



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat d'Informàtica de Barcelona



Sistemas Inteligentes Distribuidos

Agente de aprendizaje por refuerzo para el entorno Cliff Walking

Lluc Martínez Busquets
Eric Medina León
Àlex Rodríguez Rodríguez

2024/2025 Q2

Resumen

Este trabajo presenta un estudio detallado sobre la implementación de cuatro algoritmos de aprendizaje por refuerzo en el entorno *CliffWalking-v0* de la librería de Python *Gymnasium*: *Value Iteration*, *Direct Estimation*, *Q-Learning* y *REINFORCE*. El entorno se configura con el modo *is_slippery* activado, lo cual introduce estocasticidad en las transiciones de estado. El objetivo es evaluar el rendimiento de cada algoritmo, así como qué parámetros e hiperparámetros son los óptimos para su funcionamiento.

Índice

1	Introducción	3
1.1	Descripción del problema	3
1.2	Caracterización del entorno	3
1.3	Entorno experimental	3
2	Value Iteration	3
2.1	Descripción del algoritmo	3
2.2	Experimentación	3
3	Direct Estimation	3
3.1	Descripción del algoritmo	3
3.2	Experimentación	3
4	Q-learning	4
4.1	Descripción del algoritmo	4
4.2	Experimentación	4
5	Reinforce	4
5.1	Descripción del algoritmo	4
5.2	Experimentación	4
6	Conclusiones	4

1. Introducción

1.1. Descripción del problema

1.2. Caracterización del entorno

1.3. Entorno experimental

2. Value Iteration

2.1. Descripción del algoritmo

2.2. Experimentación

Observación	
Planteamiento	
Hipótesis	
Método	▪

Cuadro 1: Experimento 1

3. Direct Estimation

3.1. Descripción del algoritmo

3.2. Experimentación

Observación	
Planteamiento	
Hipótesis	
Método	▪

Cuadro 2: Experimento 1

4. Q-learning

4.1. Descripción del algoritmo

4.2. Experimentación

Observación	
Planteamiento	
Hipótesis	
Método	▪

Cuadro 3: Experimento 1

5. Reinforce

5.1. Descripción del algoritmo

5.2. Experimentación

Observación	
Planteamiento	
Hipótesis	
Método	▪

Cuadro 4: Experimento 1

6. Conclusiones