fall 2018: Artificial Intelligence

Please print neatly.

First Name:

Last Name:

Student ID:

USC email:

@usc.edu

Instructions:

14. Date: **11/26/2018 from 8:00 pm to 9:50 pm**
15. Maximum credits/points/percentage for this midterm: 100
16. The percentages for each question are indicated in square brackets [ ].
17. **No books** (or any other material) are allowed.
18. Write down your name, student ID and USC email address on both test and answer booklets.
19. Your exam will be scanned and uploaded online.
20. Write within the spaces provided for your answers in the test answer booklet.
21. **Do NOT write on the 2D barcode.**
22. **The back of the exam pages will NOT be graded. You may use it for scratch.**
23. No questions during the exam. If something is unclear to you, write that in your exam.
24. Be brief: a few words are often enough if they are precise and use the correct vocabulary studied in class.
25. When finished before 9:40 pm, raise completed exam sheets until approached by a proctor.
26. At 9:50 pm, put down your pencil. Stay seated and pass your exam to the nearest aisle or **receive 10% exam penalty.**
14. **Adhere to the Academic Integrity code.**

Problems	**100 Percent total**
PART A Short Answer	
Probabilistic Reasoning	4
PART B Multiple Choice, True/False	
Bayesian Network	8
Probability	16
Decision Network	16
Decision Tree	16
MDP and Reinforcement Learning	16
Neural Network	16
General AI Knowledge	8

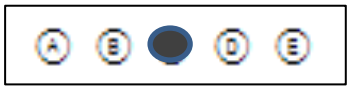
PART A – Short Answer

Probabilistic Reasoning [4%] Assume that the boolean random variables **J** and **K** are conditionally independent given **L** and with the following probabilities:

P(L) = 0.1, P(J | L) = 0.9, P(J | ¬ L) = 0.2, P(K | L) = 0.8, P(K | ¬ L) = 0.1

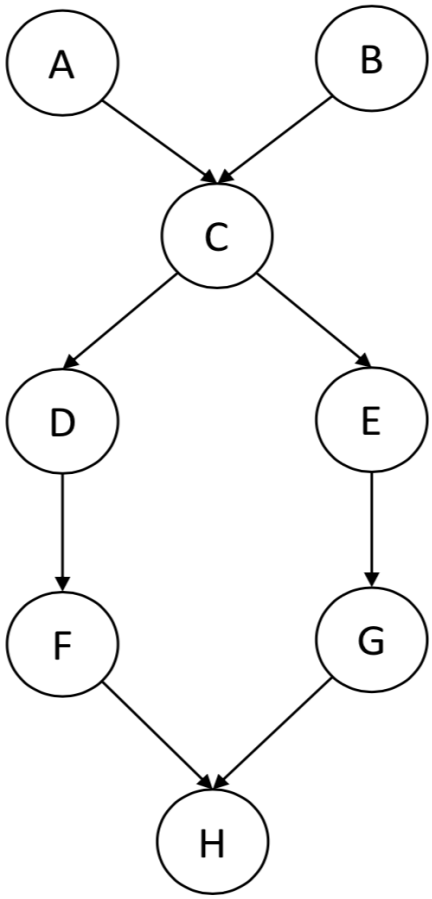
Draw the Bayesian Network in space provide in the answer sheet.

PART B – Multiple Choice & True/False



In this section choose the best option **A**, **B**, **C**, **D**, or **E** and fill in the appropriate circle in the answer sheet. Correct answers receive the % shown in the []. Answers left blank receive 0% and **wrong answers receive -1%**. Fill in at most one circle per problem.

Bayesian Network [8%]



1. [4%] For the Bayesian network given which statement[s] is true:
   A. A and B are independent given C.
   B. A and B are independent.
   C. D is conditionally independent of G given C.
   D. All of the above.
   E. None of the above.

2. [4%] Which statement[s] is true in any Bayesian Network:
   A. A node is conditionally independent of all other nodes given its Markov blanket.
   B. A node is independent of its non-descendants given its parents.
   C. At least one node in the network must have no incoming arcs.
   D. All of the above.
   E. None of the above.

Probability [16%]

For questions 3 & 4 consider the following joint probability table for the boolean random variables **Q** and **S**:

Q	**S**	
T	T	0.1
T	F	0.2
F	T	0.3
F	F	0.4

3. [4%] Which probability value is the largest?
   A. P(Q)
   B. P(¬ S)
   C. P(Q ∨ S)
   D. P(Q ⇒ S)
   E. P(Q | S)

4. [4%] Which of the following statement[s] is true:
   A. P(Q ⇒ S) > P(S | Q)
   B. P(Q ∨ S) + P(Q , S) = 1
   C. P(¬Q) = 1- P(¬ S)
   D. All of the above.
   E. None of the above.

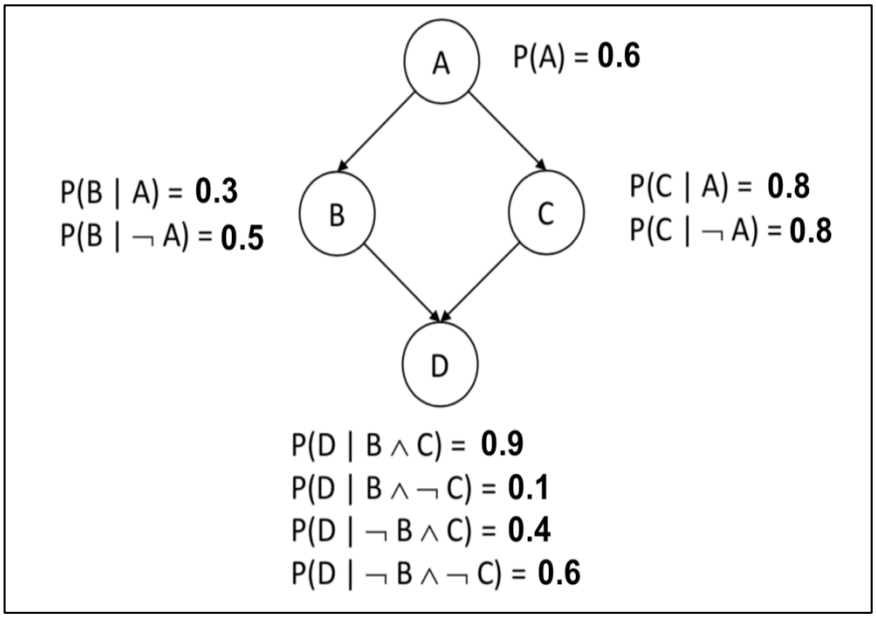
5. [4%] If random variable *X* is conditionally independent of *Y* given *Z* then
   A. P(X,Y|Z)=P(X|Z)P(Y|Z)
   B. P(X|Y,Z)=P(X|Z)
   C. P(Y|X,Z)=P(Y|Z)
   D. All of the above.
   E. None of the above

6. [4%] Assume that the boolean random variables TEST1 and TEST2 are conditionally independent given FLU and the following probabilities: P(FLU) = 0.1, P(TEST1 | FLU) = 0.9, P(TEST1 | ¬ FLU) = 0.2, P(TEST2 | FLU) = 0.8, P(TEST2 | ¬ FLU) = 0.1

What is the probability of FLU is true given TEST1 and TEST2 are both true:
   A. .25
   B. .79
   C. .0002
   D. All of the above.
   E. None of the above.

Decision Network [16%]

Consider the following Bayesian network. A, B, C, and D are boolean random variables.



7. [4%] Given that A is true and B is false, what is the probability of D being true?
   A. .004
   B. .0034
   C. .36
   D. All of the above.
   E. None of the above.

8. [4%] Converting A into a decision node in the above network and given the following utility function: if D is true then utility is 10 and if D is false utility is 100, which decision for A will give the higher expected utility?
   A. A is true
   B. A is false
   C. Utility is equal in both decisions of A.
   D. All of the above.
   E. None of the above.

9. [4%] In the decision network described in question 28 which of the following events will increase the expected utility the most?
   A. B is true
   B. B is false
   C. C is false
   D. All of the above.
   E. None of the above.

10. [4%] Variable elimination is an algorithm for:
   A. Approximate probabilistic inference in decision networks.
   B. Exact probabilistic inference.
   C. Computing optimal policies.
   D. Reinforcement learning in unknown environments.
   E. None of the above.

Decision Tree [16%] NASA wants to train a robot to recognize two species Martians (M) and Humans (H) based on the following characteristics: Green ∈ *N*,*Y*, *Legs* ∈ 2,3, *Height* ∈ *Short*, *Tall*, Smelly ∈ *N*,*Y*. The available training data is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

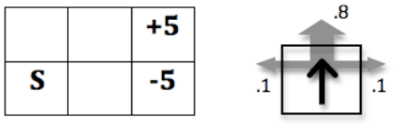
11. [4%] Using the ID3 algorithm to learn a decision tree, the root node chosen would be:
    A. Green
    B. Legs
    C. Both Height and Smelly
    D. All of the above.
    E. None of the above.

12. [4%] Once the learned concept for Martian is an ID3 decision tree, it can be converted into a set of conjunctive rules. For that tree which of following rules would be included:
    A. if (Green=Y and Legs=2 and Height=T and Smelly=N) then M;
    B. if (Green=N and Smelly=Y) then M;
    C. if (Legs=2) then M;
    D. All of the above.
    E. None of the above.

13. [4%] How is Occam's Razor employed in decision tree learning?
    A. ID3 is a greedy approach to construct the shortest tree.
    B. Occam's Razor does not apply to decision tree learning.
    C. Occam's Razor prunes nodes from the tree.
    D. All of the above.
    E. None of the above.

14. [4%] In decision tree learning, the main goal is to:
    A. Minimize the entropy of the training examples associated with each node.
    B. Maximize the entropy of the training examples associated with each node.
    C. Find the best tree structure, such that each training example is a separate tree node.
    D. All of the above.
    E. None of the above.

MDPs and Reinforcement Learning [16%] Given a Gridworld MDP as shown below, the states are grid squares, identified by row then column number. The agent always starts in the state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. The transition function is such that the intended agent movement (North, South, West, or East) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.



(a) (b)

(a) Gridworld MDP. (b) Transition function.

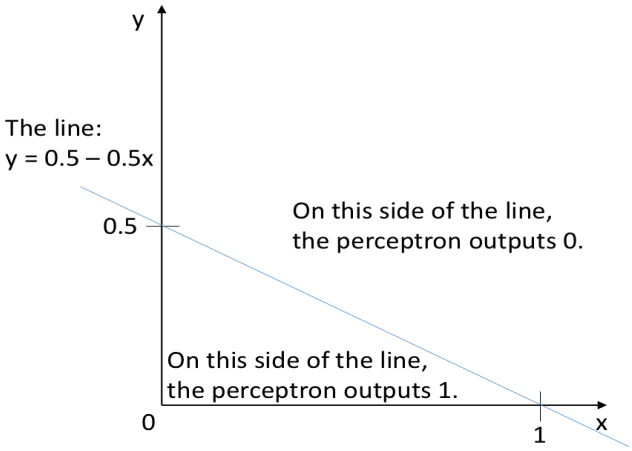
15. [4%] If the agent is deterministic, the optimal policy for this grid is:
    A. State (1,1) North, State (1,2) North, State (2,1) East, State (2,2) East
    B. State (2,1) East, State (2,2) East, State (1,1) North, State (1,2) North
    C. State (1,2) North, State (2,2) East, State (2,1) East, State (1,1) North
    D. All of the above.
    E. None of the above.

16. [4%] Now suppose the agent is stochastic and knows the transition probabilities. Give the second round of value iteration updates for each state, with a discount factor of 0.9 (Assume initial V values are 0 and compute V2 for each non-terminal state).
    A. State (1,1) is 0, State (1,2) is 0, State (2,1) is 0, State (2,2) is 3.924
    B. State (1,1) is 0, State (1,2) is 2.142, State (2,1) is 2.592, State (2,2) 3.924
    C. State (1,1) is 0, State (1,2) is 2.82528, State (2,1) is 2.82528, State (2,2) 3.924
    D. All of the above.
    E. None of the above.

17. [4%] Suppose the stochastic agent does not know the transition probabilities. What does it need in order to learn the optimal policy?
    A. Probability values
    B. Optimal Q values
    C. Training examples of the agent's movement in the grid with rewards received
    D. All of the above.
    E. None of the above.

18. [4%] Reinforcement learning can have problems with scaling because of:
    A. Sparse training data
    B. Calculations taking too much time
    C. It is hard to program
    D. All of the above
    E. None of the above

Neural Networks [16%] Design a perceptron with inputs *x* and *y* and a threshold function as activation function whose output for a given input pair (*x*, *y*) is given by:



19. [4%] The correct weights **w1** for input *x* and **w2** for input *y* and threshold **t** of the perceptron are:
   A. **w1**= -0.5, **w2**= 1, if **t**=> 0.1 outputs 1 else 0
   B. **w1**= 0.5, **w2**= 0.5, if **t**=>1 outputs 1 else 0
   C. **w1**= 0.5, **w2**= 1, if **t** <= .5 outputs 1 else 0
   D. All of the above.
   E. None of the above.

20. [4%] Given input *x* and *y* many different perceptrons can be created whose output is the line above, by making changes to:
   A. Biases
   B. Activation function
   C. Probability distribution
   D. All of the above
   E. None of the above

21. [4%] How many weights does a 2-layer feed-forward neural network with 5 input units, 3 hidden units and 2 output units contain, including the biases (dummy input weights)?
   A. 42
   B. 24
   C. 22
   D. 9
   E. None of the above

22. [4%] A perceptron is guaranteed to learn a given linearly separable function:
   A. When the activation function is a step function.
   B. Within a finite number of training steps.
   C. If all the weights are positive.
   D. All of the above.
   E. None of the above.

General AI Knowledge: True/False [8%] For each of the statements below, answer **TRUE** if the statement is **always true**, or **FALSE** otherwise.

23. [1%] **A. TRUE** **B. FALSE** Cross validation helps to avoid overfitting.

24. [1%] **A. TRUE** **B. FALSE** Unsupervised learning is learning from unlabeled data to maximize a cumulative reward.

25. [1%] **A. TRUE** **B. FALSE** Support vectors are data points that lie closest to the hyperplane.

26. [1%] **A. TRUE** **B. FALSE** A decision tree cannot represent all Boolean function.

27. [1%] **A. TRUE** **B. FALSE** Q-learning creates an explicit transition model.

28. [1%] **A. TRUE** **B. FALSE** A node X is conditionally independent of all other nodes in a Bayesian Network given its Markov Blanket.

29. [1%] **A. TRUE** **B. FALSE** A policy is a mapping from states to actions that tells the agent what action to take in any possible environment state.

30. [1%] **A. TRUE** **B. FALSE** A Bayesian Network modeling n random variables that are all mutually independent of each other would have directed arcs between every pair of variables in the network.

EXAM 5

Midterm 3 Exam

CSCI 561 Fall 2018: Artificial Intelligence

Please print neatly.

First Name:

Last Name:

Student ID:

USC email: @usc.edu

Instructions:

1. Date: **11/26/2018 from 8:00 pm to 9:50 pm**
2. Maximum credits/points/percentage for this midterm: 100
3. The percentages for each question are indicated in square brackets [ ].
4. **No books** (or any other material) are allowed.
5. Write down your name, student ID and USC email address on both test and answer booklets.
6. Your exam will be scanned and uploaded online.
7. Write within the spaces provided for your answers in the test answer booklet.
8. **Do NOT write on the 2D barcode.**
9. **The back of the exam pages will NOT be graded. You may use it for scratch.**
10. No questions during the exam. If something is unclear to you, write that in your exam.
11. Be brief: a few words are often enough if they are precise and use the correct vocabulary studied in class.
12. When finished before 9:40 pm, raise completed exam sheets until approached by a proctor.
13. At 9:50 pm, put down your pencil. Stay seated and pass your exam to the nearest aisle or **receive 10% exam penalty.**
14. **Adhere to the Academic Integrity code.**

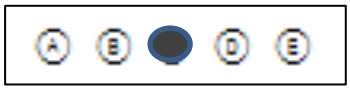
Problems	**100 Percent total**
PART A Short Answer	
AI Applications	4
PART B Multiple Choice, True/False	
General AI Knowledge	8
Bayesian Network	8
Probability	16
Decision Tree	16
MDP and Reinforcement Learning	16
Neural Network	16
Decision Network	16

PART A – Short Answer

AI Applications [4%]

During week 13 there were three presentations about AI applications: Chris Stewart from Google, Ben Marcus from AirMap, and Prof. Wei-min Shen. Choose one of these presentations and briefly describe how the application applies AI to reasoning with uncertainty. (One or two sentences max in the space provide on the answer sheet)

PART B – Multiple Choice & True/False



In this section choose the best option **A**, **B**, **C**, **D**, or **E** and fill in the appropriate circle in the answer sheet. Correct answers receive the % shown in the []. Answers left blank receive 0% and **wrong answers receive -1%**. Fill in at most one circle per problem.

General AI Knowledge: True/False [8%] For each of the statements below, answer **TRUE** if the statement is **always true**, or **FALSE** otherwise.

1. [1%] **A. TRUE B. FALSE** Cross validation helps to avoid overfitting.

2. [1%] **A. TRUE B. FALSE** Unsupervised learning is learning from unlabeled data to maximize a cumulative reward.

3. [1%] **A. TRUE B. FALSE** Support vectors are data points that lie closest to the hyperplane.

4. [1%] **A. TRUE B. FALSE** A decision tree cannot represent all Boolean function.

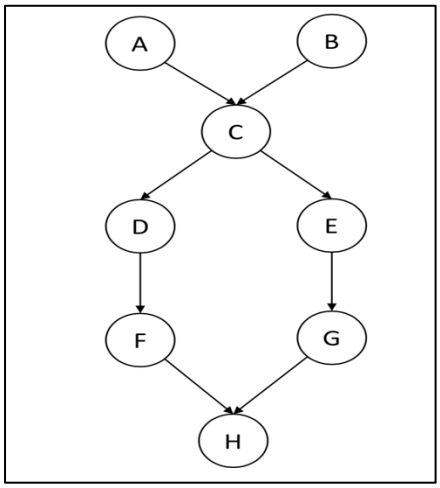
5. [1%] **A. TRUE B. FALSE** Q-learning creates an explicit transition model.

6. [1%] **A. TRUE B. FALSE** A node *X* is conditionally independent of all other nodes in a Bayesian Network given its Markov Blanket.

7. [1%] **A. TRUE B. FALSE** A policy is a mapping from states to actions that tells the agent what action to take in any possible environment state.

8. [1%] **A. TRUE B. FALSE** A Bayesian Network modeling *n* random variables that are all mutually independent of each other would have directed arcs between every pair of variables in the network.

Bayesian Network [8%]



9. [4%] For the Bayesian network given above which statement[s] is true:
   A. A and B are independent given C.
   B. A and B are independent.
   C. D is conditionally independent of G given C.
   D. All of the above.
   E. None of the above.

10. [4%] Which statement[s] is true in any Bayesian Network:
   A. A node is conditionally independent of all other nodes given its Markov blanket.
   B. A node is independent of its non-descendants given its parents.
   C. At least one node in the network must have no incoming arcs.
   D. All of the above.
   E. None of the above.

Q	S	
T	T	0.1
T	F	0.2
F	T	0.3
F	F	0.4

Probability [16%]

For questions 11 & 12 consider the following joint probability table for the boolean random variables **Q** and **S**:

11. [4%] Which probability value is the largest?
    A. P(Q)
    B. P(¬ S)
    C. P(Q ∨ S)
    D. P(Q ⇒ S)
    E. P(Q | S)

12. [4%] Which of the following statement[s] is true:
    A. P(Q ⇒ S) > P(S | Q)
    B. P(Q ∨ S) + P(Q , S) = 1
    C. P(¬Q) = 1- P(¬ S)
    D. All of the above.
    E. None of the above.

13. [4%] If random variable *X* is conditionally independent of *Y* given *Z* then
    A. P(X,Y|Z)=P(X|Z)P(Y|Z)
    B. P(X|Y,Z)=P(X|Z)
    C. P(Y|X,Z)=P(Y|Z)
    D. All of the above.
    E. None of the above

14. [4%] Assume that the boolean random variables TEST1 and TEST2 are conditionally independent given FLU and the following probabilities: P(FLU) = 0.1, P(TEST1 | FLU) = 0.9, P(TEST1 | ¬ FLU) = 0.2, P(TEST2 | FLU) = 0.8, P(TEST2 | ¬ FLU) = 0.1

What is the probability of FLU is true given TEST1 and TEST2 are both true:
    A. .25
    B. .79
    C. .0002
    D. All of the above.
    E. None of the above.

Decision Tree [16%] NASA wants to train a robot to recognize two species Martians (M) and Humans (H) based on the following characteristics: Green ∈ *N*,*Y*, *Legs* ∈ 2,3, *Height* ∈ *Short*, *Tall*, Smelly ∈ *N*,*Y*. The available training data is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

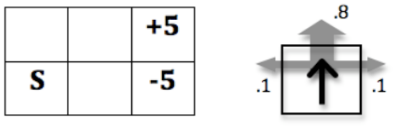
15. [4%] Using the ID3 algorithm to learn a decision tree, the root node chosen would be:
    A. Green
    B. Legs
    C. Both Height and Smelly
    D. All of the above.
    E. None of the above.

16. [4%] Once the learned concept for Martian is an ID3 decision tree, it can be converted into a set of conjunctive rules. For that tree which of following rules would be included:
    A. if (Green=Y and Legs=2 and Height=T and Smelly=N) then M;
    B. if (Green=N and Smelly=Y) then M;
    C. if (Legs=2) then M;
    D. All of the above.
    E. None of the above.

17. [4%] How is Occam's Razor employed in decision tree learning?
    A. ID3 is a greedy approach to construct the shortest tree.
    B. Occam's Razor does not apply to decision tree learning.
    C. Occam's Razor prunes nodes from the tree.
    D. All of the above.
    E. None of the above.

18. [4%] In decision tree learning, the main goal is to:
    A. Minimize the entropy of the training examples associated with each node.
    B. Maximize the entropy of the training examples associated with each node.
    C. Find the best tree structure, such that each training example is a separate tree node.
    D. All of the above.
    E. None of the above.

MDPs and Reinforcement Learning [16%] Given a Gridworld MDP as shown below, the states are grid squares, identified by row then column number. The agent always starts in the state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. The transition function is such that the intended agent movement (North, South, West, or East) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.



(a) (b)

(a) Gridworld MDP. (b) Transition function.

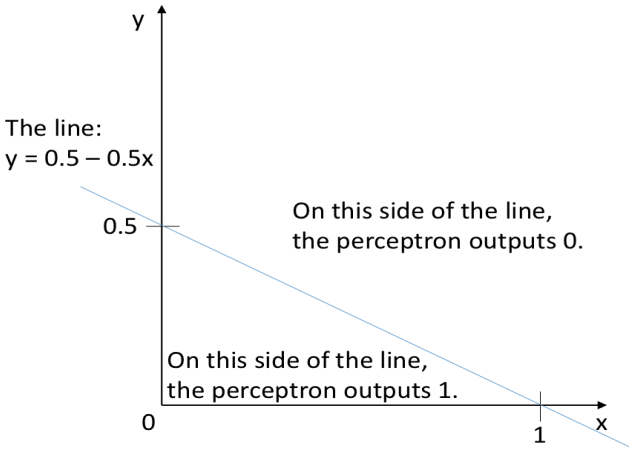
19. [4%] If the agent is deterministic, the optimal policy for this grid is:
    A. State (1,1) North, State (1,2) North, State (2,1) East, State (2,2) East
    B. State (2,1) East, State (2,2) East, State (1,1) North, State (1,2) North
    C. State (1,2) North, State (2,2) East, State (2,1) East, State (1,1) North
    D. All of the above.
    E. None of the above.

20. [4%] Now suppose the agent is stochastic and knows the transition probabilities. Give the second round of value iteration updates for each state, with a discount factor of 0.9 (Assume initial V values are 0 and compute V2 for each non-terminal state).
    A. State (1,1) is 0, State (1,2) is 0, State (2,1) is 0, State (2,2) is 3.924
    B. State (1,1) is 0, State (1,2) is 2.142, State (2,1) is 2.592, State (2,2) 3.924
    C. State (1,1) is 0, State (1,2) is 2.82528, State (2,1) is 2.82528, State (2,2) 3.924
    D. All of the above.
    E. None of the above.

21. [4%] Suppose the stochastic agent does not know the transition probabilities. What does it need in order to learn the optimal policy?
    A. Probability values
    B. Optimal Q values
    C. Training examples of the agent's movement in the grid with rewards received
    D. All of the above.
    E. None of the above.

22. [4%] Reinforcement learning can have problems with scaling because of:
    A. Sparse training data
    B. Calculations taking too much time
    C. It is hard to program
    D. All of the above
    E. None of the above

Neural Networks [16%] Design a perceptron with inputs *x* and *y* and a threshold function as activation function whose output for a given input pair (*x*, *y*) is given by:



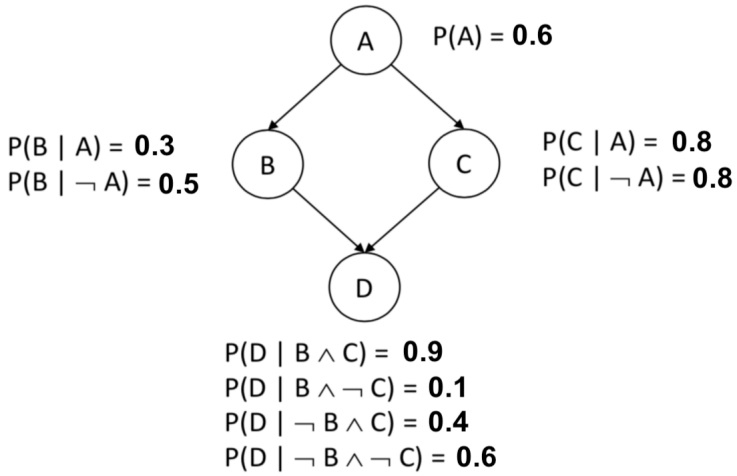
23. [4%] The correct weights **w1** for input *x* and **w2** for input *y* and threshold **t** of the perceptron are:
   A. **w1**= -0.5, **w2**= 1, if **t**=> 0.1 outputs 1 else 0
   B. **w1**= 0.5, **w2**= 0.5, if **t**=>1 outputs 1 else 0
   C. **w1**= 0.5, **w2**= 1, if **t** <= .5 outputs 1 else 0
   D. All of the above.
   E. None of the above.

24. [4%] Given input *x* and *y* many different perceptrons can be created whose output is the line above, by making changes to:
   A. Biases
   B. Activation function
   C. Probability distribution
   D. All of the above
   E. None of the above

25. [4%] How many weights does a 2-layer feed-forward neural network with 5 input units, 3 hidden units and 2 output units contain, including the biases (dummy input weights)?
   A. 42
   B. 24
   C. 22
   D. 9
   E. None of the above

26. [4%] A perceptron is guaranteed to learn a given linearly separable function:
   A. When the activation function is a step function.
   B. Within a finite number of training steps.
   C. If all the weights are positive.
   D. All of the above.
   E. None of the above.

Decision Network [16%] Consider the following Bayesian network. A, B, C, and D are boolean random variables.



27. [4%] Given that A is true and B is false, what is the probability of D being true?
    A. .004
    B. .0034
    C. .36
    D. All of the above.
    E. None of the above.

28. [4%] Converting A into a decision node in the above network and given the following utility function: if D is true then utility is 10 and if D is false utility is 100, which decision for A will give the higher expected utility?
    A. A is true
    B. A is false
    C. Utility is equal in both decisions of A.
    D. All of the above.
    E. None of the above.

29. [4%] In the decision network described in question 28 which of the following events will increase the expected utility the most?
    A. B is true
    B. B is false
    C. C is false
    D. All of the above.
    E. None of the above.

30. [4%] Variable elimination is an algorithm for:
    A. Approximate probabilistic inference in decision networks.
    B. Exact probabilistic inference.
    C. Computing optimal policies.
    D. Reinforcement learning in unknown environments.
    E. None of the above.

EXAM 6

Midterm 3 Exam

CSCI 561 Fall 2018: Artificial Intelligence

Please print neatly.

First Name:

Last Name:

Student ID:

USC email:

@usc.edu

Instructions:

14. Date: **11/26/2018 from 8:00 pm to 9:50 pm**
15. Maximum credits/points/percentage for this midterm: 100
16. The percentages for each question are indicated in square brackets [ ].
17. **No books** (or any other material) are allowed.
18. Write down your name, student ID and USC email address on both test and answer booklets.
19. Your exam will be scanned and uploaded online.
20. Write within the spaces provided for your answers in the test answer booklet.
21. **Do NOT write on the 2D barcode.**
22. **The back of the exam pages will NOT be graded. You may use it for scratch.**
23. No questions during the exam. If something is unclear to you, write that in your exam.
24. Be brief: a few words are often enough if they are precise and use the correct vocabulary studied in class.
25. When finished before 9:40 pm, raise completed exam sheets until approached by a proctor.
26. At 9:50 pm, put down your pencil. Stay seated and pass your exam to the nearest aisle or **receive 10% exam penalty.**
14. **Adhere to the Academic Integrity code.**

Problems	**100 Percent total**
PART A Short Answer	
Probabilistic Reasoning	4
PART B Multiple Choice, True/False	
Bayesian Network	8
Probability	16
Decision Network	16
Decision Tree	16
MDP and Reinforcement Learning	16
Neural Network	16
General AI Knowledge	8

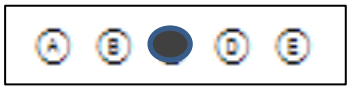
PART A – Short Answer

Probabilistic Reasoning [4%] Assume that the boolean random variables **J** and **K** are conditionally independent given **L** and with the following probabilities:

P(L) = 0.1, P(J | L) = 0.9, P(J | ¬ L) = 0.2, P(K | L) = 0.8, P(K | ¬ L) = 0.1

Draw the Bayesian Network in space provide in the answer sheet.

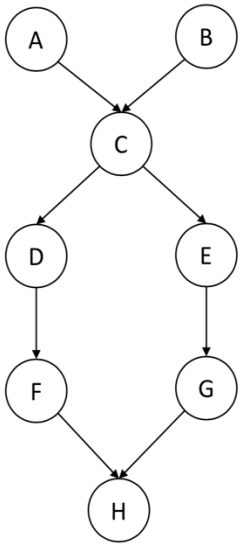
PART B – Multiple Choice & True/False



In this section choose the best option **A**, **B**, **C**, **D**, or **E** and fill in the appropriate circle in the answer sheet. Correct answers receive the % shown in the []. Answers left blank receive 0% and **wrong answers receive -1%**. Fill in at most one circle per problem.

Bayesian Network [8%]

1. [4%] For the Bayesian network given which statement[s] is true:



   A. A and B are independent given C.
   B. A and B are independent.
   C. D is conditionally independent of G given C.
   D. All of the above.
   E. None of the above.

2. [4%] Which statement[s] is true in any Bayesian Network:
   A. A node is conditionally independent of all other nodes given its Markov blanket.
   B. A node is independent of its non-descendants given its parents.
   C. At least one node in the network must have no incoming arcs.
   D. All of the above.
   E. None of the above.

Probability [16%]

Q	S	
T	T	0.1
T	F	0.2
F	T	0.3
F	F	0.4

For questions 3 & 4 consider the following joint probability table for the boolean random variables **Q** and **S**:

3. [4%] Which probability value is the largest?
   A. P(Q)
   B. P(¬ S)
   C. P(Q ∨ S)
   D. P(Q ⇒ S)
   E. P(Q | S)

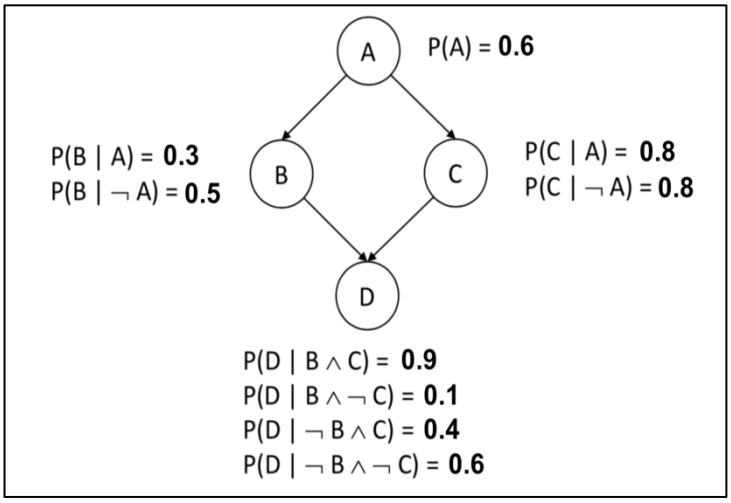
4. [4%] Which of the following statement[s] is true:
   A. P(Q ⇒ S) > P(S | Q)
   B. P(Q ∨ S) + P(Q , S) = 1
   C. P(¬Q) = 1- P(¬ S)
   D. All of the above.
   E. None of the above.

5. [4%] If random variable *X* is conditionally independent of *Y* given *Z* then
   A. P(X,Y|Z)=P(X|Z)P(Y|Z)
   B. P(X|Y,Z)=P(X|Z)
   C. P(Y|X,Z)=P(Y|Z)
   D. All of the above.
   E. None of the above

6. [4%] Assume that the boolean random variables TEST1 and TEST2 are conditionally independent given FLU and the following probabilities: P(FLU) = 0.1, P(TEST1 | FLU) = 0.9, P(TEST1 | ¬ FLU) = 0.2, P(TEST2 | FLU) = 0.8, P(TEST2 | ¬ FLU) = 0.1

What is the probability of FLU is true given TEST1 and TEST2 are both true:
   A. .25
   B. .79
   C. .0002
   D. All of the above.
   E. None of the above.

Decision Network [16%] Consider the following Bayesian network. A, B, C, and D are boolean random variables.



7. [4%] Given that A is true and B is false, what is the probability of D being true?
   A. .004
   B. .0034
   C. .36
   D. All of the above.
   E. None of the above.

8. [4%] Converting A into a decision node in the above network and given the following utility function: if D is true then utility is 10 and if D is false utility is 100, which decision for A will give the higher expected utility?
   A. A is true
   B. A is false
   C. Utility is equal in both decisions of A.
   D. All of the above.
   E. None of the above.

9. [4%] In the decision network described in question 28 which of the following events will increase the expected utility the most?
   A. B is true
   B. B is false
   C. C is false
   D. All of the above.
   E. None of the above.

10. [4%] Variable elimination is an algorithm for:
   A. Approximate probabilistic inference in decision networks.
   B. Exact probabilistic inference.
   C. Computing optimal policies.
   D. Reinforcement learning in unknown environments.
   E. None of the above.

Decision Tree [16%] NASA wants to train a robot to recognize two species Martians (M) and Humans (H) based on the following characteristics: Green ∈ *N*,*Y*, *Legs* ∈ 2,3, *Height* ∈ *Short*, *Tall*, Smelly ∈ *N*,*Y*. The available training data is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

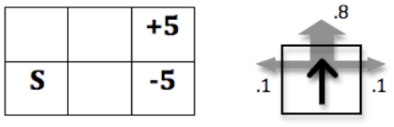
11. [4%] Using the ID3 algorithm to learn a decision tree, the root node chosen would be:
    A. Green
    B. Legs
    C. Both Height and Smelly
    D. All of the above.
    E. None of the above.

12. [4%] Once the learned concept for Martian is an ID3 decision tree, it can be converted into a set of conjunctive rules. For that tree which of following rules would be included:
    A. if (Green=Y and Legs=2 and Height=T and Smelly=N) then M;
    B. if (Green=N and Smelly=Y) then M;
    C. if (Legs=2) then M;
    D. All of the above.
    E. None of the above.

13. [4%] How is Occam's Razor employed in decision tree learning?
    A. ID3 is a greedy approach to construct the shortest tree.
    B. Occam's Razor does not apply to decision tree learning.
    C. Occam's Razor prunes nodes from the tree.
    D. All of the above.
    E. None of the above.

14. [4%] In decision tree learning, the main goal is to:
    A. Minimize the entropy of the training examples associated with each node.
    B. Maximize the entropy of the training examples associated with each node.
    C. Find the best tree structure, such that each training example is a separate tree node.
    D. All of the above.
    E. None of the above.

MDPs and Reinforcement Learning [16%] Given a Gridworld MDP as shown below, the states are grid squares, identified by row then column number. The agent always starts in the state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. The transition function is such that the intended agent movement (North, South, West, or East) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.



(a) (b)

(a) Gridworld MDP. (b) Transition function.

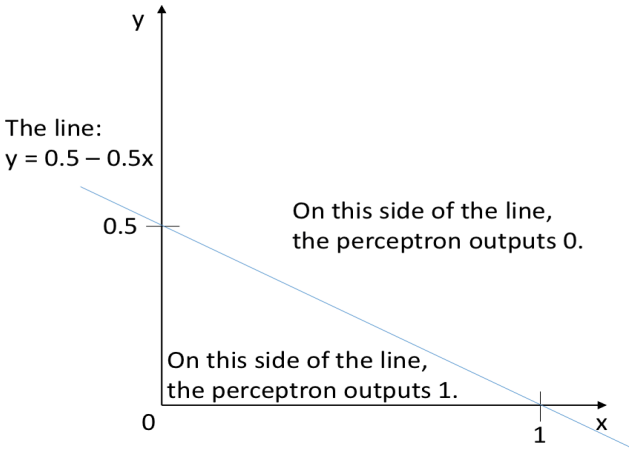
15. [4%] If the agent is deterministic, the optimal policy for this grid is:
    A. State (1,1) North, State (1,2) North, State (2,1) East, State (2,2) East
    B. State (2,1) East, State (2,2) East, State (1,1) North, State (1,2) North
    C. State (1,2) North, State (2,2) East, State (2,1) East, State (1,1) North
    D. All of the above.
    E. None of the above.

16. [4%] Now suppose the agent is stochastic and knows the transition probabilities. Give the second round of value iteration updates for each state, with a discount factor of 0.9 (Assume initial V values are 0 and compute V2 for each non-terminal state).
    A. State (1,1) is 0, State (1,2) is 0, State (2,1) is 0, State (2,2) is 3.924
    B. State (1,1) is 0, State (1,2) is 2.142, State (2,1) is 2.592, State (2,2) 3.924
    C. State (1,1) is 0, State (1,2) is 2.82528, State (2,1) is 2.82528, State (2,2) 3.924
    D. All of the above.
    E. None of the above.

17. [4%] Suppose the stochastic agent does not know the transition probabilities. What does it need in order to learn the optimal policy?
    A. Probability values
    B. Optimal Q values
    C. Training examples of the agent's movement in the grid with rewards received
    D. All of the above.
    E. None of the above.

18. [4%] Reinforcement learning can have problems with scaling because of:
    A. Sparse training data
    B. Calculations taking too much time
    C. It is hard to program
    D. All of the above
    E. None of the above

Neural Networks [16%] Design a perceptron with inputs *x* and *y* and a threshold function as activation function whose output for a given input pair (*x*, *y*) is given by:



19. [4%] The correct weights **w1** for input *x* and **w2** for input *y* and threshold **t** of the perceptron are:
   A. **w1**= -0.5, **w2**= 1, if **t**=> 0.1 outputs 1 else 0
   B. **w1**= 0.5, **w2**= 0.5, if **t**=>1 outputs 1 else 0
   C. **w1**= 0.5, **w2**= 1, if **t** <= .5 outputs 1 else 0
   D. All of the above.
   E. None of the above.

20. [4%] Given input *x* and *y* many different perceptrons can be created whose output is the line above, by making changes to:
   A. Biases
   B. Activation function
   C. Probability distribution
   D. All of the above
   E. None of the above

21. [4%] How many weights does a 2-layer feed-forward neural network with 5 input units, 3 hidden units and 2 output units contain, including the biases (dummy input weights)?
   A. 42
   B. 24
   C. 22
   D. 9
   E. None of the above

22. [4%] A perceptron is guaranteed to learn a given linearly separable function:
   A. When the activation function is a step function.
   B. Within a finite number of training steps.
   C. If all the weights are positive.
   D. All of the above.
   E. None of the above.

General AI Knowledge: True/False [8%] For each of the statements below, answer **TRUE** if the statement is **always true**, or **FALSE** otherwise.

23. [1%] **A. TRUE** **B. FALSE** Cross validation helps to avoid overfitting.

24. [1%] **A. TRUE** **B. FALSE** Unsupervised learning is learning from unlabeled data to maximize a cumulative reward.

25. [1%] **A. TRUE** **B. FALSE** Support vectors are data points that lie closest to the hyperplane.

26. [1%] **A. TRUE** **B. FALSE** A decision tree cannot represent all Boolean function.

27. [1%] **A. TRUE** **B. FALSE** Q-learning creates an explicit transition model.

28. [1%] **A. TRUE** **B. FALSE** A node X is conditionally independent of all other nodes in a Bayesian Network given its Markov Blanket.

29. [1%] **A. TRUE** **B. FALSE** A policy is a mapping from states to actions that tells the agent what action to take in any possible environment state.

30. [1%] **A. TRUE** **B. FALSE** A Bayesian Network modeling n random variables that are all mutually independent of each other would have directed arcs between every pair of variables in the network.

EXAM 7

Midterm 3 Exam

CSCI 561 Fall 2018: Artificial Intelligence

Please print neatly.

First Name:

Last Name:

Student ID:

USC email: @usc.edu

Instructions:

1. Date: **11/26/2018 from 8:00 pm to 9:50 pm**
2. Maximum credits/points/percentage for this midterm: 100
3. The percentages for each question are indicated in square brackets [ ].
4. **No books** (or any other material) are allowed.
5. Write down your name, student ID and USC email address on both test and answer booklets.
6. Your exam will be scanned and uploaded online.
7. Write within the spaces provided for your answers in the test answer booklet.
8. **Do NOT write on the 2D barcode.**
9. **The back of the exam pages will NOT be graded. You may use it for scratch.**
10. No questions during the exam. If something is unclear to you, write that in your exam.
11. Be brief: a few words are often enough if they are precise and use the correct vocabulary studied in class.
12. When finished before 9:40 pm, raise completed exam sheets until approached by a proctor.
13. At 9:50 pm, put down your pencil. Stay seated and pass your exam to the nearest aisle or **receive 10% exam penalty.**
14. **Adhere to the Academic Integrity code.**

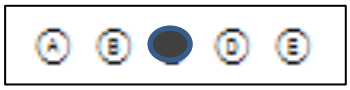
Problems	**100 Percent total**
PART A Short Answer	
AI Applications	4
PART B Multiple Choice, True/False	
General AI Knowledge	8
Bayesian Network	8
Probability	16
Decision Tree	16
MDP and Reinforcement Learning	16
Neural Network	16
Decision Network	16

PART A – Short Answer

AI Applications [4%]

During week 13 there were three presentations about AI applications: Chris Stewart from Google, Ben Marcus from AirMap, and Prof. Wei-min Shen. Choose one of these presentations and briefly describe how the application applies AI to reasoning with uncertainty. (One or two sentences max in the space provide on the answer sheet)

PART B – Multiple Choice & True/False



In this section choose the best option **A**, **B**, **C**, **D**, or **E** and fill in the appropriate circle in the answer sheet. Correct answers receive the % shown in the []. Answers left blank receive 0% and **wrong answers receive -1%**. Fill in at most one circle per problem.

General AI Knowledge: True/False [8%] For each of the statements below, answer **TRUE** if the statement is **always true**, or **FALSE** otherwise.

1. [1%] **A. TRUE B. FALSE** Cross validation helps to avoid overfitting.

2. [1%] **A. TRUE B. FALSE** Unsupervised learning is learning from unlabeled data to maximize a cumulative reward.

3. [1%] **A. TRUE B. FALSE** Support vectors are data points that lie closest to the hyperplane.

4. [1%] **A. TRUE B. FALSE** A decision tree cannot represent all Boolean function.

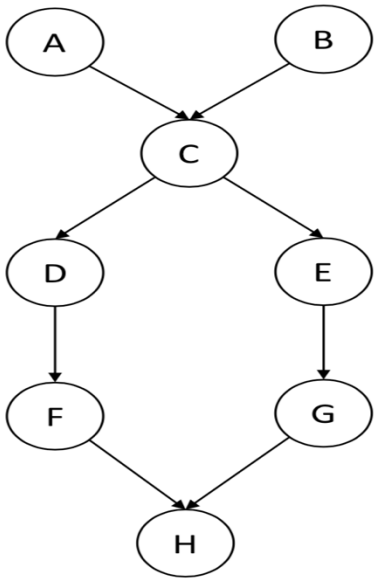
5. [1%] **A. TRUE B. FALSE** Q-learning creates an explicit transition model.

6. [1%] **A. TRUE B. FALSE** A node *X* is conditionally independent of all other nodes in a Bayesian Network given its Markov Blanket.

7. [1%] **A. TRUE B. FALSE** A policy is a mapping from states to actions that tells the agent what action to take in any possible environment state.

8. [1%] **A. TRUE B. FALSE** A Bayesian Network modeling *n* random variables that are all mutually independent of each other would have directed arcs between every pair of variables in the network.

Bayesian Network [8%]



9. [4%] For the Bayesian network given above which statement[s] is true:
   A. A and B are independent given C.
   B. A and B are independent.
   C. D is conditionally independent of G given C.
   D. All of the above.
   E. None of the above.

10. [4%] Which statement[s] is true in any Bayesian Network:
   A. A node is conditionally independent of all other nodes given its Markov blanket.
   B. A node is independent of its non-descendants given its parents.
   C. At least one node in the network must have no incoming arcs.
   D. All of the above.
   E. None of the above.

Probability [16%]

For questions 11 & 12 consider the following joint probability table for the boolean random variables **Q** and **S**:

Q	**S**	
T	T	0.1
T	F	0.2
F	T	0.3
F	F	0.4

11. [4%] Which probability value is the largest?
    A. P(Q)
    B. P(¬ S)
    C. P(Q ∨ S)
    D. P(Q ⇒ S)
    E. P(Q | S)

12. [4%] Which of the following statement[s] is true:
    A. P(Q ⇒ S) > P(S | Q)
    B. P(Q ∨ S) + P(Q , S) = 1
    C. P(¬Q) = 1- P(¬ S)
    D. All of the above.
    E. None of the above.

13. [4%] If random variable *X* is conditionally independent of *Y* given *Z* then
    A. P(X,Y|Z)=P(X|Z)P(Y|Z)
    B. P(X|Y,Z)=P(X|Z)
    C. P(Y|X,Z)=P(Y|Z)
    D. All of the above.
    E. None of the above

14. [4%] Assume that the boolean random variables TEST1 and TEST2 are conditionally independent given FLU and the following probabilities: P(FLU) = 0.1, P(TEST1 | FLU) = 0.9, P(TEST1 | ¬ FLU) = 0.2, P(TEST2 | FLU) = 0.8, P(TEST2 | ¬ FLU) = 0.1

What is the probability of FLU is true given TEST1 and TEST2 are both true:
    A. .25
    B. .79
    C. .0002
    D. All of the above.
    E. None of the above.

Decision Tree [16%] NASA wants to train a robot to recognize two species Martians (M) and Humans (H) based on the following characteristics: Green ∈ *N*,*Y*, *Legs* ∈ 2,3, *Height* ∈ *Short*, *Tall*, Smelly ∈ *N*,*Y*. The available training data is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

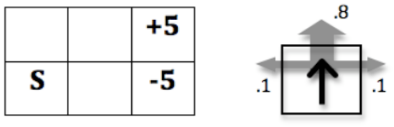
15. [4%] Using the ID3 algorithm to learn a decision tree, the root node chosen would be:
    A. Green
    B. Legs
    C. Both Height and Smelly
    D. All of the above.
    E. None of the above.

16. [4%] Once the learned concept for Martian is an ID3 decision tree, it can be converted into a set of conjunctive rules. For that tree which of following rules would be included:
    A. if (Green=Y and Legs=2 and Height=T and Smelly=N) then M;
    B. if (Green=N and Smelly=Y) then M;
    C. if (Legs=2) then M;
    D. All of the above.
    E. None of the above.

17. [4%] How is Occam's Razor employed in decision tree learning?
    A. ID3 is a greedy approach to construct the shortest tree.
    B. Occam's Razor does not apply to decision tree learning.
    C. Occam's Razor prunes nodes from the tree.
    D. All of the above.
    E. None of the above.

18. [4%] In decision tree learning, the main goal is to:
    A. Minimize the entropy of the training examples associated with each node.
    B. Maximize the entropy of the training examples associated with each node.
    C. Find the best tree structure, such that each training example is a separate tree node.
    D. All of the above.
    E. None of the above.

MDPs and Reinforcement Learning [16%] Given a Gridworld MDP as shown below, the states are grid squares, identified by row then column number. The agent always starts in the state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. The transition function is such that the intended agent movement (North, South, West, or East) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.



(a) (b)

(a) Gridworld MDP. (b) Transition function.

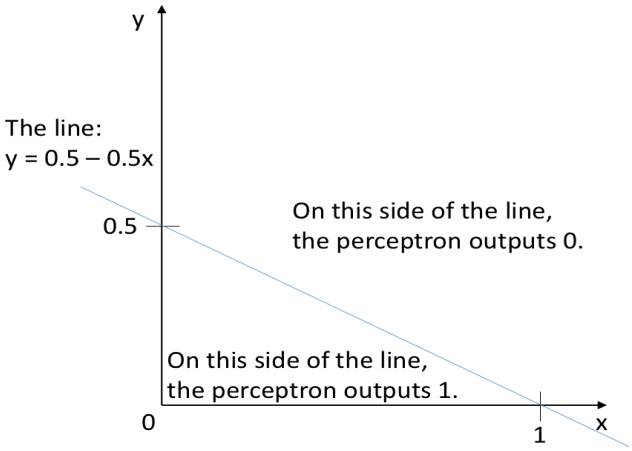
19. [4%] If the agent is deterministic, the optimal policy for this grid is:
    A. State (1,1) North, State (1,2) North, State (2,1) East, State (2,2) East
    B. State (2,1) East, State (2,2) East, State (1,1) North, State (1,2) North
    C. State (1,2) North, State (2,2) East, State (2,1) East, State (1,1) North
    D. All of the above.
    E. None of the above.

20. [4%] Now suppose the agent is stochastic and knows the transition probabilities. Give the second round of value iteration updates for each state, with a discount factor of 0.9 (Assume initial V values are 0 and compute V2 for each non-terminal state).
    A. State (1,1) is 0, State (1,2) is 0, State (2,1) is 0, State (2,2) is 3.924
    B. State (1,1) is 0, State (1,2) is 2.142, State (2,1) is 2.592, State (2,2) 3.924
    C. State (1,1) is 0, State (1,2) is 2.82528, State (2,1) is 2.82528, State (2,2) 3.924
    D. All of the above.
    E. None of the above.

21. [4%] Suppose the stochastic agent does not know the transition probabilities. What does it need in order to learn the optimal policy?
    A. Probability values
    B. Optimal Q values
    C. Training examples of the agent's movement in the grid with rewards received
    D. All of the above.
    E. None of the above.

22. [4%] Reinforcement learning can have problems with scaling because of:
    A. Sparse training data
    B. Calculations taking too much time
    C. It is hard to program
    D. All of the above
    E. None of the above

Neural Networks [16%] Design a perceptron with inputs *x* and *y* and a threshold function as activation function whose output for a given input pair (*x*, *y*) is given by:



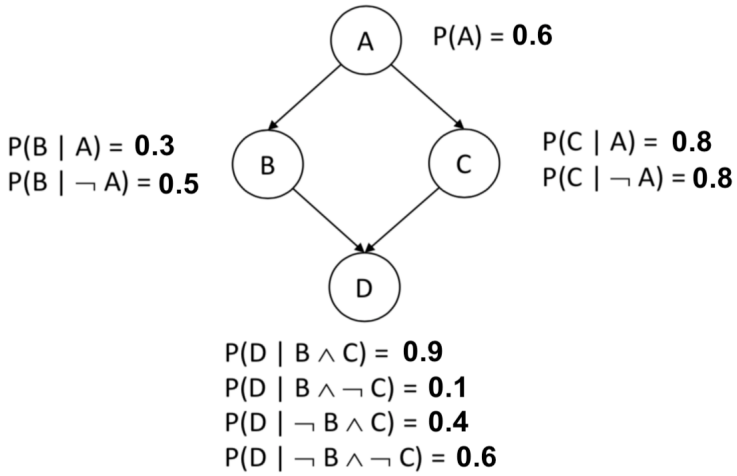
23. [4%] The correct weights **w1** for input *x* and **w2** for input *y* and threshold **t** of the perceptron are:
   A. **w1**= -0.5, **w2**= 1, if **t**=> 0.1 outputs 1 else 0
   B. **w1**= 0.5, **w2**= 0.5, if **t**=>1 outputs 1 else 0
   C. **w1**= 0.5, **w2**= 1, if **t** <= .5 outputs 1 else 0
   D. All of the above.
   E. None of the above.

24. [4%] Given input *x* and *y* many different perceptrons can be created whose output is the line above, by making changes to:
   A. Biases
   B. Activation function
   C. Probability distribution
   D. All of the above
   E. None of the above

25. [4%] How many weights does a 2-layer feed-forward neural network with 5 input units, 3 hidden units and 2 output units contain, including the biases (dummy input weights)?
   A. 42
   B. 24
   C. 22
   D. 9
   E. None of the above

26. [4%] A perceptron is guaranteed to learn a given linearly separable function:
   A. When the activation function is a step function.
   B. Within a finite number of training steps.
   C. If all the weights are positive.
   D. All of the above.
   E. None of the above.

Decision Network [16%] Consider the following Bayesian network. A, B, C, and D are boolean random variables.



27. [4%] Given that A is true and B is false, what is the probability of D being true?
    A. .004
    B. .0034
    C. .36
    D. All of the above.
    E. None of the above.

28. [4%] Converting A into a decision node in the above network and given the following utility function: if D is true then utility is 10 and if D is false utility is 100, which decision for A will give the higher expected utility?
    A. A is true
    B. A is false
    C. Utility is equal in both decisions of A.
    D. All of the above.
    E. None of the above.

29. [4%] In the decision network described in question 28 which of the following events will increase the expected utility the most?
    A. B is true
    B. B is false
    C. C is false
    D. All of the above.
    E. None of the above.

30. [4%] Variable elimination is an algorithm for:
    A. Approximate probabilistic inference in decision networks.
    B. Exact probabilistic inference.
    C. Computing optimal policies.
    D. Reinforcement learning in unknown environments.
    E. None of the above.

EXAM 8

Midterm 3 Exam

CSCI 561 Fall 2018: Artificial Intelligence

Please print neatly.

First Name:

Last Name:

Student ID:

USC email:

								@usc.edu

Instructions:

14. Date: **11/26/2018 from 8:00 pm to 9:50 pm**
15. Maximum credits/points/percentage for this midterm: 100
16. The percentages for each question are indicated in square brackets [ ].
17. **No books** (or any other material) are allowed.
18. Write down your name, student ID and USC email address on both test and answer booklets.
19. Your exam will be scanned and uploaded online.
20. Write within the spaces provided for your answers in the test answer booklet.
21. **Do NOT write on the 2D barcode.**
22. **The back of the exam pages will NOT be graded. You may use it for scratch.**
23. No questions during the exam. If something is unclear to you, write that in your exam.
24. Be brief: a few words are often enough if they are precise and use the correct vocabulary studied in class.
25. When finished before 9:40 pm, raise completed exam sheets until approached by a proctor.
26. At 9:50 pm, put down your pencil. Stay seated and pass your exam to the nearest aisle or **receive 10% exam penalty.**
14. **Adhere to the Academic Integrity code.**

Problems	**100 Percent total**
PART A Short Answer	
Probabilistic Reasoning	4
PART B Multiple Choice, True/False	
Bayesian Network	8
Probability	16
Decision Network	16
Decision Tree	16
MDP and Reinforcement Learning	16
Neural Network	16
General AI Knowledge	8

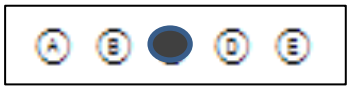
PART A – Short Answer

Probabilistic Reasoning [4%] Assume that the boolean random variables **J** and **K** are conditionally independent given **L** and with the following probabilities:

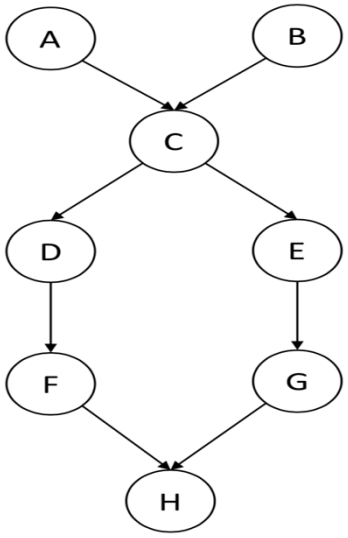
P(L) = 0.1, P(J | L) = 0.9, P(J | ¬ L) = 0.2, P(K | L) = 0.8, P(K | ¬ L) = 0.1

Draw the Bayesian Network in space provide in the answer sheet.

PART B – Multiple Choice & True/False



In this section choose the best option **A**, **B**, **C**, **D**, or **E** and fill in the appropriate circle in the answer sheet. Correct answers receive the % shown in the []. Answers left blank receive 0% and **wrong answers receive -1%**. Fill in at most one circle per problem.



Bayesian Network [8%]

1. [4%] For the Bayesian network given which statement[s] is true:
   A. A and B are independent given C.
   B. A and B are independent.
   C. D is conditionally independent of G given C.
   D. All of the above.
   E. None of the above.

2. [4%] Which statement[s] is true in any Bayesian Network:
   A. A node is conditionally independent of all other nodes given its Markov blanket.
   B. A node is independent of its non-descendants given its parents.
   C. At least one node in the network must have no incoming arcs.
   D. All of the above.
   E. None of the above.

Probability [16%]

For questions 3 & 4 consider the following joint probability table for the boolean random variables **Q** and **S**:

Q	**S**	
T	T	0.1
T	F	0.2
F	T	0.3
F	F	0.4

3. [4%] Which probability value is the largest?
   A. P(Q)
   B. P(¬ S)
   C. P(Q ∨ S)
   D. P(Q ⇒ S)
   E. P(Q | S)

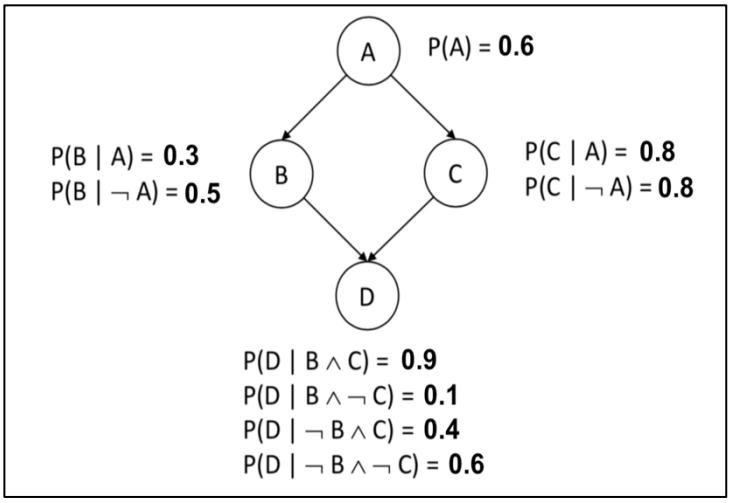
4. [4%] Which of the following statement[s] is true:
   A. P(Q ⇒ S) > P(S | Q)
   B. P(Q ∨ S) + P(Q , S) = 1
   C. P(¬Q) = 1- P(¬ S)
   D. All of the above.
   E. None of the above.

5. [4%] If random variable *X* is conditionally independent of *Y* given *Z* then
   A. P(X,Y|Z)=P(X|Z)P(Y|Z)
   B. P(X|Y,Z)=P(X|Z)
   C. P(Y|X,Z)=P(Y|Z)
   D. All of the above.
   E. None of the above

6. [4%] Assume that the boolean random variables TEST1 and TEST2 are conditionally independent given FLU and the following probabilities: P(FLU) = 0.1, P(TEST1 | FLU) = 0.9, P(TEST1 | ¬ FLU) = 0.2, P(TEST2 | FLU) = 0.8, P(TEST2 | ¬ FLU) = 0.1

What is the probability of FLU is true given TEST1 and TEST2 are both true:
   A. .25
   B. .79
   C. .0002
   D. All of the above.
   E. None of the above.

Decision Network [16%] Consider the following Bayesian network. A, B, C, and D are boolean random variables.



7. [4%] Given that A is true and B is false, what is the probability of D being true?
A. .004
B. .0034
C. .36
D. All of the above.
E. None of the above.

8. [4%] Converting A into a decision node in the above network and given the following utility function: if D is true then utility is 10 and if D is false utility is 100, which decision for A will give the higher expected utility?

A. A is true
B. A is false
C. Utility is equal in both decisions of A.
D. All of the above.
E. None of the above.

9. [4%] In the decision network described in question 28 which of the following events will increase the expected utility the most?

A. B is true
B. B is false
C. C is false
D. All of the above.
E. None of the above.

10. [4%] Variable elimination is an algorithm for:

A. Approximate probabilistic inference in decision networks.
B. Exact probabilistic inference.
C. Computing optimal policies.
D. Reinforcement learning in unknown environments.
E. None of the above.

Decision Tree [16%] NASA wants to train a robot to recognize two species Martians (M) and Humans (H) based on the following characteristics: Green ∈ *N*,*Y*, *Legs* ∈ 2,3, *Height* ∈ *Short*, *Tall*, Smelly ∈ *N*,*Y*. The available training data is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

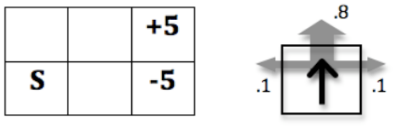
11. [4%] Using the ID3 algorithm to learn a decision tree, the root node chosen would be:
    A. Green
    B. Legs
    C. Both Height and Smelly
    D. All of the above.
    E. None of the above.

12. [4%] Once the learned concept for Martian is an ID3 decision tree, it can be converted into a set of conjunctive rules. For that tree which of following rules would be included:
    A. if (Green=Y and Legs=2 and Height=T and Smelly=N) then M;
    B. if (Green=N and Smelly=Y) then M;
    C. if (Legs=2) then M;
    D. All of the above.
    E. None of the above.

13. [4%] How is Occam's Razor employed in decision tree learning?
    A. ID3 is a greedy approach to construct the shortest tree.
    B. Occam's Razor does not apply to decision tree learning.
    C. Occam's Razor prunes nodes from the tree.
    D. All of the above.
    E. None of the above.

14. [4%] In decision tree learning, the main goal is to:
    A. Minimize the entropy of the training examples associated with each node.
    B. Maximize the entropy of the training examples associated with each node.
    C. Find the best tree structure, such that each training example is a separate tree node.
    D. All of the above.
    E. None of the above.

MDPs and Reinforcement Learning [16%] Given a Gridworld MDP as shown below, the states are grid squares, identified by row then column number. The agent always starts in the state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. The transition function is such that the intended agent movement (North, South, West, or East) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.



(a) (b)

(a) Gridworld MDP. (b) Transition function.

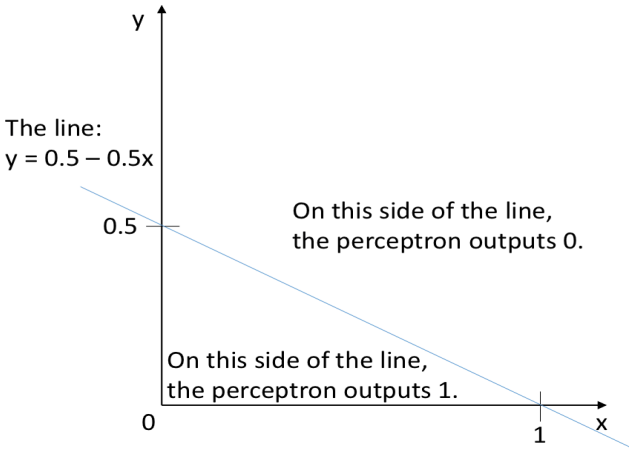
15. [4%] If the agent is deterministic, the optimal policy for this grid is:
    A. State (1,1) North, State (1,2) North, State (2,1) East, State (2,2) East
    B. State (2,1) East, State (2,2) East, State (1,1) North, State (1,2) North
    C. State (1,2) North, State (2,2) East, State (2,1) East, State (1,1) North
    D. All of the above.
    E. None of the above.

16. [4%] Now suppose the agent is stochastic and knows the transition probabilities. Give the second round of value iteration updates for each state, with a discount factor of 0.9 (Assume initial V values are 0 and compute V2 for each non-terminal state).
    A. State (1,1) is 0, State (1,2) is 0, State (2,1) is 0, State (2,2) is 3.924
    B. State (1,1) is 0, State (1,2) is 2.142, State (2,1) is 2.592, State (2,2) 3.924
    C. State (1,1) is 0, State (1,2) is 2.82528, State (2,1) is 2.82528, State (2,2) 3.924
    D. All of the above.
    E. None of the above.

17. [4%] Suppose the stochastic agent does not know the transition probabilities. What does it need in order to learn the optimal policy?
    A. Probability values
    B. Optimal Q values
    C. Training examples of the agent's movement in the grid with rewards received
    D. All of the above.
    E. None of the above.

18. [4%] Reinforcement learning can have problems with scaling because of:
    A. Sparse training data
    B. Calculations taking too much time
    C. It is hard to program
    D. All of the above
    E. None of the above

Neural Networks [16%] Design a perceptron with inputs *x* and *y* and a threshold function as activation function whose output for a given input pair (*x*, *y*) is given by:



19. [4%] The correct weights **w1** for input *x* and **w2** for input *y* and threshold **t** of the perceptron are:
   A. **w1**= -0.5, **w2**= 1, if **t**=> 0.1 outputs 1 else 0
   B. **w1**= 0.5, **w2**= 0.5, if **t**=>1 outputs 1 else 0
   C. **w1**= 0.5, **w2**= 1, if **t** <= .5 outputs 1 else 0
   D. All of the above.
   E. None of the above.

20. [4%] Given input *x* and *y* many different perceptrons can be created whose output is the line above, by making changes to:
   A. Biases
   B. Activation function
   C. Probability distribution
   D. All of the above
   E. None of the above

21. [4%] How many weights does a 2-layer feed-forward neural network with 5 input units, 3 hidden units and 2 output units contain, including the biases (dummy input weights)?
   A. 42
   B. 24
   C. 22
   D. 9
   E. None of the above

22. [4%] A perceptron is guaranteed to learn a given linearly separable function:
   A. When the activation function is a step function.
   B. Within a finite number of training steps.
   C. If all the weights are positive.
   D. All of the above.
   E. None of the above.

General AI Knowledge: True/False [8%] For each of the statements below, answer **TRUE** if the statement is **always true**, or **FALSE** otherwise.

23. [1%] **A. TRUE** **B. FALSE** Cross validation helps to avoid overfitting.

24. [1%] **A. TRUE** **B. FALSE** Unsupervised learning is learning from unlabeled data to maximize a cumulative reward.

25. [1%] **A. TRUE** **B. FALSE** Support vectors are data points that lie closest to the hyperplane.

26. [1%] **A. TRUE** **B. FALSE** A decision tree cannot represent all Boolean function.

27. [1%] **A. TRUE** **B. FALSE** Q-learning creates an explicit transition model.

28. [1%] **A. TRUE** **B. FALSE** A node X is conditionally independent of all other nodes in a Bayesian Network given its Markov Blanket.

29. [1%] **A. TRUE** **B. FALSE** A policy is a mapping from states to actions that tells the agent what action to take in any possible environment state.

30. [1%] **A. TRUE** **B. FALSE** A Bayesian Network modeling n random variables that are all mutually independent of each other would have directed arcs between every pair of variables in the network.

EXAM 9

Midterm 3 Exam

CSCI 561 Fall 2018: Artificial Intelligence

Please print neatly.

First Name:

Last Name:

Student ID:

USC email:

								@usc.edu

Instructions:

1. Date: **11/26/2018 from 8:00 pm to 9:50 pm**
2. Maximum credits/points/percentage for this midterm: 100
3. The percentages for each question are indicated in square brackets [ ].
4. **No books** (or any other material) are allowed.
5. Write down your name, student ID and USC email address on both test and answer booklets.
6. Your exam will be scanned and uploaded online.
7. Write within the spaces provided for your answers in the test answer booklet.
8. **Do NOT write on the 2D barcode.**
9. **The back of the exam pages will NOT be graded. You may use it for scratch.**
10. No questions during the exam. If something is unclear to you, write that in your exam.
11. Be brief: a few words are often enough if they are precise and use the correct vocabulary studied in class.
12. When finished before 9:40 pm, raise completed exam sheets until approached by a proctor.
13. At 9:50 pm, put down your pencil. Stay seated and pass your exam to the nearest aisle or **receive 10% exam penalty.**
14. **Adhere to the Academic Integrity code.**

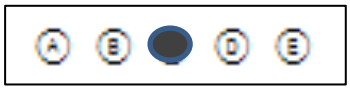
Problems	**100 Percent total**
PART A Short Answer	
AI Applications	4
PART B Multiple Choice, True/False	
General AI Knowledge	8
Bayesian Network	8
Probability	16
Decision Tree	16
MDP and Reinforcement Learning	16
Neural Network	16
Decision Network	16

PART A – Short Answer

AI Applications [4%]

During week 13 there were three presentations about AI applications: Chris Stewart from Google, Ben Marcus from AirMap, and Prof. Wei-min Shen. Choose one of these presentations and briefly describe how the application applies AI to reasoning with uncertainty. (One or two sentences max in the space provide on the answer sheet)

PART B – Multiple Choice & True/False



In this section choose the best option **A**, **B**, **C**, **D**, or **E** and fill in the appropriate circle in the answer sheet. Correct answers receive the % shown in the []. Answers left blank receive 0% and **wrong answers receive -1%**. Fill in at most one circle per problem.

General AI Knowledge: True/False [8%] For each of the statements below, answer **TRUE** if the statement is **always true**, or **FALSE** otherwise.

1. [1%] **A. TRUE B. FALSE** Cross validation helps to avoid overfitting.

2. [1%] **A. TRUE B. FALSE** Unsupervised learning is learning from unlabeled data to maximize a cumulative reward.

3. [1%] **A. TRUE B. FALSE** Support vectors are data points that lie closest to the hyperplane.

4. [1%] **A. TRUE B. FALSE** A decision tree cannot represent all Boolean function.

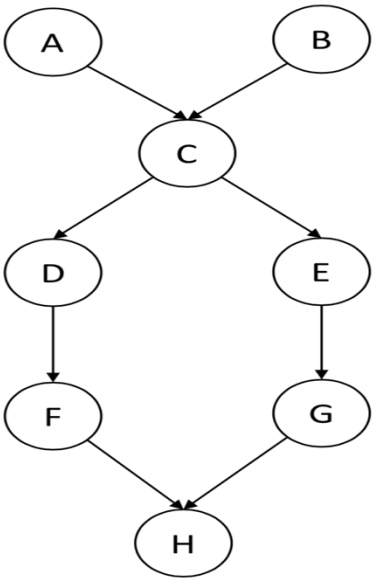
5. [1%] **A. TRUE B. FALSE** Q-learning creates an explicit transition model.

6. [1%] **A. TRUE B. FALSE** A node *X* is conditionally independent of all other nodes in a Bayesian Network given its Markov Blanket.

7. [1%] **A. TRUE B. FALSE** A policy is a mapping from states to actions that tells the agent what action to take in any possible environment state.

8. [1%] **A. TRUE B. FALSE** A Bayesian Network modeling *n* random variables that are all mutually independent of each other would have directed arcs between every pair of variables in the network.

Bayesian Network [8%]



9. [4%] For the Bayesian network given above which statement[s] is true:
   A. A and B are independent given C.
   B. A and B are independent.
   C. D is conditionally independent of G given C.
   D. All of the above.
   E. None of the above.

10. [4%] Which statement[s] is true in any Bayesian Network:
   A. A node is conditionally independent of all other nodes given its Markov blanket.
   B. A node is independent of its non-descendants given its parents.
   C. At least one node in the network must have no incoming arcs.
   D. All of the above.
   E. None of the above.

Probability [16%]

For questions 11 & 12 consider the following joint probability table for the boolean random variables **Q** and **S**:

Q	**S**	
T	T	0.1
T	F	0.2
F	T	0.3
F	F	0.4

11. [4%] Which probability value is the largest?
    A. P(Q)
    B. P(¬ S)
    C. P(Q ∨ S)
    D. P(Q ⇒ S)
    E. P(Q | S)

12. [4%] Which of the following statement[s] is true:
    A. P(Q ⇒ S) > P(S | Q)
    B. P(Q ∨ S) + P(Q , S) = 1
    C. P(¬Q) = 1- P(¬ S)
    D. All of the above.
    E. None of the above.

13. [4%] If random variable *X* is conditionally independent of *Y* given *Z* then
    A. P(X,Y|Z)=P(X|Z)P(Y|Z)
    B. P(X|Y,Z)=P(X|Z)
    C. P(Y|X,Z)=P(Y|Z)
    D. All of the above.
    E. None of the above

14. [4%] Assume that the boolean random variables TEST1 and TEST2 are conditionally independent given FLU and the following probabilities: P(FLU) = 0.1, P(TEST1 | FLU) = 0.9, P(TEST1 | ¬ FLU) = 0.2, P(TEST2 | FLU) = 0.8, P(TEST2 | ¬ FLU) = 0.1

What is the probability of FLU is true given TEST1 and TEST2 are both true:
    A. .25
    B. .79
    C. .0002
    D. All of the above.
    E. None of the above.

Decision Tree [16%] NASA wants to train a robot to recognize two species Martians (M) and Humans (H) based on the following characteristics: Green ∈ *N*,*Y*, *Legs* ∈ 2,3, *Height* ∈ *Short*, *Tall*, Smelly ∈ *N*,*Y*. The available training data is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

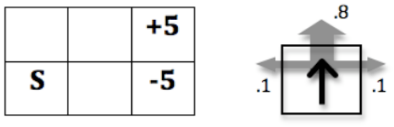
15. [4%] Using the ID3 algorithm to learn a decision tree, the root node chosen would be:
    A. Green
    B. Legs
    C. Both Height and Smelly
    D. All of the above.
    E. None of the above.

16. [4%] Once the learned concept for Martian is an ID3 decision tree, it can be converted into a set of conjunctive rules. For that tree which of following rules would be included:
    A. if (Green=Y and Legs=2 and Height=T and Smelly=N) then M;
    B. if (Green=N and Smelly=Y) then M;
    C. if (Legs=2) then M;
    D. All of the above.
    E. None of the above.

17. [4%] How is Occam's Razor employed in decision tree learning?
    A. ID3 is a greedy approach to construct the shortest tree.
    B. Occam's Razor does not apply to decision tree learning.
    C. Occam's Razor prunes nodes from the tree.
    D. All of the above.
    E. None of the above.

18. [4%] In decision tree learning, the main goal is to:
    A. Minimize the entropy of the training examples associated with each node.
    B. Maximize the entropy of the training examples associated with each node.
    C. Find the best tree structure, such that each training example is a separate tree node.
    D. All of the above.
    E. None of the above.

MDPs and Reinforcement Learning [16%] Given a Gridworld MDP as shown below, the states are grid squares, identified by row then column number. The agent always starts in the state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. The transition function is such that the intended agent movement (North, South, West, or East) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.



(a) (b)

(a) Gridworld MDP. (b) Transition function.

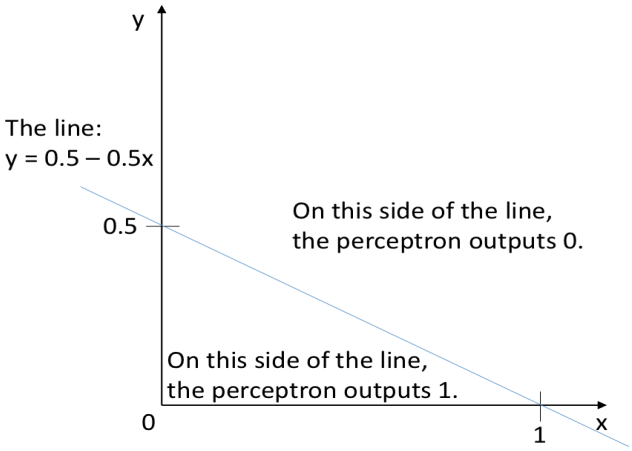
19. [4%] If the agent is deterministic, the optimal policy for this grid is:
    A. State (1,1) North, State (1,2) North, State (2,1) East, State (2,2) East
    B. State (2,1) East, State (2,2) East, State (1,1) North, State (1,2) North
    C. State (1,2) North, State (2,2) East, State (2,1) East, State (1,1) North
    D. All of the above.
    E. None of the above.

20. [4%] Now suppose the agent is stochastic and knows the transition probabilities. Give the second round of value iteration updates for each state, with a discount factor of 0.9 (Assume initial V values are 0 and compute V2 for each non-terminal state).
    A. State (1,1) is 0, State (1,2) is 0, State (2,1) is 0, State (2,2) is 3.924
    B. State (1,1) is 0, State (1,2) is 2.142, State (2,1) is 2.592, State (2,2) 3.924
    C. State (1,1) is 0, State (1,2) is 2.82528, State (2,1) is 2.82528, State (2,2) 3.924
    D. All of the above.
    E. None of the above.

21. [4%] Suppose the stochastic agent does not know the transition probabilities. What does it need in order to learn the optimal policy?
    A. Probability values
    B. Optimal Q values
    C. Training examples of the agent's movement in the grid with rewards received
    D. All of the above.
    E. None of the above.

22. [4%] Reinforcement learning can have problems with scaling because of:
    A. Sparse training data
    B. Calculations taking too much time
    C. It is hard to program
    D. All of the above
    E. None of the above

Neural Networks [16%] Design a perceptron with inputs *x* and *y* and a threshold function as activation function whose output for a given input pair (*x*, *y*) is given by:



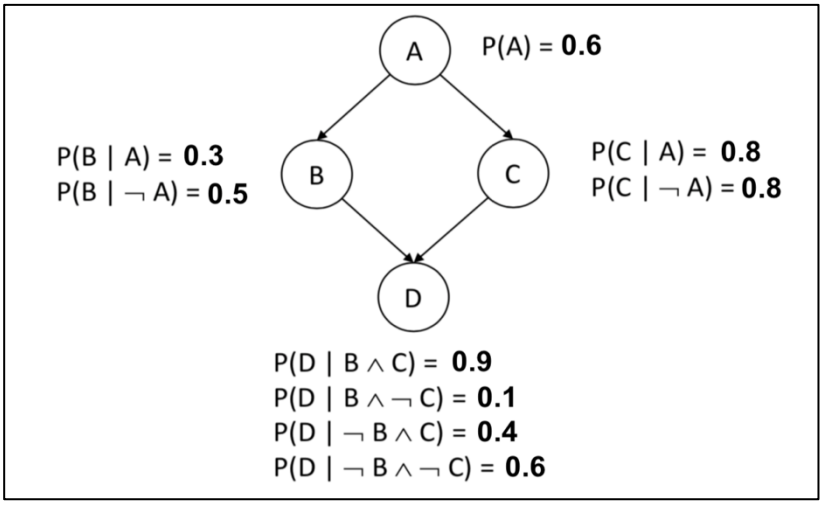
23. [4%] The correct weights **w1** for input *x* and **w2** for input *y* and threshold **t** of the perceptron are:
   A. **w1**= -0.5, **w2**= 1, if **t**=> 0.1 outputs 1 else 0
   B. **w1**= 0.5, **w2**= 0.5, if **t**=>1 outputs 1 else 0
   C. **w1**= 0.5, **w2**= 1, if **t** <= .5 outputs 1 else 0
   D. All of the above.
   E. None of the above.

24. [4%] Given input *x* and *y* many different perceptrons can be created whose output is the line above, by making changes to:
   A. Biases
   B. Activation function
   C. Probability distribution
   D. All of the above
   E. None of the above

25. [4%] How many weights does a 2-layer feed-forward neural network with 5 input units, 3 hidden units and 2 output units contain, including the biases (dummy input weights)?
   A. 42
   B. 24
   C. 22
   D. 9
   E. None of the above

26. [4%] A perceptron is guaranteed to learn a given linearly separable function:
   A. When the activation function is a step function.
   B. Within a finite number of training steps.
   C. If all the weights are positive.
   D. All of the above.
   E. None of the above.

Decision Network [16%] Consider the following Bayesian network. A, B, C, and D are boolean random variables.



27. [4%] Given that A is true and B is false, what is the probability of D being true?
    A. .004
    B. .0034
    C. .36
    D. All of the above.
    E. None of the above.

28. [4%] Converting A into a decision node in the above network and given the following utility function: if D is true then utility is 10 and if D is false utility is 100, which decision for A will give the higher expected utility?
    A. A is true
    B. A is false
    C. Utility is equal in both decisions of A.
    D. All of the above.
    E. None of the above.

29. [4%] In the decision network described in question 28 which of the following events will increase the expected utility the most?
    A. B is true
    B. B is false
    C. C is false
    D. All of the above.
    E. None of the above.

30. [4%] Variable elimination is an algorithm for:
    A. Approximate probabilistic inference in decision networks.
    B. Exact probabilistic inference.
    C. Computing optimal policies.
    D. Reinforcement learning in unknown environments.
    E. None of the above.

EXAM 10

Midterm 3 Exam

CSCI 561 Fall 2018: Artificial Intelligence

Please print neatly.

First Name:

Last Name:

Student ID:

USC email:

								@usc.edu

Instructions:

14. Date: **11/26/2018 from 8:00 pm to 9:50 pm**
15. Maximum credits/points/percentage for this midterm: 100
16. The percentages for each question are indicated in square brackets [ ].
17. **No books** (or any other material) are allowed.
18. Write down your name, student ID and USC email address on both test and answer booklets.
19. Your exam will be scanned and uploaded online.
20. Write within the spaces provided for your answers in the test answer booklet.
21. **Do NOT write on the 2D barcode.**
22. **The back of the exam pages will NOT be graded. You may use it for scratch.**
23. No questions during the exam. If something is unclear to you, write that in your exam.
24. Be brief: a few words are often enough if they are precise and use the correct vocabulary studied in class.
25. When finished before 9:40 pm, raise completed exam sheets until approached by a proctor.
26. At 9:50 pm, put down your pencil. Stay seated and pass your exam to the nearest aisle or **receive 10% exam penalty.**
14. **Adhere to the Academic Integrity code.**

Problems	**100 Percent total**
PART A Short Answer	
Probabilistic Reasoning	4
PART B Multiple Choice, True/False	
Bayesian Network	8
Probability	16
Decision Network	16
Decision Tree	16
MDP and Reinforcement Learning	16
Neural Network	16
General AI Knowledge	8

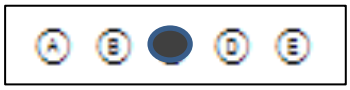
PART A – Short Answer

Probabilistic Reasoning [4%] Assume that the boolean random variables **J** and **K** are conditionally independent given **L** and with the following probabilities:

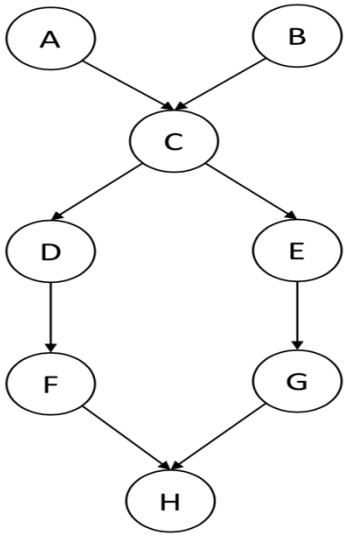
P(L) = 0.1, P(J | L) = 0.9, P(J | ¬ L) = 0.2, P(K | L) = 0.8, P(K | ¬ L) = 0.1

Draw the Bayesian Network in space provide in the answer sheet.

PART B – Multiple Choice & True/False



In this section choose the best option **A**, **B**, **C**, **D**, or **E** and fill in the appropriate circle in the answer sheet. Correct answers receive the % shown in the []. Answers left blank receive 0% and **wrong answers receive -1%**. Fill in at most one circle per problem.



Bayesian Network [8%]

1. [4%] For the Bayesian network given which statement[s] is true:
   A. A and B are independent given C.
   B. A and B are independent.
   C. D is conditionally independent of G given C.
   D. All of the above.
   E. None of the above.

2. [4%] Which statement[s] is true in any Bayesian Network:
   A. A node is conditionally independent of all other nodes given its Markov blanket.
   B. A node is independent of its non-descendants given its parents.
   C. At least one node in the network must have no incoming arcs.
   D. All of the above.
   E. None of the above.

Probability [16%]

For questions 3 & 4 consider the following joint probability table for the boolean random variables **Q** and **S**:

Q	**S**	
T	T	0.1
T	F	0.2
F	T	0.3
F	F	0.4

3. [4%] Which probability value is the largest?
   A. P(Q)
   B. P(¬ S)
   C. P(Q ∨ S)
   D. P(Q ⇒ S)
   E. P(Q | S)

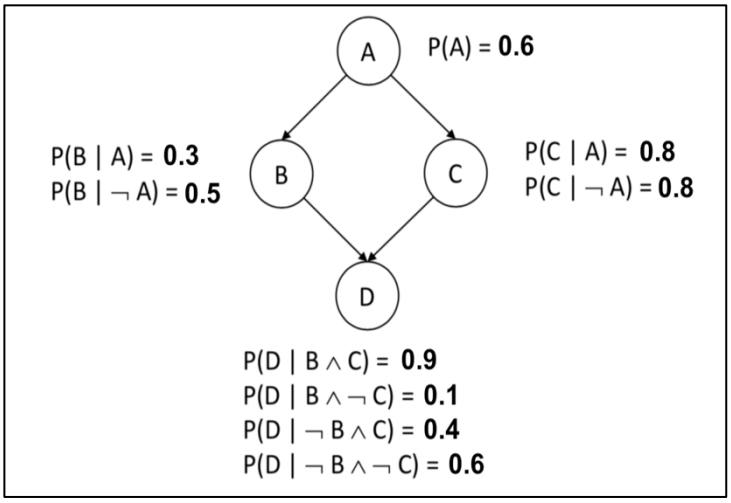
4. [4%] Which of the following statement[s] is true:
   A. P(Q ⇒ S) > P(S | Q)
   B. P(Q ∨ S) + P(Q , S) = 1
   C. P(¬Q) = 1- P(¬ S)
   D. All of the above.
   E. None of the above.

5. [4%] If random variable *X* is conditionally independent of *Y* given *Z* then
   A. P(X,Y|Z)=P(X|Z)P(Y|Z)
   B. P(X|Y,Z)=P(X|Z)
   C. P(Y|X,Z)=P(Y|Z)
   D. All of the above.
   E. None of the above

6. [4%] Assume that the boolean random variables TEST1 and TEST2 are conditionally independent given FLU and the following probabilities: P(FLU) = 0.1, P(TEST1 | FLU) = 0.9, P(TEST1 | ¬ FLU) = 0.2, P(TEST2 | FLU) = 0.8, P(TEST2 | ¬ FLU) = 0.1

What is the probability of FLU is true given TEST1 and TEST2 are both true:
   A. .25
   B. .79
   C. .0002
   D. All of the above.
   E. None of the above.

Decision Network [16%] Consider the following Bayesian network. A, B, C, and D are boolean random variables.



7. [4%] Given that A is true and B is false, what is the probability of D being true?
A. .004
B. .0034
C. .36
D. All of the above.
E. None of the above.

8. [4%] Converting A into a decision node in the above network and given the following utility function: if D is true then utility is 10 and if D is false utility is 100, which decision for A will give the higher expected utility?

A. A is true
B. A is false
C. Utility is equal in both decisions of A.
D. All of the above.
E. None of the above.

9. [4%] In the decision network described in question 28 which of the following events will increase the expected utility the most?

A. B is true
B. B is false
C. C is false
D. All of the above.
E. None of the above.

10. [4%] Variable elimination is an algorithm for:

A. Approximate probabilistic inference in decision networks.
B. Exact probabilistic inference.
C. Computing optimal policies.
D. Reinforcement learning in unknown environments.
E. None of the above.

Decision Tree [16%] NASA wants to train a robot to recognize two species Martians (M) and Humans (H) based on the following characteristics: Green ∈ *N*,*Y*, *Legs* ∈ 2,3, *Height* ∈ *Short*, *Tall*, Smelly ∈ *N*,*Y*. The available training data is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2	S	N
10)	H	N	2	T	Y

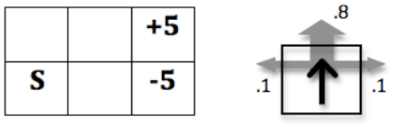
11. [4%] Using the ID3 algorithm to learn a decision tree, the root node chosen would be:
    A. Green
    B. Legs
    C. Both Height and Smelly
    D. All of the above.
    E. None of the above.

12. [4%] Once the learned concept for Martian is an ID3 decision tree, it can be converted into a set of conjunctive rules. For that tree which of following rules would be included:
    A. if (Green=Y and Legs=2 and Height=T and Smelly=N) then M;
    B. if (Green=N and Smelly=Y) then M;
    C. if (Legs=2) then M;
    D. All of the above.
    E. None of the above.

13. [4%] How is Occam's Razor employed in decision tree learning?
    A. ID3 is a greedy approach to construct the shortest tree.
    B. Occam's Razor does not apply to decision tree learning.
    C. Occam's Razor prunes nodes from the tree.
    D. All of the above.
    E. None of the above.

14. [4%] In decision tree learning, the main goal is to:
    A. Minimize the entropy of the training examples associated with each node.
    B. Maximize the entropy of the training examples associated with each node.
    C. Find the best tree structure, such that each training example is a separate tree node.
    D. All of the above.
    E. None of the above.

MDPs and Reinforcement Learning [16%] Given a Gridworld MDP as shown below, the states are grid squares, identified by row then column number. The agent always starts in the state (1,1), marked with the letter S. There are two terminal goal states, (2,3) with reward +5 and (1,3) with reward -5. Rewards are 0 in non-terminal states. The transition function is such that the intended agent movement (North, South, West, or East) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.



(a) (b)

(a) Gridworld MDP. (b) Transition function.

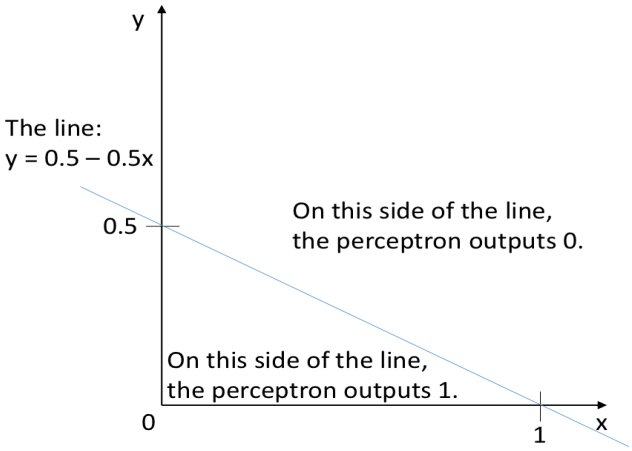
15. [4%] If the agent is deterministic, the optimal policy for this grid is:
    A. State (1,1) North, State (1,2) North, State (2,1) East, State (2,2) East
    B. State (2,1) East, State (2,2) East, State (1,1) North, State (1,2) North
    C. State (1,2) North, State (2,2) East, State (2,1) East, State (1,1) North
    D. All of the above.
    E. None of the above.

16. [4%] Now suppose the agent is stochastic and knows the transition probabilities. Give the second round of value iteration updates for each state, with a discount factor of 0.9 (Assume initial V values are 0 and compute V2 for each non-terminal state).
    A. State (1,1) is 0, State (1,2) is 0, State (2,1) is 0, State (2,2) is 3.924
    B. State (1,1) is 0, State (1,2) is 2.142, State (2,1) is 2.592, State (2,2) 3.924
    C. State (1,1) is 0, State (1,2) is 2.82528, State (2,1) is 2.82528, State (2,2) 3.924
    D. All of the above.
    E. None of the above.

17. [4%] Suppose the stochastic agent does not know the transition probabilities. What does it need in order to learn the optimal policy?
    A. Probability values
    B. Optimal Q values
    C. Training examples of the agent's movement in the grid with rewards received
    D. All of the above.
    E. None of the above.

18. [4%] Reinforcement learning can have problems with scaling because of:
    A. Sparse training data
    B. Calculations taking too much time
    C. It is hard to program
    D. All of the above
    E. None of the above

Neural Networks [16%] Design a perceptron with inputs *x* and *y* and a threshold function as activation function whose output for a given input pair (*x*, *y*) is given by:



19. [4%] The correct weights **w1** for input *x* and **w2** for input *y* and threshold **t** of the perceptron are:
   A. **w1**= -0.5, **w2**= 1, if **t**=> 0.1 outputs 1 else 0
   B. **w1**= 0.5, **w2**= 0.5, if **t**=>1 outputs 1 else 0
   C. **w1**= 0.5, **w2**= 1, if **t** <= .5 outputs 1 else 0
   D. All of the above.
   E. None of the above.

20. [4%] Given input *x* and *y* many different perceptrons can be created whose output is the line above, by making changes to:
   A. Biases
   B. Activation function
   C. Probability distribution
   D. All of the above
   E. None of the above

21. [4%] How many weights does a 2-layer feed-forward neural network with 5 input units, 3 hidden units and 2 output units contain, including the biases (dummy input weights)?
   A. 42
   B. 24
   C. 22
   D. 9
   E. None of the above

22. [4%] A perceptron is guaranteed to learn a given linearly separable function:
   A. When the activation function is a step function.
   B. Within a finite number of training steps.
   C. If all the weights are positive.
   D. All of the above.
   E. None of the above.

General AI Knowledge: True/False [8%] For each of the statements below, answer **TRUE** if the statement is **always true**, or **FALSE** otherwise.

23. [1%] **A. TRUE** **B. FALSE** Cross validation helps to avoid overfitting.

24. [1%] **A. TRUE** **B. FALSE** Unsupervised learning is learning from unlabeled data to maximize a cumulative reward.

25. [1%] **A. TRUE** **B. FALSE** Support vectors are data points that lie closest to the hyperplane.

26. [1%] **A. TRUE** **B. FALSE** A decision tree cannot represent all Boolean function.

27. [1%] **A. TRUE** **B. FALSE** Q-learning creates an explicit transition model.

28. [1%] **A. TRUE** **B. FALSE** A node X is conditionally independent of all other nodes in a Bayesian Network given its Markov Blanket.

29. [1%] **A. TRUE** **B. FALSE** A policy is a mapping from states to actions that tells the agent what action to take in any possible environment state.

30. [1%] **A. TRUE** **B. FALSE** A Bayesian Network modeling n random variables that are all mutually independent of each other would have directed arcs between every pair of variables in the network.