**Avance Tarea 1:** Programación dinámica

**Código:** EL7021-1

**Nombre:** José Luis Cádiz Sejas

**Pregunta 1:**

* **Espacio de estados:** Dado el espacio , definimos el espacio de estados:

Donde 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.

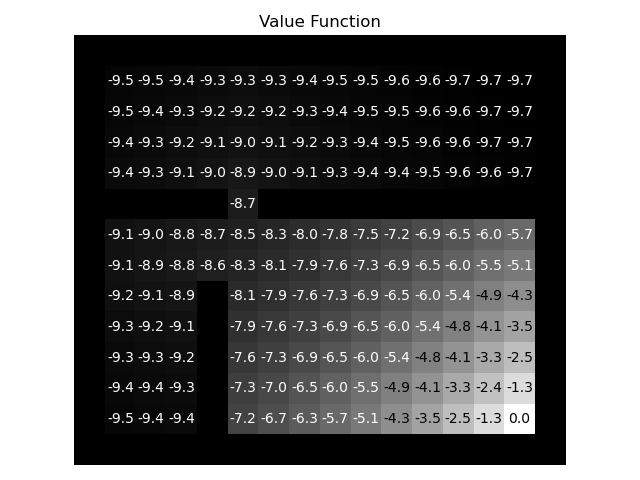
En particular si , estamos hablando del estado terminal:

* **Espacio de acciones**: donde {"0":"up","1":"down","2":"right","3":"left"}
* **Función de recompensa**: Función independiente de las acciones.
* **Función de transición de estados**: Solo pasa al siguiente estado si esta dentro del dominio, si no, se mantiene en el estado inicial.

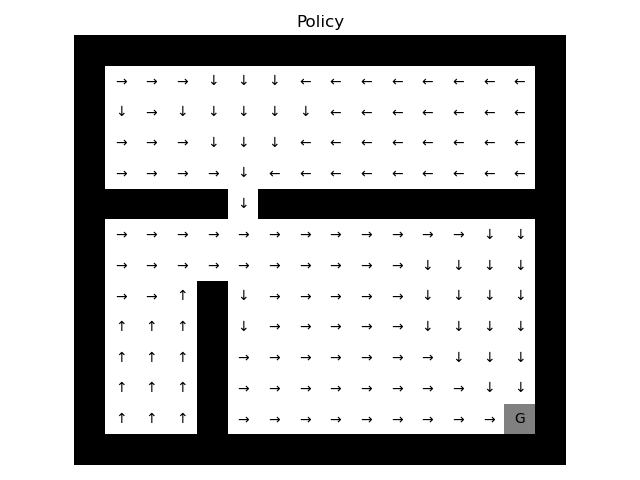
**Pregunta 2:** Código adjunto.

**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 |