

Procesos de decisión de Markov.

Los MDP (Procesos de Decision de Markov) una forma idealizada matematicamente del problema de aprendizaje por refuerzo, para el cual se podría encontrar un enunciado teórico preciso que pueda describirla.

Propiedad de Markov. ¶

Esta propiedad de Markov nos muestra que el futuro es independiente del pasado, dado el presente, expresado de la siguiente forma.



$$P[\mathbf{S}_{t+1}] = P[\mathbf{S}_{t+1} | \mathbf{S}_1, \dots, \mathbf{S}_t]$$

Donde \mathbf{S}_t representa el estado actual y contiene toda la información relevante de los estados pasados $\mathbf{S}_1, \dots, \mathbf{S}_t$, la que es información suficiente y no se requiere mayor información.

Matriz de transacción de estados.

En una matriz de transacción pasaremos a ver como pasa de un estado S a un S' , en donde cada fila sumaría uno, se veria así:

$$\begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & & \\ P_{n1} & & P_{nn} \end{bmatrix}$$



Simplificando es una secuencia de estados aleatorios de Markov, se puede definir como una tupla $\langle S, P \rangle$.

Donde:

- $S \rightarrow$ Es una lista de estados de las que puede pertenecer.
- $P \rightarrow$ Es una matriz de transacción de estados.