

Ejercicios de Aprendizaje por Refuerzo

J. F. González

*Facultad de Ingeniería
Universidad de Buenos Aires*

E-mail: jfgonzalez@fi.uba.ar

Índice

1. Cadenas de Markov	1
1.1. Ejercicio 1.1 - Estudiante	1
1.2. Ejercicio 1.2 - Recompensas	1
1.3. Ejercicio 1.3 - La Rata	1

1. Cadenas de Markov

1.1. Ejercicio 1.1 - Estudiante

Ver el notebook aqui:

1.2. Ejercicio 1.2 - Recompensas

Dada la cadena de Markov de la Figura 1 encontrar la esperanza de la ganancia en dos pasos $\mathbb{E}_1[g(S_1, S_2)]$ cuando las recompensas son $r(1) = -2$, $r(2) = 3$ y $r(3) = 5$ cuando $g(S_1, S_2) = r(S_1) + r(S_2)$.

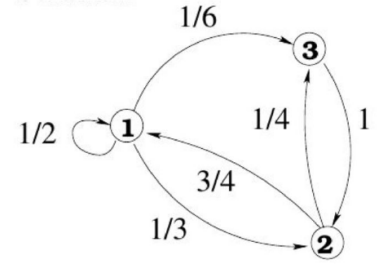


Figura 1. Cadena de Markov con recompensa

$$\begin{aligned}\mathbb{E}_1[g(\vec{S})] &= \sum_{\vec{s}} g(\vec{s}) p_{\vec{s}}(\vec{s}) = \sum_{\vec{s}} g(\vec{s}) \mathbb{P}(S_1 = s_1, S_2 = s_2 | S_0 = 1) \\ &= (-2 - 2)p(1, 1) + (-2 + 3)p(1, 2) + (-2 + 5)p(1, 3) + \\ &\quad + (3 - 2)p(2, 1) + (3 + 3)p(2, 2) + (3 + 5)p(2, 3) + \\ &\quad + (5 - 2)p(3, 1) + (5 + 3)p(3, 2) + (5 + 5)p(3, 3) \\ &= -4 \cdot \left(\frac{1}{2} \cdot \frac{1}{2}\right) + 1 \cdot \left(\frac{1}{2} \cdot \frac{1}{3}\right) + 3 \cdot \left(\frac{1}{2} \cdot \frac{1}{6}\right) + \\ &\quad + 1 \cdot \left(\frac{1}{3} \cdot \frac{3}{4}\right) + 6 \cdot \left(\frac{1}{2} \cdot 0\right) + 8 \cdot \left(\frac{1}{3} \cdot \frac{1}{4}\right) + \\ &\quad + 3 \cdot \left(\frac{1}{6} \cdot 0\right) + 8 \cdot \left(\frac{1}{6} \cdot \frac{1}{1}\right) + 10 \cdot \left(\frac{1}{6} \cdot 0\right) \\ &\approx 1,66\end{aligned}$$

1.3. Ejercicio 1.3 - La Rata

Una rata está atrapada ...

Definimos un retorno cuando se tarda T pasos en finalizar como

$$G_0 = R_1 + G_1$$

$$\begin{aligned}\mathbb{E}[G_0] &= 3\left(\frac{1}{3}\right) + 13\left(\frac{1}{3}\right) + 8\left(\frac{1}{3}\right) + \\ &\quad + 13\left(\frac{1}{3} \cdot \frac{1}{2}\right) + 5\left(\frac{1}{3} \cdot \frac{1}{2}\right) + 8\left(\frac{1}{3} \cdot \frac{1}{2}\right) + 5\left(\frac{1}{3} \cdot \frac{1}{2}\right) + \\ &\quad + 8\left(\frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2}\right) + 5\left(\frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2}\right) + 13\left(\frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2}\right) + 5\left(\frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2}\right) + \dots \\ &= \frac{26}{3} + 13 \frac{1}{3} \frac{1}{2} \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i + 5 \frac{1}{3} \frac{1}{2} \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i + 8 \frac{1}{3} \frac{1}{2} \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i \\ &= \frac{26}{3} + (13 + 5 + 8) \frac{1}{3} \frac{1}{2} \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i \\ &= \frac{26}{3} + \frac{26}{6} \frac{1}{1 - 0,5} \\ &\approx 10,8\end{aligned}$$

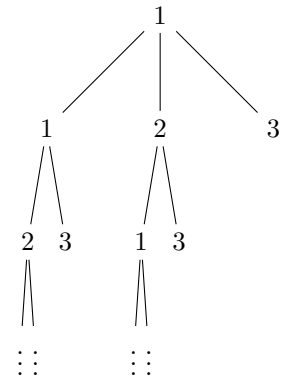


Figura 2.