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:

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

  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**: $a \in \{0,1,2,3\}$ donde
  `{"0":"up","1":"down","2":"right","3":"left"}`

- **Función de recompensa**: Función independiente de las acciones.

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

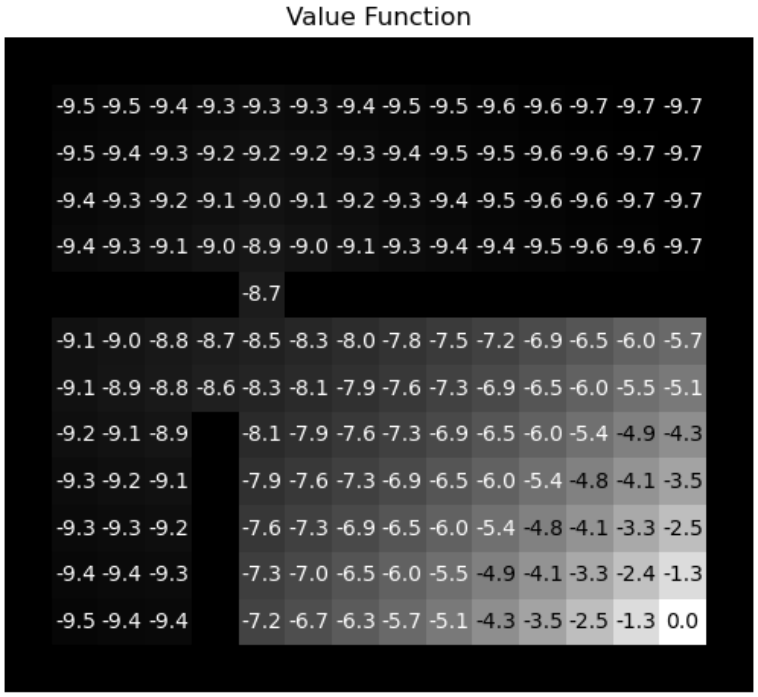
- **Función de transición de estados**: Solo pasa al siguiente estado si esta dentro del dominio, si no, se mantiene en el estado inicial.

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

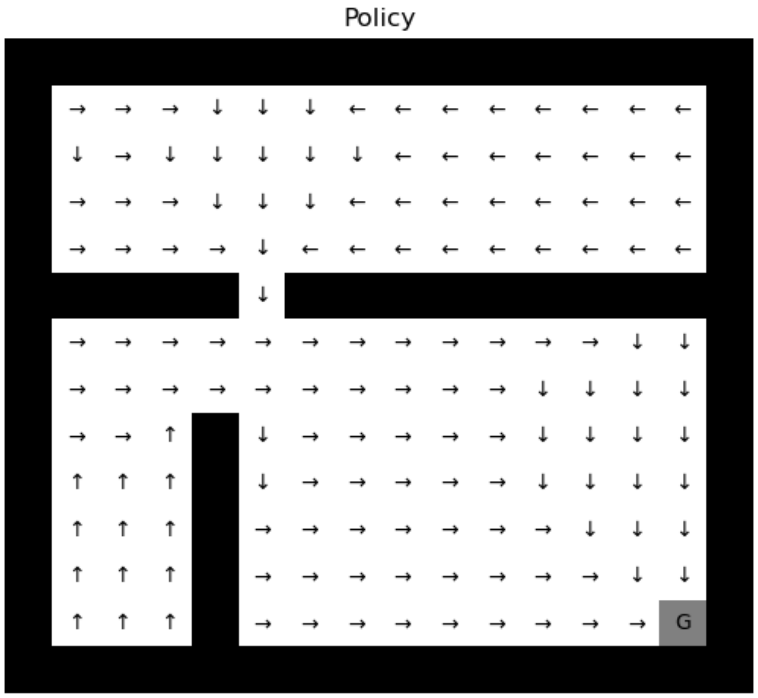
Pregunta 2: Código adjunto.

Pregunta 3:

- **Función de valor**:



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