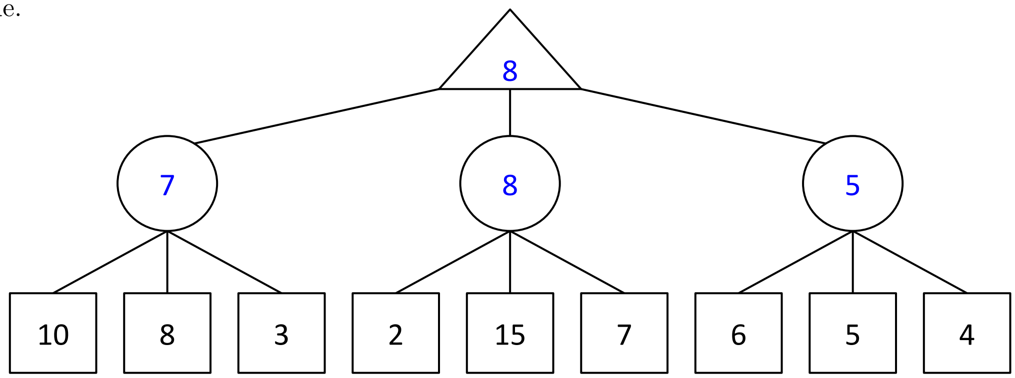
CS188 Fall 2017 Section 4: Games and MDPs

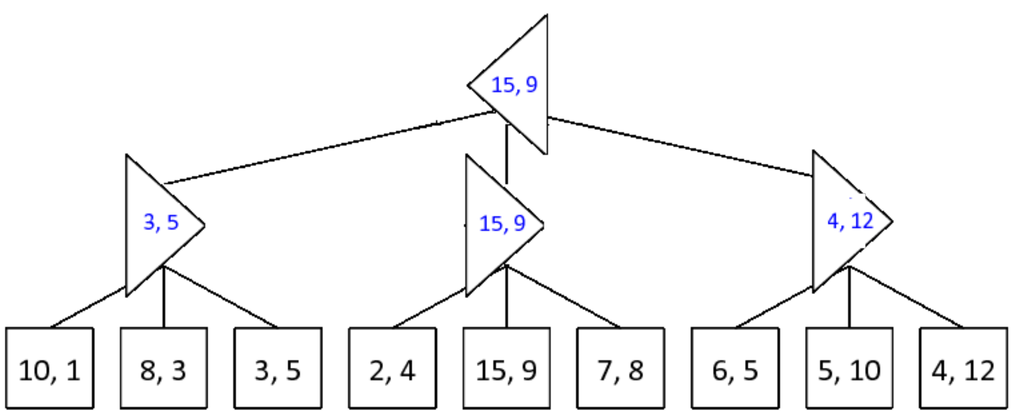
1 Expectimax and Non-Zero-Sum Games

1. Consider the zero-sum game tree below. Instead of a minimizing player, we have a chance node (marked with a circle) that will select one of the three values uniformly at random. Fill in the expectimax value of each node.



2. Which nodes can be pruned from the game tree above through alpha-beta pruning? If no nodes can be pruned, explain why not. No nodes can be pruned. There will always be the possibility that some leaf further down the branch will have a very high value, which increases the overall average value.

3. Let's now look at a non-zero-sum version of this game. In this formulation, player A's utility will be represented as the first of the two leaf numbers, and player B's utility will be represented as the second of the two leaf numbers. Player A plays first (the triangle pointing left), and Player B plays second (the triangles pointing right). Fill in this non-zero game tree assuming each player is acting optimally.



4. Which nodes can be pruned from the game tree above through alpha-beta pruning? If no nodes can be pruned, explain why not. No nodes can be pruned. Because this game is non-zero-sum, there can exist a leaf node anywhere in the tree that is good for both player A and player B.

2 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.

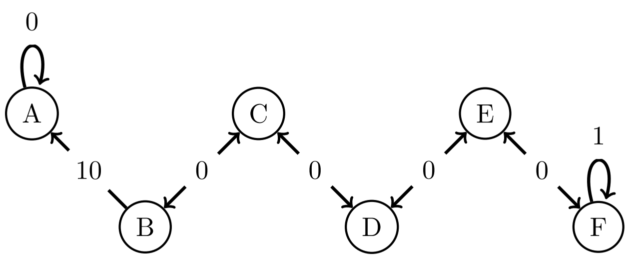
   (a) U is an arbitrary function.
   $U(L)$? $U(EMV(L))$

(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 $U(EMV(L)) = \sqrt{4} = 2$.

3 MDP



Consider the MDP above, with states represented as nodes and transitions as edges between nodes. The rewards for the transitions are indicated by the numbers on the edges. For example, going from state B to state A gives a reward of 10, but going from state A to itself gives a reward of 0. Some transitions are not allowed, such as from state A to state B. Transitions are deterministic (if there is an edge between two states, the agent can choose to go from one to the other and will reach the other state with probability 1).

1. Write down the optimal action at each step if the discount factor is $\gamma = 1$.

   A: Go to A
   B: Go to C
   C: Go to D
   D: Go to E
   E: Go to F
   F: Go to F

2. For each state, does the optimal action depend on γ? If so, for each state, write an equation that would let you determine the value for γ at which the optimal action changes.

   A: Only staying at A is a possible action. For the other states, let n be the number of steps to B, and m be the number of steps to F. Then, the value of going left is $10\gamma^n$ and the value of going right is $\sum_{k=m}^{\infty} \gamma^k = \frac{1}{1-\gamma} - \frac{1-\gamma^m}{1-\gamma}$ because of the geometric series. Now we find the value of γ at which these are equal.

$$10\gamma^n = \frac{1}{1-\gamma} - \frac{1-\gamma^m}{1-\gamma} = \frac{\gamma^m}{1-\gamma}$$
$$10 - 10\gamma = \gamma^{m-n}$$
$$\gamma^{m-n} + 10\gamma - 10 = 0$$

The roots of the above polynomial are the points at which the optimal action changes.