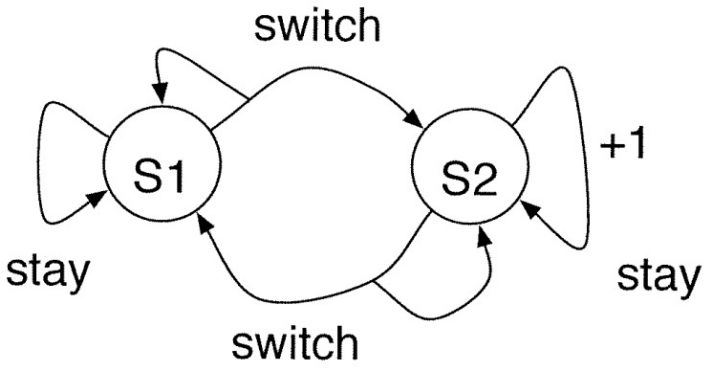
UNIVERSITÀ DEGLI STUDI DI SALERNO
DIPARTIMENTO DI INFORMATICA

Intelligenza Artificiale

Esercizi RL

a.a. 2022/2023

Esercizio 1



Consideriamo il seguente MDP.

Ci sono due stati S1 ed S2 e due azioni *switch* e *stay*. L'azione di *switch* sposta l'agente nell'altro stato con probabilità 0,8 e rimane nello stesso stato con probabilità 0,2.

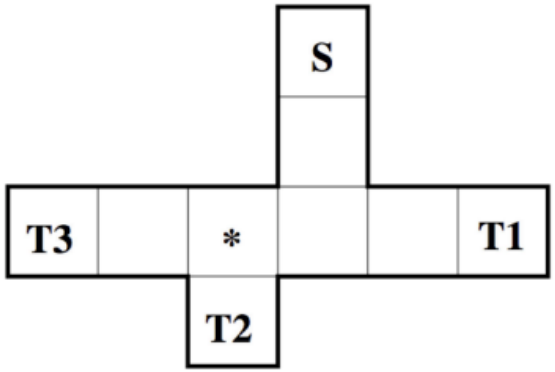
L'azione *stay* lascia l'agente nello stesso stato con probabilità 1.

La ricompensa per l'azione *stay* nello stato S2 è 1. Tutte le altre ricompense sono 0. Il fattore di sconto $\gamma = 0.5$.

- Qual è la policy ottimale?
- Calcola la value function ottimale risolvendo il sistema di equazioni lineari che corrispondono alla politica ottimale.

Esercizio 2

- Considera il gridworld mostrato di seguito. C'è uno stato iniziale S e tre stati terminali, T1, T2 e T3. Il task è episodico e scontato, con ogni episodio che inizia in S e termina in uno degli stati terminali. Gli stati terminali T1, T2 e T3 forniscono rispettivamente ricompense di 2, 4 e 6. Passare allo stato contrassegnato con * fornisce una ricompensa di 1. Tutte le altre ricompense sono 0. La politica ottimale dipende dal valore del tasso di sconto γ, $0 \leq \gamma \leq 1$
- Per quali valori di gamma una politica ottimale porta l'agente a T1?
- Per quali valori di gamma una politica ottimale porta l'agente a T2?
- Per quali valori di gamma una politica ottimale porta l'agente a T3?



Esercizio 3

- Supponiamo di osservare i seguenti 9 episodi generati da un processo di ricompensa Markov sconosciuto, dove A e B sono stati e i numeri sono ricompense:

A,0,B,4	B,0,A,1,B,2	B,4
A,2	B,0,A,2	B,2
A,1,B,0,A,2	B,2	B,4

- Fornire i valori per gli stati A e B che sarebbero ottenuti dal metodo first-visit Monte-Carlo batch utilizzando questo dataset (supponendo no sconti).
- Fornire i valori per gli stati A e B con TD batch.

Esercizio 4

- Supponiamo di avere un fattore di sconto di 0,5 ed osserviamo la seguente sequenza di ricompense:

R1 = -1, R2 =2, R3 =6, R4 =3, R5 =2, con T =5

Seguito dallo stato terminale. Quali sono i valori di ritorno di:

G_5 =

G_4 =

G_3 =

G_2 =

G_1 =

G_0 =

Esercizio 5

- Supponiamo di avere $\gamma = 0{,}5$ ed osserviamo la seguente sequenza di ricompense:

$$R_1 = 7;\ R_2 = 6;\ R_3 = -4;\ R_4 = 4;\ R_5 = 8;\ R_6 = 2;$$

- Seguito dallo stato terminale. Quali sono i valori di ritorno di:

$G_6 =$

$G_5 =$

$G_4 =$

$G_3 =$

$G_2 =$

$G_1 =$

$G_0 =$