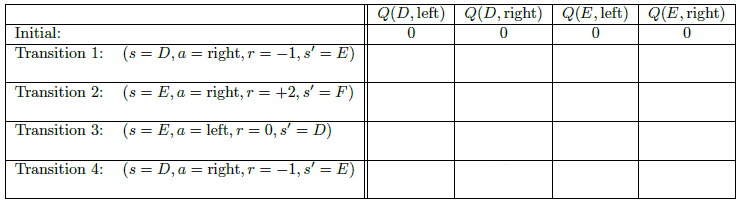
Waterpark World

Part II: Q-Learning

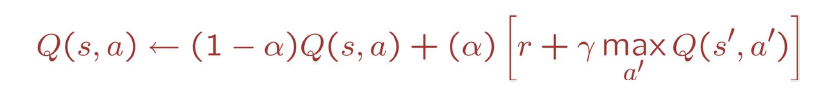
Here is the Waterpark World MDP with seven of its 14 states labeled.



Fill in the blank cells of the following table with the Q-values that result from applying the Q-update for the transition specified on each row. You may leave blank those Q-values that are unaffected by the current update. Use discount γ = 1.0 and learning rate α = 0.5. Assume all Q-values are initialized to 0. (Note: the specified transitions would not arise from a single episode.) For your convenience, the Q-update formula is provided beneath the table below.



|  |  |  |  |
| --- | --- | --- | --- |
|  | -0.5 |  |  |
|  |  |  | +1 |
|  |  | 0 |  |
|  | -0.25 |  |  |



(V-ST-17-05-19. This exercise is based on an example used at U.C. Berkeley)