1 First Order Logic

1. Vocab check: are you familiar with the following terms?
   (a) Objects
   (b) Relations
   (c) Functions
   (d) Constants
   (e) Variables
   (f) Interpretation
   (g) Connectives
   (h) Equality
   (i) Quantifiers
   (j) Atomic Sentence
   (k) Unification

2. Which of the following FOL sentences correctly expresses its corresponding English sentence?
   (a) There was a student at CMU who never did 281 homework but passed the class.
   $\exists x, \text{IsStudent}(x, \text{CMU}) \wedge \neg\text{DoesHW}(x, 281) \implies \text{Pass}(x, 281)$

   (b) If a student likes Pat, they'll pass the class.
   $\forall x, \text{Student}(x) \wedge \text{Likes}(x, \text{Pat}) \implies \text{Pass}(x, 281)$

   (c) All students at CMU who never did 281 homework passed the class.
   $\forall x, \text{IsStudent}(x, \text{CMU}) \wedge \neg\text{DoesHW}(x, 281) \wedge \text{Pass}(x, 281)$

3. Which of the following is true with respect to the English sentences (a) and (c) above?

   (a) $\models$ (c) (c) $\models$ (a) Both Neither

4. Forward Chaining
   (a) What are the requirements for Knowledge Base in Forward Chaining?

   (b) What does FOL Forward Chaining return?

5. DPLL: determine whether the following are satisfiable or unsatisfiable using DPLL.
   (a) $(P \vee Q \vee \neg R) \wedge (P \vee \neg Q) \wedge \neg P \wedge R \wedge U$

(b) $(P \vee Q) \wedge \neg Q \wedge (\neg P \vee Q \vee \neg R)$

2 Planning

1. Vocab check: are you familiar with the following terms?

   (a) Predicates
   (b) Operators
   (c) Linear Planning
   (d) Non-linear Planning
   (e) Inconsistency
   (f) Interference
   (g) Competing Needs
   (h) Complete
   (i) Sound
   (j) Optimal

2. What are in the knowledge base of logic agents and classical planning problem, respectively?

3. What are the 3 components when defining an operator?

4. Is linear planning complete? Optimal? What about non-linear planning?

5. The monkey-and-bananas problem is faced by a monkey in a laboratory with some bananas hanging out of reach from the ceiling. A box is available that will enable the monkey to reach the bananas if he climbs on it. Initially, the monkey is at A, the bananas at B, and the box at C. The monkey and box have height *Low*, but if the monkey climbs onto the box he will have height *High*, the same as the bananas. The actions available to the monkey include *Go* from one place to another, *Push* an object from one place to another, *ClimbUp* onto or *ClimbDown* from an object, and *Grasp* or *Ungrasp* an object. The result of a *Grasp* is that the monkey holds the object if the monkey and object are in the same place at the same height. We want to formulate this problem using GraphPlan.

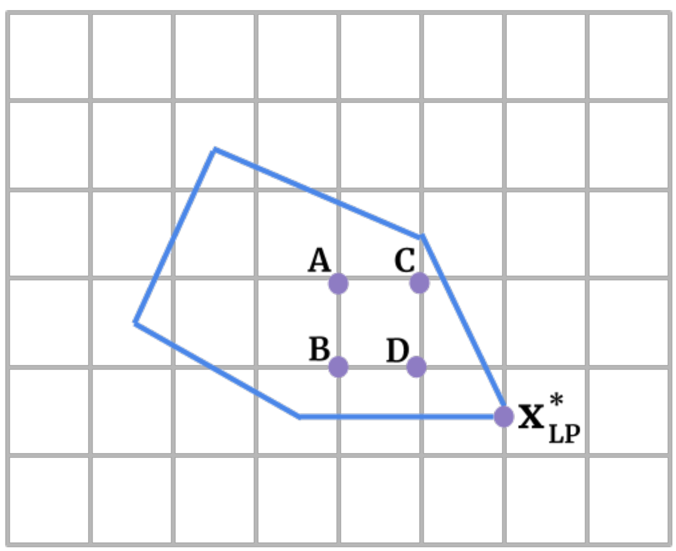
   (a) Define the initial and goal states.

   (b) Write the six action schemas, i.e., define the preconditions and effects for each possible action the monkey can take.

3 IP

1. What is the relationship between branch and bound and pruning?

2. Consider the integer programming problem illustrated by the figure below.
   - The blue lines represent the boundaries of the feasible region and the gray vertical and horizontal lines represent the integer values for each axis.
   - Let x^*_{LP} be the unique point that minimizes the relaxed linear program. It happens to lie on a vertical gray line and a horizontal blue line.
   - Let x^*_{IP} (unlabeled) be the unique point that minimizes the integer program.
   - When running branch and bound, we will explore the x_i less than constraint subtree before the x_i greater than constraint subtree ("less" being to the left and down).



   (a) Which of the points (A, B, C, or D) can possibly be x^*_{IP}?

   (b) Which of the points from your answer to part (a) needs to be x^*_{IP} in order for branch and bound to find the IP solution at the minimum possible depth?

4 MDPs

1. Concepts:

   (a) What does the Markov Property state?

   (b) What are the Bellman Equations, and when are they used?

   (c) What is a policy? What is an optimal policy?

   (d) How does the discount factor γ affect the agent's policy search? Why is it important?

   (e) What are the two steps to Policy Iteration?

   (f) What is the relationship between $V^*(s)$ and $Q^*(s, a)$?

2. In a certain country there are N cities, all connected by roads in a circular fashion. A wandering poet is travelling around the country and staging shows in its different cities. She can choose to move from a city to each of the neighboring ones or she can stay in her current city i and perform, getting a reward r_i. If she chooses to travel, she will have a success probability of p_i. There is a $1 - p_i$ chance she will encounter a dragon along the way, which means she will have to turn back and wait the next day. If she is successful in travelling, she gains a reward of 0 for the day. And if she is unsuccessful at travelling, she can still perform a little bit when she gets back, giving her a reward of $r_i/2$.

   Let $r_i = 1$ and $p_i = 0.5$ for all i and let $\gamma = 0.5$. For $1 \leq i \leq N$, answer the following questions with concrete numbers:

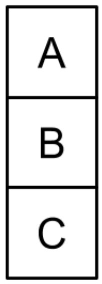
   i) What is the value $V^{\text{stay}}(i)$ under the policy the wandering poet always chooses to stay?

   ii) What is the value $V^{\text{next}}(i)$ under the policy the wandering poet always chooses to go to the next city? You may assume that the values of each state converge to the same value.

5 Reinforcement Learning: Just Add Water

Rikki the mermaid is trying to learn a general policy for swimming to Mako Island while avoiding Charlotte[1], the inferior and annoying mermaid. Rikki always starts in grid A shown below.

Rikki's possible environment states are of the form (x, y), where $x \in \{A, B, C\}$, representing her grid position, and y is 1 if Charlotte is present in grid C, 0 if she is nowhere on the map. Rikki's possible actions are swim (S) or heat (H). She observes the episodes shown to the right:



s	a	s'	r
(A, 1)	S	(C, 1)	-4
(A, 1)	S	(B, 1)	0
(B, 1)	S	(C, 1)	-4
(A, 0)	S	(C, 0)	2
(A, 0)	S	(B, 0)	0
(B, 0)	S	(C, 0)	2
(A, 1)	S	(B, 1)	0
(B, 1)	H	(B, 0)	0
(B, 0)	S	(C, 0)	2

1. Which of the following are techniques Rikki can use to try finding an optimal policy (without having to use any other process in addition to the technique)?
   - Policy iteration
   - TD learning
   - Estimate T and R, then policy iteration
   - Q-learning
   - Value iteration, then policy extraction
   - Approximate Q-learning

2. Let $Q(s, a) = 0$ initially for all Q-states (s, a), and that $(C, 0)$ and $(C, 1)$ are terminal states. Let $\gamma = \alpha = 0.5$. Now calculate $Q(s, a)$ for the entries in the following table after Q-learning.

s	(A, 1)	(A, 0)	(B, 1)	(B, 0)	(A, 1)	(B, 1)
a	S	S	S	S	H	H
$Q(s, a)$						

3. According to the observed episodes, what is Rikki's optimal policy?

4. Suppose Rikki uses ϵ-greedy search, and that her environment is deterministic. She starts with $\epsilon = 0$ and doesn't change it. Is she guaranteed to learn the optimal policy? Why or why not?

[1] who sucks

5. Now suppose Rikki sets $\epsilon = 1$ and doesn't change it. Is she guaranteed to learn the optimal policy? Why/why not?

6. Rikki's no ordinary girl, so she decides to use approximate Q-learning, with $\alpha = \gamma = 0.5$. Suppose she has three features f_1, f_2, f_3, and corresponding weights w_1, w_2, w_3.
   Her approximate function is

$$Q_w(s, a) = w_1^2 f_1(s, a) + e^{w_2 f_2(s,a)} + w_3^3 f_3(s, a).$$

   Initially, $w_1 = 2, w_2 = 0, w_3 = 3$. Suppose $f_1((A, 1), S) = f_2((A, 1), S) = f_3((A, 1), S) = 1$.

   (a) What is the approximate Q-value for $((A, 1), S)$?

   (b) What is the value of $((A, 1), S)$ according to Rikki's first observation $((A, 1), S, (C, 1), -4)$? Suppose $max_{a'} Q_w((C, 1), a') = 4$.

   (c) What are the new weights after the first observation?

6 Probability

Recall the example of *marginalization*, which means summing out variables from a joint distribution. Consider three random variables A, B, and C with domains $\{a_1, a_2\}$, $\{b_1\}$, and $\{c_1, c_2, c_3\}$, respectively. Remember that $P(C)$ refers to the table of probabilities of all the elements of the domain C.

1. Express $P(C)$ in terms of the joint distribution $P(A, B, C)$. Your answers should contain summations.

2. Express $P(C)$ in terms of $P(B)$, $P(A \mid B)$ and $P(C \mid A, B)$. Reorder any summations in your solution to minimize the number of arithmetic operations.

3. Expand the sums from part (b) to show the two elements of $P(C)$ ($P(c_1)$,$P(c_2)$,and $P(c_3)$) in terms of the individual probabilities (e.g. $P(a_2)$, $P(b_1)$ instead of $P(A)$ or $P(B)$).

4. Each of the following terms represents a conditional probability table. What do the entries for each table sum to? (For example, the entries of the table $P(A)$ sum to 1.)

   (a) $P(A \mid B)$

   (b) $P(A \mid C)$

   (c) $P(C \mid a_1, b_1)$

   (d) $P(C \mid a_1, B)$

   (e) $P(C \mid A, b_1)$