

Tarea 1: Reinforcement Learning
Universidad de Concepcion
Diego Neira
19-10-2025

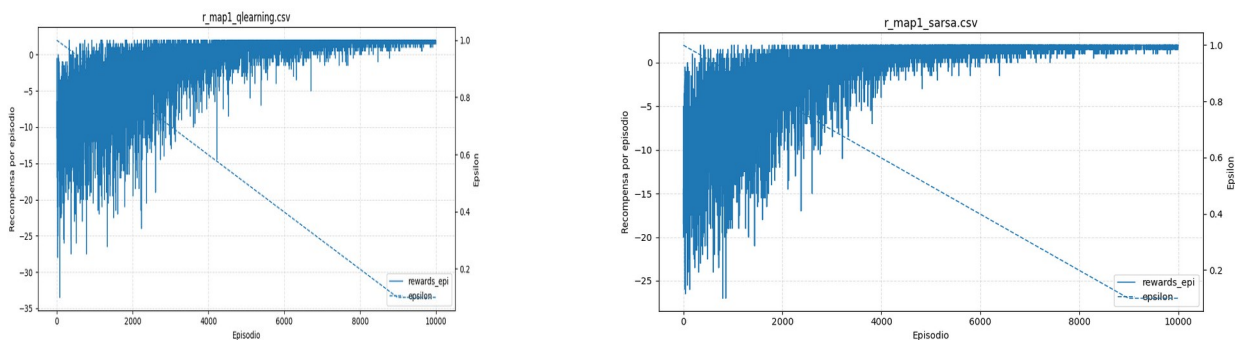
Pregunta 1:

El segundo mapa tiene una mayor complejidad por lo que Q no se completa con los hiperparametros originales. La primera observación fue que el camino óptimo era largo por lo que se procedió a aumentar el numero de pasos que el modelo podía tomar en cada iteración.

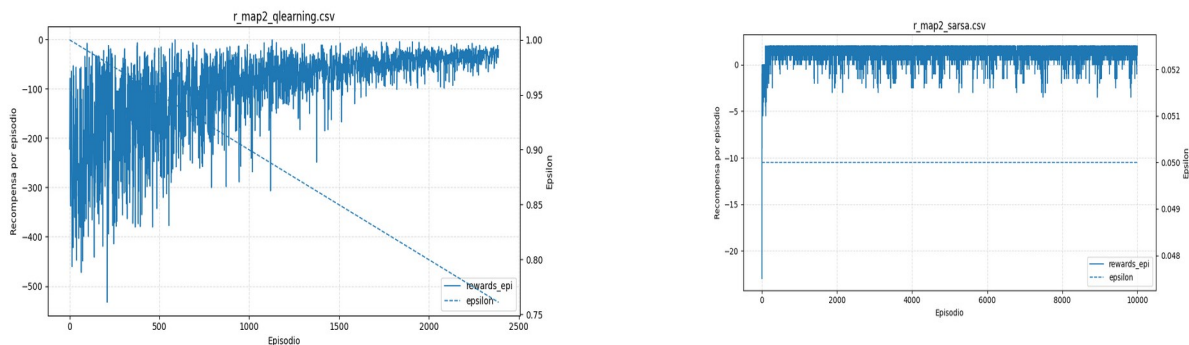
Esto no resulto en cambios en ninguno de los algoritmos por lo que se consideró que lo largo del camino hacía que la tasa de descuento se aplicara en una proporción muy grande sobre la recompensa del estado de término. Por lo que se aumentó el valor a 0,99 que es prácticamente como si las recompensas futuras fueran inmediatas.

Así se obtuvo logró que Q-learning pudiera encontrar un camino pero solo SARSA pudo encontrar el camino óptimo.

Mapa 1:

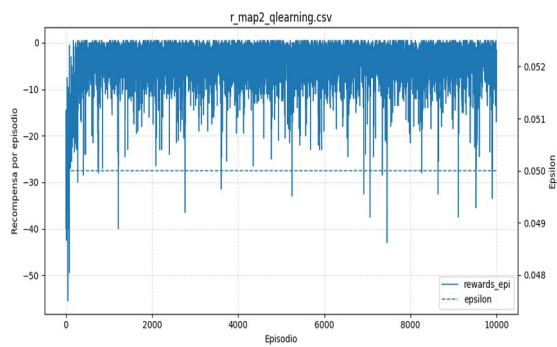
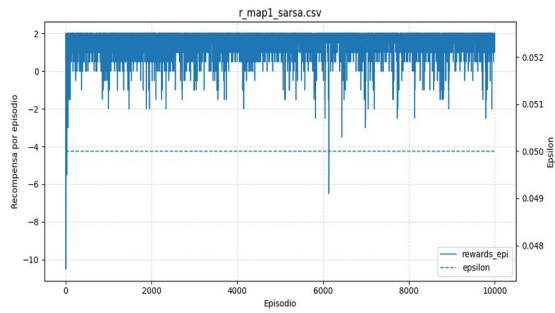
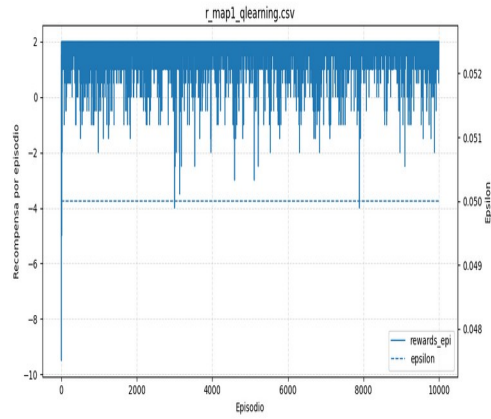


Mapa 2:



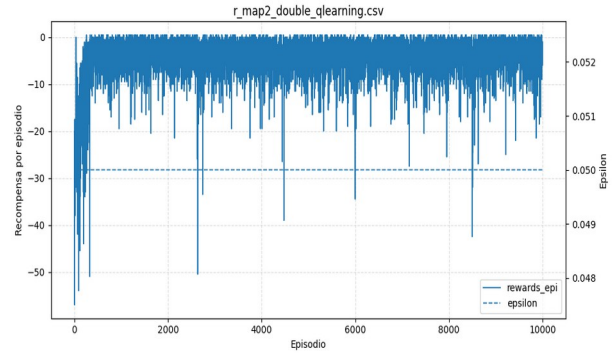
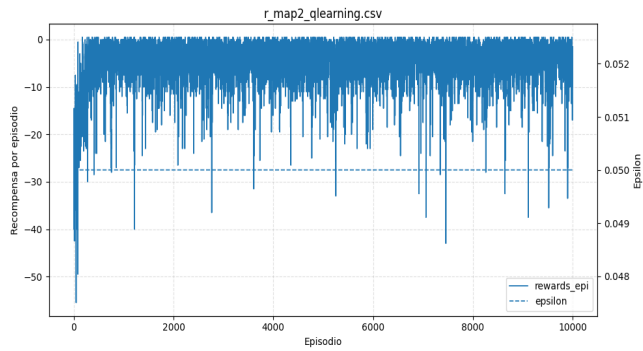
Pregunta 2:

Con un modelo no estocástico el desempeño cae y si bien los modelos llegan a los mismos caminos óptimos, estos obtienen resultados mucho menores:



Pregunta 3

Se puede apreciar como los valores se acercan mucho más rápida en double q learning pero son mas inestables, teniendo picos mas lejanos a los valores óptimos.



Pregunta 4

