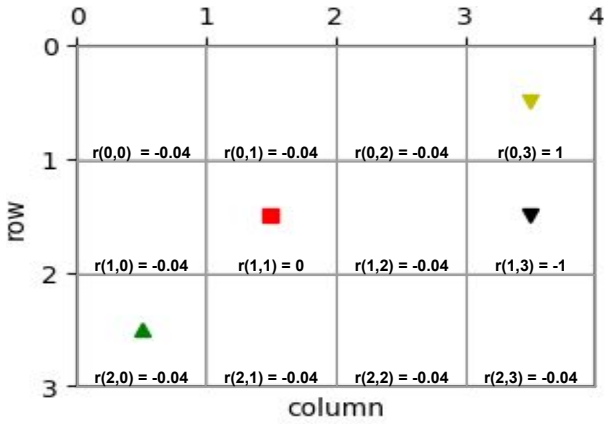

There, are 12 states in this system and 4 possible actions amd each episode ends at "n" when reached terminal state, n <= 20 (episode length)

At, J = 0, X_0 = [2,0], γ = 0.99, α=0.01, $V(X_k) = V(X_k) + \alpha * [r_{k+1} + \gamma * V(X_{k+1}) - V(X_k)]$

While goal is not reached:

- X_0 = [2,0] → perform next action (a_0) → X_1 = [1,0] → Got reward, r_1
- $V(X_0) = V(X_0) + \alpha * [r_1 + \gamma * V(X_1) - V(X_0)] = 0 + 0.01*[-0.04 + 0.99*0 - 0] = -0.0004$
- Update $V(X_0$=[2,0]) = -0.0004

- X_1 = [1,0] → perform next action (a_1) → X_2 = [0,0] → Got reward, r_2
- $V(X_1) = V(X_1) + \alpha * [r_2 + \gamma * V(X_2) - V(X_1)] = 0 + 0.01*[-0.04 + 0.99*0 - 0] = -0.0004$
- Update $V(X_1$=[1,0]) = -0.0004

- X_2 = [0,0] → perform next action (a_2) → X_3 = [0,1] → Got reward, r_3
- $V(X_2) = V(X_2) + \alpha * [r_3 + \gamma * V(X_3) - V(X_2)] = 0 + 0.01*[-0.04 + 0.99*0 - 0] = -0.0004$
- Update $V(X_2$=[0,0]) = -0.0004

- X_3 = [0,1] → perform next action (a_3) → X_4 = [0,1] → Got reward, r_4
- $V(X_3) = V(X_3) + \alpha * [r_4 + \gamma * V(X_4) - V(X_0)] = 0 + 0.01*[-0.04 + 0.99*0 - 0] = -0.0004$
- Update $V(X_3$=[0,1]) = -0.0004

- X_4 = [0,1] → perform next action (a_4) → X_5 = [0,2] → Got reward, r_5
- $V(X_4) = V(X_4) + \alpha * [r_5 + \gamma * V(X_5) - V(X_4)] = -0.0004 + 0.01*[-0.04 + 0.99*0 + 0.0004] = -0.000796$
- Update $V(X_4$=[0,1]) = -0.000796

- X_5 = [0,2] → perform next action (a_5) → X_6 = [0,3] → Got reward, r_6
- $V(X_5) = V(X_5) + \alpha * [r_6 + \gamma * V(X_6) - V(X_5)] = 0 + 0.01*[1 + 0.99*0 - 0] = 0.01$
- Update $V(X_5$=[0,2]) = 0.01

Before j=0 started

V(0,0)=0	V(0,1)=0	V(0,2)=0	V(0,3)=0
V(1,0)=0	V(1,1)=0	V(1,2)=0	V(1,3)=0
V(2,0)=0	V(2,1)=0	V(2,2)=0	V(2,3)=0

After J=0 finished