Total 110 points. 1-11 each 3 point: True or False, give brief explanation (one or two sentence).

1. The discounted model of optimality in MDPs is inappropriate for agents that have a known finite lifetime.

True

2. A reinforcement learning agent, having estimated a utility function, should simply adopt the policy of choosing the action that maximize the expected estimated utility.

3. In reinforcement learning, the agent needs to know the reward function in advance.

False

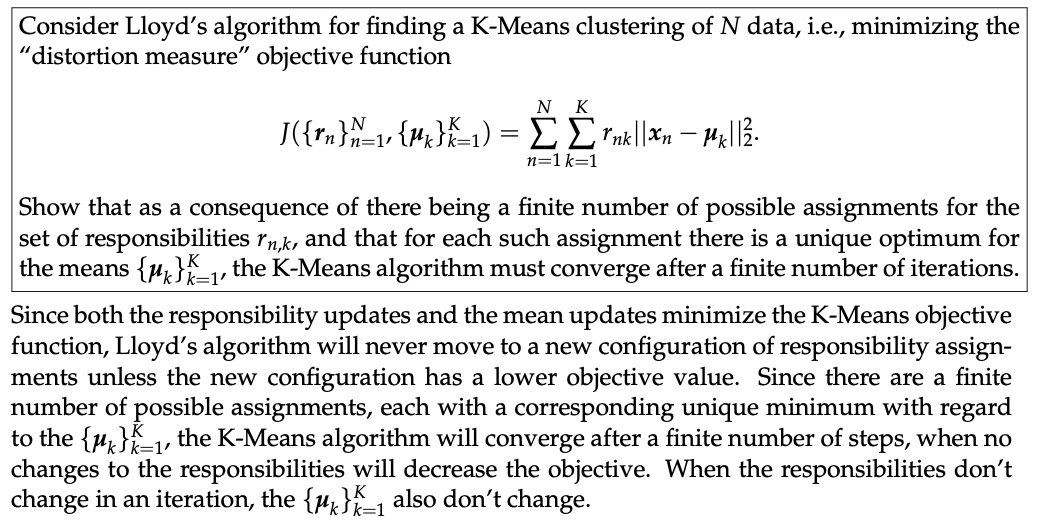
4. Under the undiscounted criterion(γ=1), a Markov decision process can have unbounded long-term reward.

5. Following the strategy prescribed by a Nash equilibrium is an optimal strategy for a group of rational agent.

6. If implemented properly, K-means must converge after a finite number of iterations.

7.If implemented properly, EM must converge after a finite number of iterations.

True



8. Consider the following feature selection algorithm (called The Uninformative Feature Function, or TUFF): remove features in your data which have entropy below some suitably low threshold and keep the rest. TUFF is an effective feature selection algorithm for the task of classification because it throws away the least informative features.

True. TUFF can be considered as a filter-style feature selection algorithm that utilizes entropy. The good side of it is, it decides which feature to keep (or remove) purely on the entropy of each feature, so the speed is quick. However, such algorithm could be not very powerful compared with a wrapper feature selection algorithm.

9. Scale invariance means the clustering algorithm produces the same clusters for a distance matrix and a positive multiple of the same distance matrix. Scale invariance can be archived for a single-linked clustering when it terminates as the number of clusters equals to the half of the number of objects.

True. In single-linkage clustering, the distance among points decides whether any given pair of points should be connected into the same cluster. Multiplying a positive value to the distance matrix will not affect the relative distance among all the points.

10. The no free lunch theorem means it doesn't matter what algorithm is chosen for any particular problem.

False, the no free lunch theorem indicates that for all possible problems, these is no such algorithm that outperforms other algorithms. However, a particular problem could belong to a particular subset of all problems, and some algorithms WILL outperform other algorithms in solving this particular subset of problems.

False, it will only work for optimization and search problem.

11. Both supervised learning algorithms (e.g. neural networks) and reinforcement learning algorithms (e.g. TD) are trying to solve credit assignment problems.

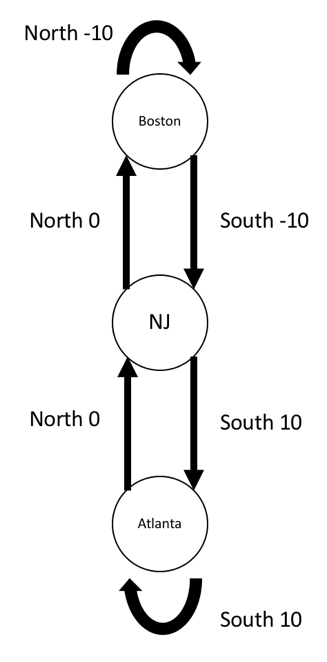
True. In supervised learning, the algorithm is trying to distinguish which features, and what value to that feature, are more valuable (or, worth more credit) than others. In reinforcement learning, the algorithm will evaluate (i.e. assign values to) all actions from all states.

12. (12 points) Compare principal components analysis and independent components analysis. For example: what do they do? For what sort of problems is each suited? What are the strengths and weaknesses of each? Be specific. Compare both to any of your favorite clustering algorithms. How is clustering similar and different to dimensionality reduction?

13. (15 points) Deterministic Markov Decision Processes

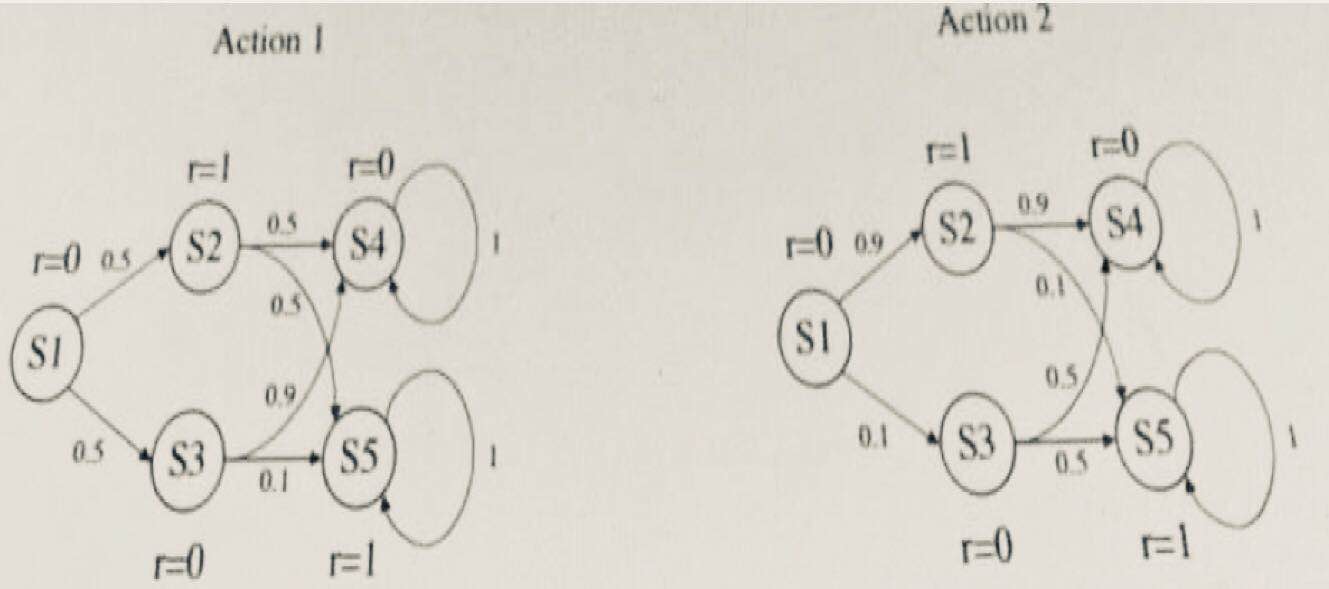
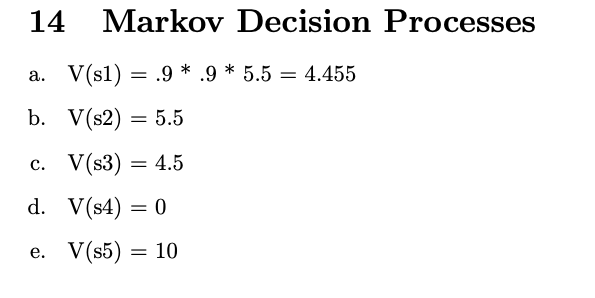
Consider the remarkably simple three-state deterministic world described by the graph on the right. There are two actions for each state: move north or south. Rewards are marked on edges.

1. How many possible policies are there in this world?
2. Now using actual math to justify your answer, explain whether the policy move north for all states is better or worse than the policy move south for all states, assuming that γ = 0.5.
3. Would a different value of γ < 1 change your answer? How so? Again, justify your answer.



14. (16 points) Markov Decision Processes

Consider the single MDP at each state you can take one of two actions. Just for ease of reading we’ve drawn the single MDP as two graphs. One with the transition probabilities for action 1 and the other with the transition probabilities for action 2. The transition probabilities are on the transition arcs between states and the reward for entering each state is written above or below the state as r=#. What are the utilities of each of the states, assuming γ=0.9?

15. (16 points) Consider the game described on the right. Players A and B have strategies a1, a2, b1, and b2. The payoffs (indicated by A, B in the cells) are known to everyone in advance.

(a) For x=2: What are the dominated strategies? What are the pure-strategy Nash equilibria? What are the mixed-strategy Nash equilibria?

(b) For x=6: What are the dominated strategies? What are the pure strategy Nash equilibria? What are the mixed-strategy Nash equilibria?

|  |  |  |
| --- | --- | --- |
|  | b1 | b2 |
| a1 | 6, 4 | 5, x |
| a2 | 0, 8 | 9, 0 |

16. (18 points) You are asked to work during summer time for a communication giant company. The CEO has a virtual agent with virtual actuators and virtual sensors, but he cannot explain to you how it works. However, he knowns when the agent is doing well or poorly.

(a) Design a reinforcement learning method to solve this problem. Be specific on you design. How do you obtain the parameters and how do you tell your system is working well? Explain why you cannot use supervised learning method to solve this problem.

(b) Can you use unsupervised learning method to solve this problem? Which algorithm will you use?

(c) If you start your work by the last month of summer, do you think your system will work well? You can also use your personal experience to address this.