חלק א' יבש

1. 1. נוסחה מקורית:

נוסחה חדשה:

**Function** VALUE-ITERATION(*mdp*, *ε*) **returns** a utility function

**Inputs**: *mdp*, an MDP with states S, actions A(s), transition model P(s’|s,a), rewards R(s,a,s’),

discount

*ε*, the maximum error allowed in the utility of any state

**local variables**: U,U’, vectors of utilities for states in S, initially zero

, the maximum change in the utility of any state in an iteration

**Repeat**

**For each** *s* **in** *S* **do**

**If** **then**

**Until**

**Return** U

**Function** POLICY-ITERATION(*mdp*) **returns** a policy

**Inputs**: *mdp*, an MDP with states S, actions A(s), transition model P(s’|s,a)

**Local variables**: U, a vector of utilities for states in S, initially zero

, a policy vector indexed by state, initially random

**Repeat**

**for each** s **in** *S* **do**

**if**

**Until** unchanged?

**Return**

להשלים: התנאים שצריכים להתקיים במקרה של כדי למצוא מדיניות אופטימלית.



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | U0(si) | U1(si) | U2(si) | U3(si) | U4(si) | U5(si) | U6(si) |
| s1 | 0 | -1 | -1 | -1 | -1 | -1 | -1 |
| s2 | 0 | -1 | -2 | -2 | -2 | -2 | -2 |
| s3 | 0 | -1 | -2 | -3 | -3 | -3 | -3 |
| s4 | 0 | -1 | -2 | -3 | -4 | -4 | -4 |
| s5 | 0 | -1 | -2 | -3 | -4 | -5 | -5 |
| s6 | 0 | -1 | -2 | -3 | -4 | -4 | -4 |
| s7 | 0 | -1 | -2 | -3 | -3 | -3 | -3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| s1 | ↑ | ↑ | ↑ | ↑ |
| s2 | ↑ | ↑ | ↑ | ↑ |
| s3 | ← | ← | ← | ← |
| s4 | ↑ | ↑ | ↑ | ↑ |
| s5 | → | → | → | → |
| s6 | → | → | ↑ | ↑ |
| s7 | ↓ | → | → | → |