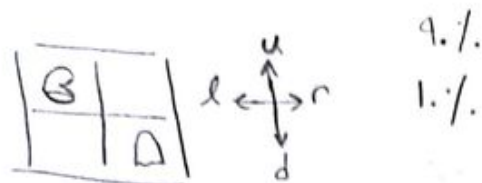


نیمه

سارا دانش ۹۸۱۷۵۴۱۸



رنده مادل +1

MDP \rightarrow 3 state

State, Action, transition Probability

$$M = (S, A, R, P, \gamma), 0 < \gamma < 1$$

$$\forall s \in S: \sqrt{V^{\pi_1}(s)} > \sqrt{V^{\pi_2}(s)} \rightarrow \pi_1 > \pi_2, \sqrt{V} = \text{value function}$$

V^{π}

(ا) جدا کردن سیاست و مقداردهی از ارزش سیاست ها

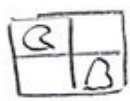
deterministic

دقیقت به کار بردن و در عمل است



(4) یا در کار بردن سیاست

بسیار از کار بردن سیاست

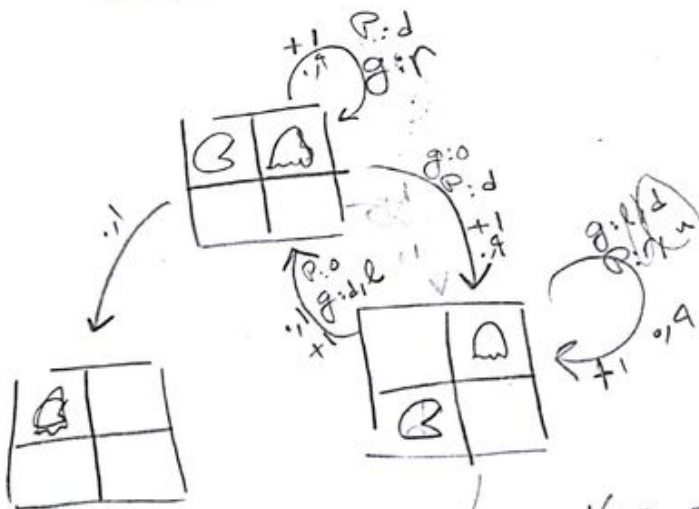


(3) یا به صورت قطعی قرار دادن



(3) یا در کار بردن سیاست

ghost = g
Paeman = P



MDP است که به نام است در تئوری تیرم
reward ها برای آن ها است که می تواند به صورت ریاضی

$$V_{k+1}^{\pi}(s) \leftarrow \sum_{s'} T(s, \pi(s), s') [R(s, \pi(s), s') + \gamma V_k^{\pi}(s')]$$

ب آسان تر است