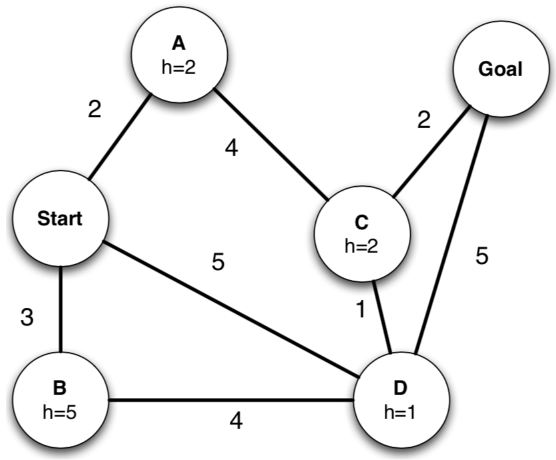
1 Search



For each of the following graph search strategies, work out the order in which states are expanded, as well as the path returned by graph search. In all cases, break ties in alphabetical order. The start and goal state use letter S and G, respectively. Remember that in graph search, a state is expanded only once.

(a) Depth-first search.

(b) Breadth-first search.

(c) Uniform cost search.

(d) Greedy search with the heuristic h shown on the graph.

(e) $A^{\star}$ search with the same heuristic.

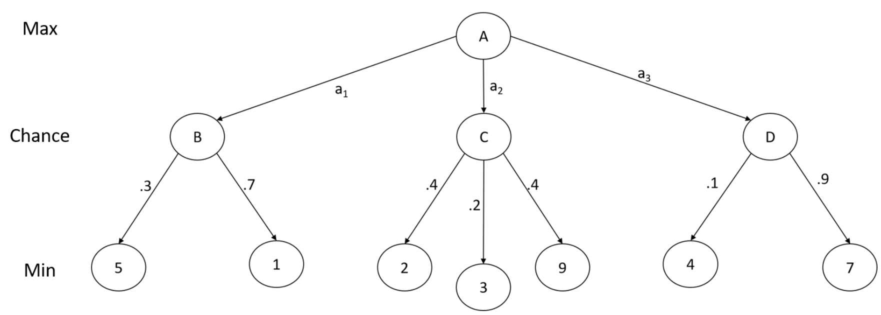
2 Adversarial Search

Warm up

1. What is the advantage of adding alpha-beta pruning to a minimax algorithm

2. Give two advantages of Iterative Deepening minimax algorithms over Depth Limited minimax algorithms.

The three questions are about the following adversarial "chance" tree.



Expectiminimax

1. Calculate the EXPECTIMINIMAX values for nodes B, C and D in the above adversarial "chance"tree. Show your calculations!

2. Which action will MAX choose, a_1, a_2, or a_3? Explain your answer!

3. If the utility values given for MIN were multiplied with a positive constant c, which action would MAX then choose?

3 CSP Backtracking Search

In this problem, you are given a 3×3 grid with some numbers filled in. The squares can only be filled with the numbers $\{2, 3, \ldots, 10\}$, with each number being used once and only once. The grid must be filled such that adjacent squares (horizontally and vertically adjacent, but not diagonally) are relatively prime.

x_1	x_2	x_3
x_4	x_5	3
4	x_6	2

We will use backtracking search to solve the CSP with the following heuristics:

- Use the Minimal Remaining Values (MRV) heuristic when choosing which variable to assign next.
- Break ties with the Most Constraining Variable (MCV) heuristic.
- If there are still ties, break ties between variables x_i, x_j with $i < j$ by choosing x_i.
- Once a variable is chosen, assign the minimal value from the set of feasible values.
- For any variable x_i, a value v is infeasible if and only if: (i) v already appears elsewhere in the grid, or (ii) a variable in a neighboring square to x_i has been assigned a value u where $\gcd(v, u) > 1$, which is to say, they are not relatively prime.

Fill out the table below with the appropriate values.

- Give initial feasible values in set form; x_1 has already been filled out for you.
- Assignment order refers to the order in which the final value assignments are given. If x_i is the j^{th} variable on the path to the goal state, then the assignment order for x_i is j.
- In the branching column, write "yes" if the algorithm branches (considers more than one value) at that node in the search tree, and write "B" if the algorithm backtracks at that node, meaning it is the highest node in its subtree that fails for a value, and has to be chosen again. Also write the values it tried then failed.

Variable	Initial Feasible Values	Assignment Order	Final Value	Branch or Backtrack?
x_1	$\{5, 6, 7, 8, 9, 10\}$	________	______	__________
x_2	__________	________	______	__________
x_3	__________	________	______	__________
x_4	__________	________	______	__________
x_5	__________	________	______	__________
x_6	__________	________	______	__________

4 Local Search

(a) Which of the following local search algorithm are complete and/or optimal? If necessary, specify the conditions that must be true for completeness or optimality.

- First-choice Hill Climbing
  (i) Complete?
  (ii) Optimal?
- Random-restart Hill Climbing
  (i) Complete?
  (ii) Optimal?
- Simulated Annealing
  (i) Complete?
  (ii) Optimal?
- Genetic Algorithm
  (i) Complete?
  (ii) Optimal?
- Local Beam Search
  (i) Complete?
  (ii) Optimal?

(b) Of the local search algorithms above, which one(s) would perform best in a continuous state space and why?

(c) What are the disadvantages and advantages of allowing sideways moves? How can we modify our search algorithm to address the disadvantages?

5 Propositional Logic

1. Warm Up: Are you familiar with these terms?

    - Symbols
    - Operators
    - Sentences
    - Equivalence
    - Literals
    - Knowledge Base
    - Entailment
    - Query
    - Satisfiable
    - Valid
    - Clause - Definite, Horn clauses
    - Model Checking
    - Theorem Proving
    - Modus Ponens

2. Indicate whether the following sentence is *valid*, *satisfiable*, or *unsatisfiable*. If satisfiable, give a model such that the sentence is satisfied. Prove your answer by reducing the sentence to its simplest form. Remember to **show all the steps and write down an explanation of each step.** Let T stand for the atomic sentence *True* and F for the atomic sentence *False*.

    $((T \Leftrightarrow \neg(x \vee \neg x)) \vee z) \wedge \neg(z \wedge ((z \wedge \neg z) \Rightarrow x))$

6 Satisfiability and Planning

With the holidays coming up, Santa needs to start making a plan (and checking it twice) to deliver presents. First he tries taking a SATplan (logical planning) approach, and formulates the following propositions:

- $at(loc, t)$: Santa's sled is at location loc at time t
- $reindeerHunger(x, t)$: the reindeers' hunger level is at x at time t, $x \in [0, 5]$.
- $hasCarrots(loc, t)$: location loc has carrots to feed reindeer with at time t
- $hasPresents(loc, t)$: location loc has presents at time t

His starting state is $at(NorthPole, 0) \wedge reindeerHunger(0, 0)$.

1. Using the above predicates, formulate successor-state axioms for the actions $feedReindeer(t)$, $deliver(t)$, and $fly(origin, destination, t)$. Santa can only feed the reindeer at a location that has carrots, and their hunger level drops to 0 as a result of feeding. Santa can fly between any distinct locations, as long as the reindeers' hunger level is less than 3; after flying, their hunger level increases by 1. Santa can deliver presents anywhere, and the result is that the location of his sled now has presents.

2. Convert your $deliver$ axiom into conjunctive normal form. You may want to abbreviate each proposition. What's the purpose of converting logical sentences into CNF (besides solving recitation problems)?

3. Suppose Santa's goal is to deliver presents to NorthPole. Describe an algorithm which uses a SAT solver to find a plan for this goal.

4. Run DPLL to determine whether the goal $hasPresents(NorthPole, 1)$ is feasible with our knowledge base $at(NorthPole, 0) \wedge reindeerHunger(0, 0) \wedge D$, where D is your $deliver$ axiom instantiated with $loc = NorthPole$, $t = 0$ (we can leave the other axioms out because they won't be relevant). What's a possible model found by DPLL?

5. Suppose DPLL returned $False$ on some sentence $A \wedge B$. What entailment conclusions can we draw involving A and B?

6. Now Santa tries taking a GraphPlan approach. Define each action as an operator in the following table (note that we can drop the t parameter from each predicate and action):

	$feedReindeer$	$fly(o, d)$	$deliver$
Precondition			
Add			
Delete			

7. Which of the operators are mutually exclusive? What type of mutex relation does each pair have?

8. Now draw the GraphPlan graph up to proposition level S_1. Suppose Pittsburgh is the only other location besides NorthPole.

9. Which operators are mutually exclusive in A_0? Which propositions are mutually exclusive in S_1?

10. In general, when does GraphPlan stop extending the planning graph?

11. Is GraphPlan sound? complete? optimal? What about the SATPlan algorithm you described above?

7 First Order Logic

(a) For each of the logical expressions, state whether it correctly expresses the English sentence and explain.

i. Everyone in 281 is awesome: $\forall x$ In(x, 281) $\wedge$ Awesome(x)

ii. Someone taking 281 actually dislikes the class: $\exists x$ In(x, 281) => Dislikes(x, 281).

(b) For each pair of atomic sentences, give the most general unifier if it exists:

i. ***Q(f(a), f(b)), Q(f(x),f(x))***

ii. ***R(a, a, b, f(b)), R(f(x), f(m(5)), x, y)***

8 Bayes' Nets: Representation, Independence

For this problem, any answers that require division can be left written as a fraction.
PacLabs has just created a new type of mini power pellet that is small enough for Pacman to carry around with him when he's running around mazes. Unfortunately, these mini-pellets don't guarantee that Pacman will win all his fights with ghosts, and they look just like the regular dots Pacman carried around to snack on.
Pacman (P) just ate a snack , which was either a mini-pellet ($+p$), or a regular dot ($-p$), and is about to get into a fight (W), which he can win ($+w$) or lose ($-w$). Both these variables are unknown, but fortunately, Pacman is a master of probability. He knows that his bag of snacks has 5 mini-pellets and 15 regular dots. He also knows that if he ate a mini-pellet, he has a 70% chance of winning, but if he ate a regular dot, he only has a 20% chance.

(a) What is $P(+w)$, the marginal probability that Pacman will win?

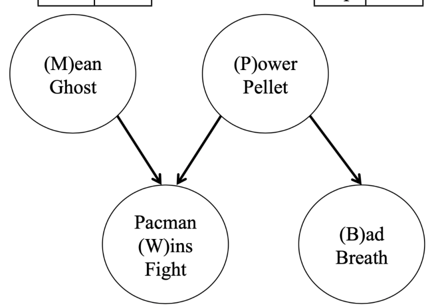
(b) Pacman won! Hooray! What is the conditional probability $P(+p \mid +w)$ that the food he ate was a mini-pellet, given that he won?

Pacman can make better probability estimates if he takes more information into account. First, Pacman's breath, B, can be bad ($+b$) or fresh ($-b$). Second, there are two types of ghost (M): mean ($+m$) and nice ($-m$). Pacman has encoded his knowledge about the situation in the following Bayes' Net:

P(M)	
+m	0.50
-m	0.50

P(P)	
+p	0.25
-p	0.75

P(W|M, P)			
+m	+p	+w	0.60
+m	+p	-w	0.40
+m	-p	+w	0.10
+m	-p	-w	0.90
-m	+p	+w	0.80
-m	+p	-w	0.20
-m	-p	+w	0.30
-m	-p	-w	0.70



P(B|P)		
+p	+b	0.80
+p	-b	0.20
-p	+b	0.40
-p	-b	0.60

(c) What is the probability of the event $(-m, +p, +w, -b)$, where Pacman eats a mini-pellet and has fresh breath before winning a fight against a nice ghost?

(d) Which of the following conditional independence statements are guaranteed to be true by the Bayes' Net graph structure?

i. $W \perp\!\!\!\perp B$

ii. $W \perp\!\!\!\perp B|P$

iii. $M \perp\!\!\!\perp P$

iv. $M \perp\!\!\!\perp P|W$

v. $M \perp\!\!\!\perp B$

vi. $M \perp\!\!\!\perp B|P$

vii. $M \perp\!\!\!\perp B|W$

For the remaining of this question, use the half of the joint probability table that has been computed for you below:

$P(M, P, W, B)$				
$+m$	$+p$	$+w$	$+b$	0.0800
$+m$	$+p$	$+w$	$-b$	0.0150
$+m$	$+p$	$-w$	$+b$	0.0400
$+m$	$+p$	$-w$	$-b$	0.0100
$+m$	$-p$	$+w$	$+b$	0.0150
$+m$	$-p$	$+w$	$-b$	0.0225
$+m$	$-p$	$-w$	$+b$	0.1350
$+m$	$-p$	$-w$	$-b$	0.2025

(e) What is the marginal probability, $P(+m, +b)$ that Pacman encounters a mean ghost and has bad breath?

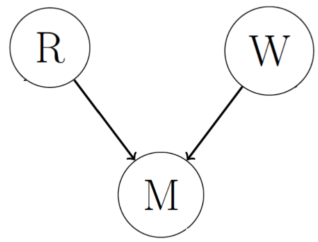
(f) Pacman observes that he has bad breath and that the ghost he's facing is mean. What is the conditional probability, $P(+w \mid +m, +b)$, that he will win the fight, given his observations?

(g) Pacman's utility is +10 for winning a fight, -5 for losing a fight, and -1 for running away from a fight. Pacman wants to maximize his expected utility. Given that he has bad breath and is facing a mean ghost, should he stay and fight, or run away? Justify your answer.

9 Bayes' Nets: Sampling

Consider the following Bayes Net and corresponding probability tables.

R	P(R)
+r	0.4
-r	0.6



W	P(W)
+w	0.2
-w	0.8

M	R	W	P(M | R,W)
+m	+r	+w	0.1
-m	+r	+w	0.9
+m	+r	-w	0.45
-m	+r	-w	0.55
+m	-r	+w	0.35
-m	-r	+w	0.65
+m	-r	-w	0.9
-m	-r	-w	0.1

Consider the case where we are sampling to approximate the query $P(R \mid +m)$.

Fill in the following table with the probabilities of drawing each respective sample given that we are using each of the following sampling techniques. Let $P(+m) = a$.

Method	$< +r, -w, +m >$	$< +r, +w, -m >$
Prior sampling		
Rejection sampling		
Likelihood weighting		

We are going to use Gibbs sampling to estimate the probability of getting the sample $< +r, -w, +m >$. We will start from the sample $< -r, +w, +m >$ and resample W first then R. What is the probability of drawing sample $< +r, -w, +m >$?

10 HMMs and Particle Filtering

Consider the following Markov Model with a binary state X (i.e. X_t is either 0 or 1). The transition probabilities and initial distribution are as follows:

X_0	$P(X_0)$
0	0.5
1	0.5

X_t	X_{t+1}	$P(X_{t+1}\|X_t)$
0	0	0.9
0	1	0.1
1	0	0.5
1	1	0.5

(a) After one timestep, what is the new belief distribution $P(X_1)$?

X_1	$P(X_1)$
0	
1	

Now, we incorporate sensor readings as our observations. The sensor model is parameterized by some value $\beta \in [0, 1]$:

X_t	E_t	$P(E_t\|X_t)$
0	0	β
0	1	$1 - \beta$
1	0	$1 - \beta$
1	1	β

(b) At $t = 1$, we get the first sensor reading, $E_1 = 0$. Find $P(X_1 = 0|E_1 = 0)$ in terms of β.

(c) For what range of values of β will a sensor reading $E_1 = 0$ increase our belief that $X_1 = 0$? In other words, what is the range of β for which $P(X_1 = 0|E_1 = 0) > P(X_1 = 0)$?

(d) Now, we want to use particle filtering to predict what state our model. At time t, there are 2 particles in state value 0, and 3 particles in state value 1. What is the prior belief distribution $\hat{P}(X_t)$?

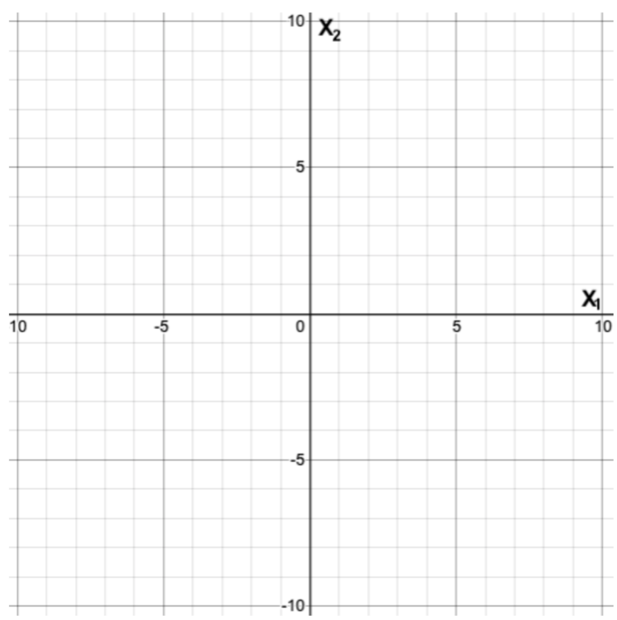
(e) At time t, we receive our first sensor reading $E_t = 1$. Given $\beta = 0.6$ and the previous table for $P(E_t|X_t)$, how many particles will be in each state value after resampling? For a source of randomness, use this list of numbers: [0.182, 0.703, 0.471, 0.859, 0.382]

11 LP

Santa is struggling to pack all the presents that he wants to deliver onto his sled. This year, Santa is giving out massive amounts of chocolate and peppermints.

Let x_1 represent pounds of chocolate and let x_2 represent pounds of peppermints. We assume we can deliver a fraction of a pound of chocolate or peppermints. Unfortunately, Santa's sled can only fit 10 pounds of sweets. Furthermore, Santa wants to provide enough presents for 10 kids. Each pound of chocolate is enough for 2 kids while a pound of peppermint is only enough for 1 kid. However, Santa also wants to maximize the children's happiness. Chocolate brings 4 units of happiness while peppermints only bring 1.

1. Represent the following problem as an LP and graph the constraints in the provided graph.



2. What would the optimal solution be?

3. If Santa didn't know how much happiness chocolate and peppermints bring, what would be a cost vector that makes the optimal solution (0, 10)?

4. List three cost vectors that will lead to an infinite number of solutions.

12 MDPs/RL

1. Warm Up

   - What does the Markov Property state?

   - What are the Bellman Equations, and when are they used?

   - What is a policy? What is an optimal policy?

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

   - What are the two steps to Policy Iteration?

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

   - Exploration, exploitation, and the difference between them? Why are they both useful?

   - What is the difference between on-policy and off-policy learning?

   - What is the difference between model-based and model-free learning?

   - We are given a pre-existing table of Q-values (and its corresponding policy), and asked to perform ϵ-greedy Q-learning. Individually, what effect does setting each of the following constants to 0 have on this process?

     (i) α:

     (iii) ϵ:

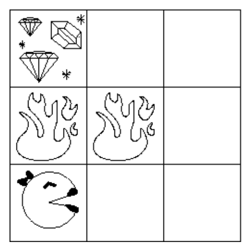
   - For each of the following functions, write which MDP/RL value the function computes, or none if none apply. We are given an MDP (S, A, T, γ, R), where R is only a function of the current state s. We are also given an arbitrary policy π.

     Possible choices: $V^*, Q^*, \pi^*, V^\pi, Q^\pi$.

(i) $f(s) = R(s) + \sum_{s'} \gamma T(s, \pi(s), s') f(s')$

(ii) $g(s) = \max_{a} \sum_{s'} T(s, a, s')[R(s) + \gamma \max_{a'} Q^*(s', a')]$

2. Ms.Pacman: While Pacman is busting ghosts, Ms. Pacman goes treasure hunting on GridWorld Island. She has a map showing where the hazards are, and where the treasure is. From any unmarked square, Ms. Pacman can take any of the deterministic actions (N, S, E, W) that doesn't lead off the island. If she lands in a hazard square or a treasure square, her only action is to call for an airlift (X), which takes her to the terminal *Done* state; this results in a reward of -64 if she's escaping a hazard, or +128 if she reached the treasure. There is no living reward.



(a) Let $\gamma = 0.5$. What are the optimal values V^* of each state in the grid above?

(b) How would we compute the Q-values for each state-action pair?

(c) What's the optimal policy?

Call this policy π_0.

Ms. Pacman realizes that her map might be out of date, so she uses Q-learning to see what the island is really like. She believes π_0 is close to correct, so she follows an ϵ-random policy, ie., with probability ϵ she picks a legal action uniformly at random (otherwise, she does what π_0 recommends). Call this policy π_ϵ.

π_ϵ is known as a *stochastic* policy, which assigns probabilities to actions rather than recommending a single one. A stochastic policy can be defined with $\pi(s, a)$, the probability of taking action a when the agent is in state s.

(d) Write a modified Bellman update equation for policy evaluation when using a stochastic policy $\pi(s, a)$ (this is similar to a problem seen on midterm 2).

3. Consider the fictitious play matching pennies example from lecture. A (rows) wants to match and B (columns) does not.

A/B	Heads	Tails
Heads	1,-1	-1,1
Tails	-1,1	1,-1

We will play a few more rounds to understand how fictitious play can arrive at a Nash equilibrium. Remember that w(a) refers to the number of times the opponent plays action a.

Round	A action	B action	A's belief for w_B	B's belief for w_A
0			(1,1)	(1,1)
1	H	H	(2,1)	(2,1)
2				
3				
4				

(a) Play round 2. What action will A choose? What action will B choose? Update the beliefs accordingly.

(b) Fill out the table completely by simulating all the rounds. Assume that both players break ties by choosing heads. What are the final beliefs? Do you notice interesting changes in the players' strategies? What do you think the equilibrium state is?

(c) Will this game converge? What do you think is the Nash equilibrium?

13 Game Theory

1. Warm Up
   (a) What is a strategy?
   (b) What is the difference between a pure and mixed strategy?
   (c) What is a Nash Equilibrium?
   (d) Does a Nash Equilibrium always exist?
2. Nash Equilibria
   With the Grinch now reformed, he has started helping Santa deliver presents. They have different preferred towns to deliver to but they are both happier if they are delivering presents together. The following table denotes the expected payoffs for Santa and the Grinch for the two different directions they can travel in.

		The Grinch	
		North	South
SANTA	North	(8, 8)	(1, 1)
	South	(3, 4)	(2, 9)

   (a) Identify the pure strategy Nash Equilibrium in this game.
   (b) Determine the mixed strategy Nash Equilibrium in this game.

3. Voting Rules

(a) Match each voting rule, axiom, or property to its description.

(i) ______ For $m-1$ rounds, each voter gets 1 vote per round and alternative with least plurality votes per round is eliminated. Alternative left is the winner.

(ii) ______ Each voter give one vote to top ranked preference, alternative with most votes wins.

(iii) ______ Voter can never benefit from lying about preferences.

(iv) ______ Alternative that beats every other alternative in pairwise election.

(v) ______ Alternative x beats y if majority of voters prefer x to y.

(vi) ______ First round: top two plurality winners advance to second round. Second round: pairwise election between the two.

(vii) ______ Each voter awards $m - k$ votes to their rank k, alternative with most votes wins.

i. Plurality
ii. Borda Count
iii. Plurality with runoff
iv. Single Transferable Vote
v. Pairwise election
vi. Condorcet winner
vii. Strategyproof

(b) Which voting rule(s) (Plurality, Borda Count, Plurality with Runoff, Single Transferable Vote) are **Condorcet consistent**? It may help to consider the following two examples.

3 voters	2 voters
A	B
B	C
C	A

3 voters	2 voters	2 voters
A	B	C
B	C	B
C	A	A

(c) Which voting rule(s) (Plurality, Borda Count, Plurality with Runoff, Single Transferable Vote) are **strategyproof**?