Q-Learning: Obtener la Q-table

DATA SCIENCE
ONLINE

¿Cómo se obtiene la tabla-Q?

- Inicializamos la tabla-Q intermedia con valores nulos
- El agente explorará de forma más o menos aleatoria el entorno
- Obteniendo valores de recompensa y actualizando una tabla-Q intermedia
- Continúa haciéndolo hasta alcanzar un criterio de parada

El agente “explora” el entorno

- Puede hacerlo de forma completamente aleatoria o

EXPLORATION 

- … combinando esta aleatoriedad con seguir la tabla-Q intermedia

EXPLOTATION 

- Épsilon (ε) es el hiperparámetro que regula la dicotomía Exploration vs Explotation
- ε -greedy

Actualización de la tabla-Q

$$Q(s, a) = (1 - \alpha) * Q(s, a) + \alpha * (r + \gamma * \max(Q(s', a')))$$

- Q(s, a) representa el valor Q para el estado s y la acción a.
- α (alfa) es la tasa de aprendizaje.
- r es la recompensa obtenida al tomar la acción a en el estado s.
- s' representa el nuevo estado resultante
- max(Q(s', a')) es el valor máximo de Q para el nuevo estado s' considerando todas las acciones posibles a’.
- γ (gamma) es el factor de descuento.

Algoritmo

1. Inicializamos la tabla Q

2. Dar valores a los hiperparámetros. Valores de referencia:
   $\alpha = 0.05$; $\gamma = 0.9$; $\varepsilon = 0.1$

3. Ejecutar esta secuencia:
   I. Comprobar la condición de parada, si no se cumple ir a II
   II. Observar el estado actual (s)
   III. Escoger una acción aleatoria (explorar) o la acción con mayor Q(s,a) actual (explotar), en función de épsilon
   IV. Ejecutar la acción
   V. Actualizar Q(s,a) según la ecuación
   VI. Volver a I