CS188 Fall 2018 Section 4: Games and MDPs

1 Utilities

1. Consider a utility function of $U(x) = 2x$. What is the utility for each of the following outcomes?

   (a) 3
   U(3) = 2(3) = 6

   (b) L($\frac{2}{3}$, 3; $\frac{1}{3}$, 6)
   U(L($\frac{2}{3}$, 3; $\frac{1}{3}$, 6)) = $\frac{2}{3}$U(3) + $\frac{1}{3}$U(6) = 8

   (c) -2
   U(-2) = 2(-2) = -4

   (d) L(0.5, 2; 0.5, L(0.5, 4; 0.5, 6)) U(L(0.5, 2; 0.5, L(0.5, 4; 0.5, 6))) = 0.5U(2) + 0.5(0.5U(4) + 0.5U(6)) = 2 + 0.5(4 + 6) = 7

2. Consider a utility function of $U(x) = x^2$. What is the utility for each of the following outcomes?

   (a) 3
   U(3) = 3^2 = 9

   (b) L($\frac{2}{3}$, 3; $\frac{1}{3}$, 6)
   U(L($\frac{2}{3}$, 3; $\frac{1}{3}$, 6)) = $\frac{2}{3}$U(3) + $\frac{1}{3}$U(6) = 6 + 12 = 18

   (c) -2
   U(-2) = $(-2)^2$ = 4

   (d) L(0.5, 2; 0.5, L(0.5, 4; 0.5, 6))
   U(L(0.5, 2; 0.5, L(0.5, 4; 0.5, 6))) = 0.5U(2) + 0.5(0.5U(4) + 0.5U(6)) = 2 + 0.5(8 + 18) = 15

3. What is the expected monetary value (EMV) of the lottery $L(\frac{2}{3}, \$3; \frac{1}{3}, \$6)$?
   $\frac{2}{3} \cdot \$3 + \frac{1}{3} \cdot \$6 = \$4$

4. For each of the following types of utility function, state how the utility of the lottery $U(L)$ compares to the utility of the amount of money equal to the EMV of the lottery, $U(EMV(L))$. Write $<, >, =$, or ? for can't tell.

   (a) U is an arbitrary function.
   $U(L)$? $U(EMV(L))$

   (b) U is monotonically increasing and its rate of increase is increasing (its second derivative is positive).
   $U(L) > U(EMV(L))$.
   As an example, consider $U = x^2$ from Q2. Then $U(L) = 18$ and $U(EMV(L)) = 4^2 = 16$.

   (c) U is monotonically increasing and linear (its second derivative is zero).
   $U(L) = U(EMV(L))$

(d) U is monotonically increasing and its rate of increase is decreasing (its second derivative is negative).
$U(L) < U(EMV(L))$.
Consider $U = \sqrt{x}$. Then $U(L) = \frac{2}{3} \cdot \sqrt{3} + \frac{1}{3} \cdot \sqrt{6} \approx 1.97$, and $U(EMV(L)) = \sqrt{4} = 2$.

2 MDPs: Micro-Blackjack

In micro-blackjack, you repeatedly draw a card (with replacement) that is equally likely to be a 2, 3, or 4. You can either Draw or Stop if the total score of the cards you have drawn is less than 6. If your total score is 6 or higher, the game ends, and you receive a utility of 0. When you Stop, your utility is equal to your total score (up to 5), and the game ends. When you Draw, you receive no utility. There is no discount ($\gamma = 1$). Let's formulate this problem as an MDP with the following states: $0, 2, 3, 4, 5$ and a *Done* state, for when the game ends.

1. What is the transition function and the reward function for this MDP? The transition function is

$$\begin{aligned}
T(s, Stop, Done) &= 1 \\
T(0, Draw, s') &= 1/3 \text{ for } s' \in \{2, 3, 4\} \\
T(2, Draw, s') &= 1/3 \text{ for } s' \in \{4, 5, Done\} \\
T(3, Draw, s') &= \begin{matrix} 1/3 \text{ if } s' = 5 \\ 2/3 \text{ if } s' = Done \end{matrix} \\
T(4, Draw, Done) &= 1 \\
T(5, Draw, Done) &= 1 \\
T(s, a, s') &= 0 \text{ otherwise}
\end{aligned}$$

The reward function is

$$R(s, Stop, Done) = s, s \leq 5$$
$$R(s, a, s') = 0 \text{ otherwise}$$

2. Fill in the following table of value iteration values for the first 4 iterations.

States	0	2	3	4	5
V_0	0	0	0	0	0
V_1	0	2	3	4	5
V_2	3	3	3	4	5
V_3	10/3	3	3	4	5
V_4	10/3	3	3	4	5

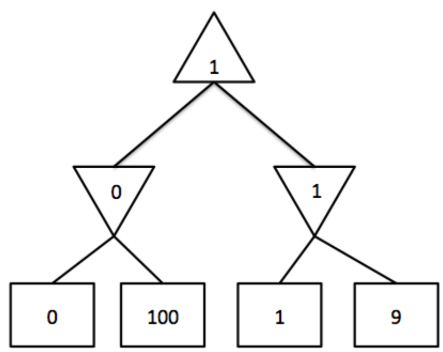
3. You should have noticed that value iteration converged above. What is the optimal policy for the MDP?

States	0	2	3	4	5
π^*	Draw	Draw	Stop	Stop	Stop

3 (Optional) Minimax and Expectimax

In this problem, you will investigate the relationship between expectimax trees and minimax trees for zero-sum two player games. Imagine you have a game which alternates between player 1 (max) and player 2. The game begins in state s_0, with player 1 to move. Player 1 can either choose a move using minimax search, or expectimax search, where player 2's nodes are chance rather than min nodes.

1. Draw a (small) game tree in which the root node has a larger value if expectimax search is used than if minimax is used, or argue why it is not possible.



We can see here that the above game tree has a root value of 1 for the minimax strategy. If we instead switch to expectimax and replace the min nodes with chance nodes, the root of the tree takes on a value of 50 and the optimal action changes for MAX.

2. Draw a (small) game tree in which the root node has a larger value if minimax search is used than if expectimax is used, or argue why it is not possible.

   Optimal play for MIN, by definition, means the best moves for MIN to obtain the lowest value possible. Random play includes moves that are not optimal. Assuming there are no ties (no two leaves have the same value), expectimax will always average in suboptimal moves. Averaging a suboptimal move (for MIN) against an optimal move (for MIN) will always increase the expected outcome.

   With this in mind, we can see how there is no game tree where the value of the root for expectimax is lower than the value of the root for minimax. One is optimal play – the other is suboptimal play averaged with optimal play, which by definiton leads to a higher value for MIN.

3. Under what assumptions about player 2 should player 1 use minimax search rather than expectimax search to select a move?

   Player 1 should use minimax search if he/she expects player 2 to move optimally.

4. Under what assumptions about player 2 should player 1 use expectimax search rather than minimax search?

   If player 1 expects player 2 to move randomly, he/she should use expectimax search. This will optimize for the maximum expected value.

5. Imagine that player 1 wishes to act optimally (rationally), and player 1 knows that player 2 also intends to act optimally. However, player 1 also knows that player 2 (mistakenly) believes that player 1 is moving uniformly at random rather than optimally. Explain how player 1 should use this knowledge to select a move. Your answer should be a precise algorithm involving a game tree search, and should include a sketch of an appropriate game tree with player 1's move at the root. Be clear what type of nodes are at each ply and whose turn each ply represents.

   Use two games trees:

   Game tree 1: max is replaced by a chance node. Solve this tree to find the policy of MIN.

   Game tree 2: the original tree, but MIN doesn't have any choices now, instead is constrained to follow the policy found from Game Tree 1.