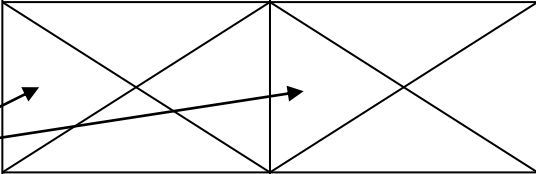


TAREFA

- Ambiente discreto 4x4, com obstáculos.
- Agente deve alcançar posição destino **D** a partir de **qualquer lugar** do ambiente.
- **D** é um estado absorvente (ao atingir **D**, o episódio termina): $V^*(D) = 0$
- Ações que o agente pode realizar: N, S, L, O
- Penalidade por executar uma **ação** (qualquer) = -1
 - ◆ Melhor política \Rightarrow caminho mais curto
- Considerar $\gamma = 1$ e MDP determinístico ($p=1$)

Ambiente

(1,4)	(2,4)	(3,4)	(4,4)
(1,3)	(2,3)	(3,3)	(4,3)
		(3,2)	(4,2)
D (1,1)	(2,1)	(3,1)	(4,1)

Obstáculos

Ações = {N, S, L, O}

Tarefa: algoritmo de iteração de valor para MDP determinístico

- Cálculo iterativo da função valor ótima.

$$V(s) \leftarrow r_{s,a} + \max_a (V(s'))$$

Repetir até $V(s)$ estabilizar.

Sendo:

s – estado atual, s' – próximo estado,

$r_{s,a}$ – reforço recebido por executar a em s

$V(.)$ – valor do estado

Exemplo de cálculo de $V(s)$

Início

0	0	s'_2 0	0
0	s'_1 0	s 0	s'_3 0
		s'_4 0	0
D 0	0	0	0

Iteração 1 (quando calculous para todos os estados)

		-1	
D 0			

■ ■ ■

$$V(s) = \max_a ((r(s, O) + V(s'_1)), \\ (r(s, N) + V(s'_2)), \\ (r(s, L) + V(s'_3)), \\ (r(s, S) + V(s'_4)))$$

$$= \max_a ((-1+0), (-1+0), (-1+0), (-1+0)) \\ = -1$$

Tarefa

- Entrega: Mostrar o valor no espaço de estados (grade com o valor em cada célula) após CADA ITERAÇÃO, até a convergência do algoritmo VI
- Responder:
 - ◆ Qual estado tem valor **mínimo**? Qual o valor deste estado?
 - ◆ Qual estado tem valor **máximo**? Qual o valor deste estado?
 - ◆ Mostrar na grade qual é a **política ótima** em cada célula.