Q3: X-Values

Problem 3: X Values

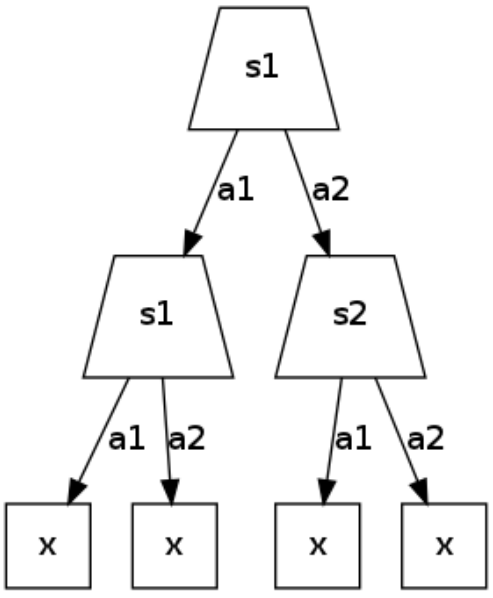
Part 1

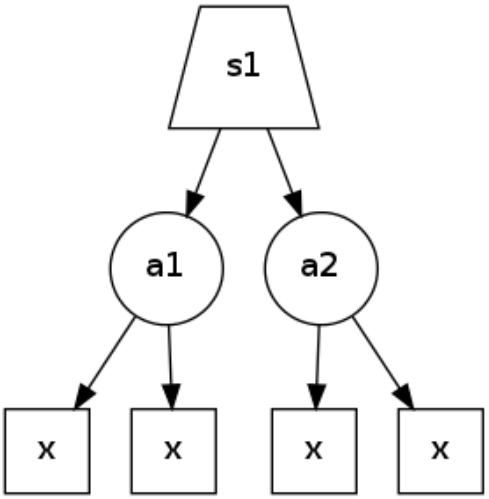
6/6 points (ungraded)

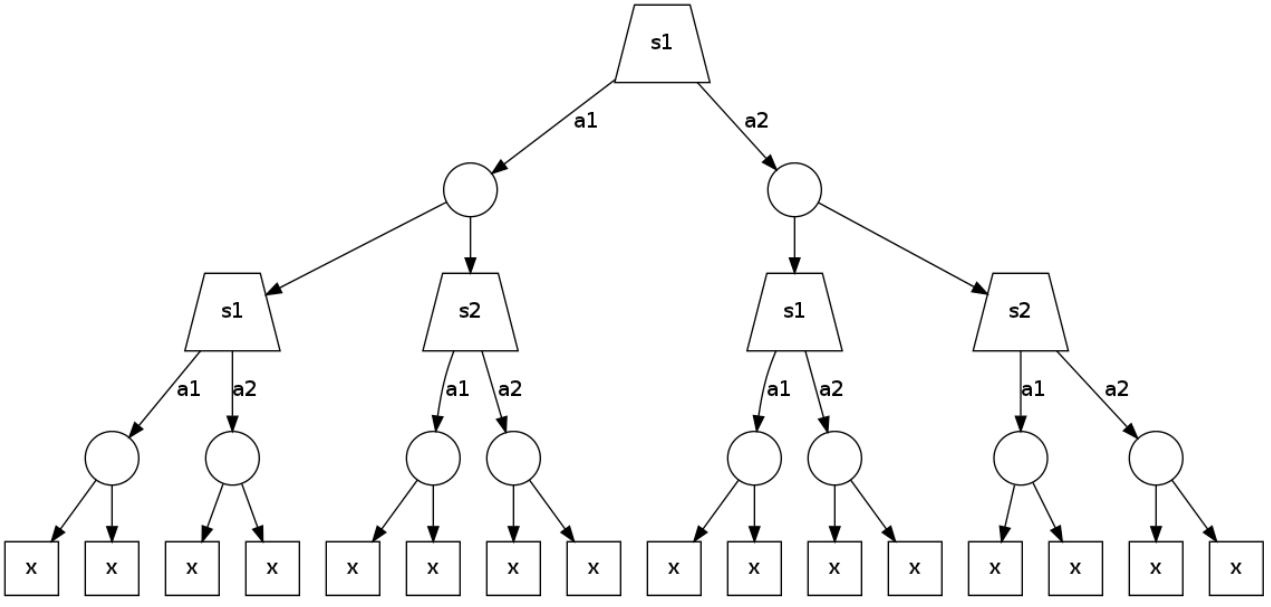
Instead of the Bellman update equation, consider an alternative update equation, which learns the X value function. The update equation, assuming a discount factor $\gamma = 1$, is shown below:

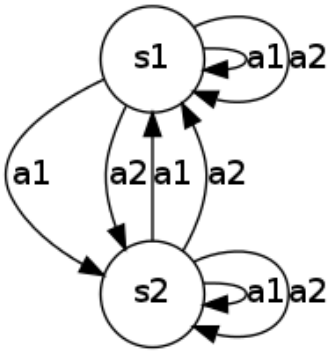
$$X_{k+1}(s) \leftarrow \max_a \sum_{s'} T(s,a,s') \left[R(s,a,s') + \max_{a'} \sum_{s''} T(s',a',s'') \left[R(s',a',s'') + X_k(s'')\right]\right]$$

Assuming we have an MDP with two states, S_1, S_2 and two actions, a_1, a_2, select the expectimax tree rooted at S_1 that corresponds to the alternative update equation




s1
a1
a2
x
x
x
x


s1
a1
a2
s1
s2
s1
s2
a1
a2
a1
a2
a1
a2
a1
a2
x
x
x
x
x
x
x
x
x
x
x
x
x
x
x
x



Submit

✔ Correct (6/6 points)

Part 2

6/6 points (ungraded)

Select the mathematical relationship between the X_k-values learned using the alternative update equation and the V_k-values learned using a Bellamn update equation.

- $X_k(s) = V_{k+1}(s)$
- $X_{k+1}(s) = V_k(s)$
- $X_k(s) = V_{k+2}(s)$
- $X_{k+2}(s) = V_k(s)$
- $X_k(s) = V_k(s)$
- $X_k(s) = V_k(V_{k+1}(s))$
- $X_k(s) = V_{2k}(s)$ ✔
- $X_{2k}(s) = V_k(s)$
- $X_k(s) = V_k(s) + V_{k+1}(s)$

- $X_k(s) = V_k(s) + V_k(s')$

Submit

✔ Correct (6/6 points)