Avance Tarea 1: Programación dinámica

Código: EL7021-1

Nombre: José Luis Cádiz Sejas

Pregunta 1:

• **Espacio de estados:** Dado el espacio $(x,y) \in \Re^2$, definimos el espacio de estados:

$$S = \{(RewardGgrid(x, y) = -1) \text{ or } (RewardGgrid(x, y) = 0)\}$$

Donde RewardGgrid(x,y) es la función de recompensa que puede generar valores -1, 0 o NULL según si el estado es de transición, terminal o no factible respectivamente.

En particular si (RewardGgrid(x, y) = 0), estamos hablando del estado terminal:

$$s_t \in S \mid (RewardGgrid(x, y) = 0)$$

- Espacio de acciones: α ∈ {0,1,2,3} donde
 {"0":"up","1":"down","2":"right","3":"left"}
- Función de recompensa: Función independiente de las acciones.

$$R(s,a) = \begin{cases} -1 & \text{if } s \neq s_t \\ 0 & \text{if } s = s_t \end{cases} donde s \in S$$

• **Función de transición de estados**: Solo pasa al siguiente estado si esta dentro del dominio, si no, se mantiene en el estado inicial.

$$T(s', s, a) = \begin{cases} s & \text{if } s' \notin S \\ s' & \text{if } s' \in S \end{cases}$$

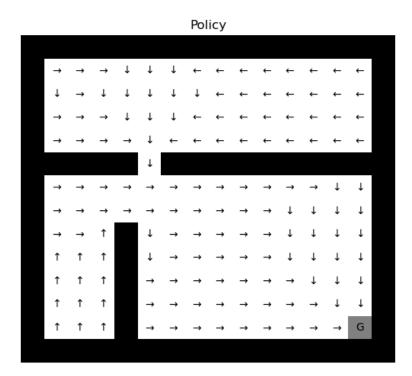
Pregunta 2: Código adjunto.

Pregunta 3:

• Función de valor:

Value Function -9.5 -9.5 -9.4 -9.3 -9.3 -9.3 -9.4 -9.5 -9.5 -9.6 -9.6 -9.7 -9.7 -9.7 -9.5 -9.4 -9.3 -9.2 -9.2 -9.2 -9.3 -9.4 -9.5 -9.5 -9.6 -9.6 -9.7 -9.7 -9.4 -9.3 -9.2 -9.1 -9.0 -9.1 -9.2 -9.3 -9.4 -9.5 -9.6 -9.6 -9.7 -9.7 -9.4 -9.3 -9.1 -9.0 -8.9 -9.0 -9.1 -9.3 -9.4 -9.4 -9.5 -9.6 -9.6 -9.7 -8.7 -9.1 -9.0 -8.8 -8.7 -8.5 -8.3 -8.0 -7.8 -7.5 -7.2 -6.9 -6.5 -6.0 -5.7 -9.1 -8.9 -8.8 -8.6 -8.3 -8.1 -7.9 -7.6 -7.3 -6.9 -6.5 -6.0 -5.5 -5.1 -9.2 -9.1 -8.9 -8.1 -7.9 -7.6 -7.3 -6.9 -6.5 -6.0 -5.4 -4.9 -4.3 -9.3 -9.2 -9.1 -7.9 -7.6 -7.3 -6.9 -6.5 -6.0 -5.4 -4.8 -4.1 -3.5 -7.6 -7.3 -6.9 -6.5 -6.0 -5.4 -4.8 -4.1 -3.3 **-2.5** -9.3 -9.3 -9.2 -9.4 -9.4 -9.3 -7.3 -7.0 -6.5 -6.0 -5.5 -4.9 -4.1 -3.3 -2.4 -1.3 -9.5 -9.4 -9.4 -7.2 -6.7 -6.3 -5.7 -5.1 -4.3 -3.5 -2.5 -1.3 0.0

Política aprendida:



• **Número de iteraciones sobre la función de valor**: 253 iteraciones en los 11 llamados que se hizo a la función policy_evaluation.

Iteración Policy	Iteraciones dentro de
evaluation	Policy evaluation
1	84
2	16
3	16
4	41
5	20
6	19
7	34
8	15
9	6
10	1
11	1