

Q3 №2 RL Сложная ЮА.

Найти точное символьное выражение для оптимальной в-ти выбора действия right

С вероятностью p движения вправо по боковым аткам мы можем записать матрицу перехода

$$Q = \begin{bmatrix} 1-p & p & 0 \\ p & 0 & 1-p \\ 0 & 1-p & 0 \end{bmatrix}$$

Из этого у нас есть метод расчёта ожидаемых шагов между состояниями, что будет суммой 1ой строки

$$M = (I - Q)^{-1}$$

$$M = (I - Q)^{-1} = \begin{bmatrix} \frac{2}{p} + \frac{1}{1-p} & \frac{1}{p} + \frac{1}{1-p} & \frac{1}{p} \\ \frac{1}{p} + \frac{1}{1-p} & \frac{1}{p} + \frac{1}{1-p} & \frac{1}{p} \\ \frac{1}{p} & \frac{1}{p} & \frac{1}{p} \end{bmatrix}$$

Т.е. ожидаемое число шагов

$$E(t) = \frac{4}{p} + \frac{2}{1-p}$$

и учитывая, что награда = 1

$$E(t) = \frac{4}{p} + \frac{2}{1-p}$$

$$J(p) = -E(t) = -\frac{4}{p} - \frac{2}{1-p}$$

$$\frac{dJ(p)}{dp} = \frac{4}{p^2} - \frac{2}{(1-p)^2} = 0$$

$$2p^2 = 4(1-p)^2$$

$$p^2 - 4p + 2 = 0$$

и решением будет являться

$$p = 2 \pm \sqrt{2}, \text{ ну и из условия, что } p \in [0, 1] - \text{вероятность}$$

$$\text{Отв: } \hat{p} = 2 - \sqrt{2} \approx 0.586$$