

Tabla de Contenido

1. Introducción	1
1.1. Contexto y justificación	1
1.2. Objetivos	2
1.2.1. Objetivo general	2
1.2.2. Objetivos específicos	2
2. Marco teórico y estado del arte	3
2.1. Control de sistemas	3
2.1.1. Control PID	3
2.1.2. Regulador lineal cuadrático	4
2.1.3. Control predictivo por modelo	4
2.2. Aprendizaje reforzado	5
2.2.1. Proceso de decisión de Markov	5
2.2.2. Tipos de algoritmos de aprendizaje reforzado	7
2.2.2.1. Q-Learning	8
2.2.2.2. Actor-Critic	8
2.3. Aprendizaje reforzado profundo	9
2.3.1. Deep Q Learning	9
2.3.1.1. Double Deep Q Learning	9
2.3.1.2. Dueling Deep Q Learning	10
2.3.1.3. Deep Q Learning from Demonstration	11
2.3.1.4. Neural Fitted Q Iteration	11
2.3.2. Soft Actor Critic	12
2.4. Aprendizaje reforzado fuera de línea	12
2.4.1. Estrategias de aprendizaje reforzado offline	14
2.4.1.1. Conservative Q Learning	14
2.4.1.2. Implicit Q Learning	15
2.4.1.3. Batch Constrained Q Learning	16
3. Formulación del problema y metodología de trabajo	17
3.1. Proceso dinámico a controlar: El péndulo invertido	17
3.2. Python y Gym	18
3.3. Datos para el entrenamiento fuera de línea	19
3.3.1. Función de recompensa	20
3.4. Formulación y configuración de algoritmos	20
3.5. Entrenamiento y evaluación de las implementaciones	21
3.5.1. Entrenamiento de agentes	21

3.5.1.1.	Balaneo de péndulo	21
3.5.1.2.	Elevación del péndulo	22
3.5.2.	Evaluación de los agentes	22
3.5.2.1.	Balaneo del péndulo	23
3.5.2.2.	Elevación del péndulo	23
4.	Resultados	24
4.1.	Experimento 1: Balaneo del péndulo invertido	25
4.1.1.	Entrenamiento con base de datos generada por política experta y función de recompensa default	25
4.1.1.1.	Deep Q-Learning	25
4.1.1.2.	Deep Q-Learning from Demonstration	28
4.1.1.3.	Neural Fitted Q Iteration	30
4.1.1.4.	Conservative Q-Learning	31
4.1.1.5.	Implicit Q-Learning	33
4.1.1.6.	Batch-Constrained Deep Q-Learning	35
4.1.2.	Entrenamiento con base de datos generada por política experta y función de recompensa customizada	37
4.1.2.1.	Deep Q-Learning	37
4.1.2.2.	Deep Q-Learning from Demonstration	39
4.1.2.3.	Neural Fitted Q Iteration	41
4.1.2.4.	Conservative Q-Learning	42
4.1.2.5.	Implicit Q-Learning	44
4.1.2.6.	Batch-Constrained Deep Q-Learning	46
4.1.3.	Entrenamiento con base de datos generada por política mixta y función de recompensa default	48
4.1.3.1.	Deep Q-Learning	48
4.1.3.2.	Deep Q-Learning from Demonstration	50
4.1.3.3.	Neural Fitted Q Iteration	52
4.1.3.4.	Conservative Q-Learning	53
4.1.3.5.	Implicit Q-Learning	55
4.1.3.6.	Batch-Constrained Deep Q-Learning	57
4.1.4.	Entrenamiento con base de datos generada por política mixta y función de recompensa customizada	59
4.1.4.1.	Deep Q-Learning	59
4.1.4.2.	Deep Q-Learning from Demonstration	61
4.1.4.3.	Neural Fitted Q Iteration	63
4.1.4.4.	Conservative Q-Learning	64
4.1.4.5.	Implicit Q-Learning	66
4.1.4.6.	Batch-Constrained Deep Q-Learning	68
4.1.5.	Análisis y comparativa de algoritmos	70
4.2.	Experimento 2: Elevación del péndulo invertido	73
4.2.1.	Deep Q-Learning	73
4.2.2.	Deep Q-Learning from Demonstration	76
4.2.3.	Neural Fitted Q Iteration	78
4.2.4.	Conservative Q-Learning	79
4.2.5.	Implicit Q-Learning	81

4.2.6. Batch-Constrained Deep Q-Learning	83
4.2.7. Análisis y comparativa de algoritmos	85
5. Conclusiones	86
Bibliografía	88