COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI ICADE CIHS

GUÍA DOCENTE
2025 - 2026

FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura

Nombre completo	Aprendizaje por Refuerzo Profundo
Código	DEAC-MIA-523
Impartido en	Máster Universitario en Inteligencia Artificial [Primer Curso]
Nivel	Postgrado Oficial Master
Cuatrimestre	Semestral
Créditos	6,0 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Electrónica, Automática y Comunicaciones
Responsable	Lucía Güitta López
Horario de tutorías	Contactar por email con la profesora.

Datos del profesorado

Profesor

Nombre	Lucía Güitta López
Departamento / Área	Departamento de Electrónica, Automática y Comunicaciones
Despacho	AA25 - D507
Correo electrónico	lguitta@comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura

Aportación al perfil profesional de la titulación

El aprendizaje por refuerzo profundo (Deep Reinforcement Learning, DRL) combina los principios del aprendizaje por refuerzo tradicional con el poder representacional de las redes neuronales profundas, permitiendo abordar problemas de toma de decisiones secuenciales en entornos complejos, de alta dimensionalidad y parcialmente observables. Esta asignatura proporciona al estudiante una comprensión avanzada de cómo diseñar y entrenar agentes inteligentes capaces de aprender directamente de datos sensoriales brutos mediante la interacción con su entorno. La integración de técnicas de deep learning con estrategias de exploración, políticas estocásticas y aprendizaje fuera de línea ofrece una base sólida para aplicar soluciones en áreas como la robótica autónoma, los sistemas de recomendación adaptativos, los videojuegos, la automatización industrial y la toma de decisiones estratégicas. Al finalizar la asignatura, el estudiante, gracias a la combinación de clases teóricas y sesiones prácticas, habrá desarrollado competencias clave en el uso de herramientas actuales para implementar algoritmos de DRL, analizando su comportamiento e idoneidad para diferentes escenarios prácticos. Esto contribuye directamente a su perfil profesional como experto en inteligencia artificial, capacitándolo para abordar desafíos reales que requieren sistemas inteligentes con capacidad de aprendizaje autónomo y adaptativo en entornos dinámicos.

Prerrequisitos

Fundamentos de programación en Python y familiaridad en el uso de librerías como PyTorch.

Competencias - Objetivos

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

Teoría

Tema 1: Repaso de aprendizaje por refuerzo

- Procesos de decisión markovianos: ecuaciones de Bellman.
- Programación dinámica: ecuaciones de optimalidad de Bellman, y proceso de mejora general de la política.
- Aproximaciones sin modelo: Monte Carlo, diferencia temporal, y aprendizaje *on-policy* y *off-policy*.
- Representación aproximada del entorno: métodos no lineales, aproximación directa de la política y métodos *actor-critic*.

Tema 2: Aprendizaje de representación

- Motivación y fundamentos del aprendizaje de representación.
- Objetivos del aprendizaje de representación.
- Propiedades de una buena representación.
- Tipos de representaciones.
- Métodos para el aprendizaje.
- Casos de estudio y aplicaciones.

Tema 3: Aproximación de la función de valor y de valor de la acción

- Repaso de los métodos de aproximación lineal.
- Aproximación de la función de valor con redes neuronales.
- La triada mortal y la estabilización del aprendizaje.
- *Deep Q-Network* (DQN).
- *Double DQN*, *Prioritized Experience Replay*, *Dueling DQN* y *Multi-step bootstrapping*.
- Casos de estudio y aplicaciones.

Tema 4: Aproximación de la política

- Introducción a métodos basados en la aproximación de la política.
- Algoritmos básicos de gradiente de la política: *REINFORCE*.
- Estrategias de reducción de la varianza.
- Métodos *actor-critic*.
- Estrategias de optimización y mejora del rendimiento.
- Casos de estudio y aplicaciones.

Tema 5: Aprendizaje por refuerzo basado en el modelo (Planificación)

- Introducción al aprendizaje por refuerzo basado en el modelo del entorno: *model-based* vs *model-free*, modelos deterministas vs modelos estocásticos.
- Componentes: modelo, planificación, aprendizaje y actualización del modelo, y ejecución de la política.
- Planificación: algoritmos clásicos y algoritmos de búsqueda.
- *Self-Play* y *Fictitious Play* (modelado y exploración).
- Aprendizaje por refuerzo profundo basado en el modelo del entorno.
- Retos, problemas y aplicaciones.

Tema 6: Aprendizaje por imitación y aprendizaje por refuerzo inverso

- Motivación y fundamentos del aprendizaje de imitación.
- *Behavioral Cloning* y variantes.
- Aprendizaje por refuerzo inverso: objetivo, supuestos, algoritmos clásicos y extensiones.
- Aprendizaje de imitación adversario y variantes.
- Imitación a partir de la observación.
- Casos de estudio y aplicaciones.

Tema 7: Aprendizaje de la recompensa y aprendizaje condicionado por objetivos

- Introducción al diseño de funciones de recompensa.
- Aprendizaje de la recompensa.
- Consideraciones del aprendizaje de la recompensa.
- Aprendizaje basado en objetivos (*Goal-Conditioned RL*).
- Casos de estudio y aplicaciones.

Tema 8: Transferencia del aprendizaje

- Introducción a la problemática de la transferencia del aprendizaje.
- Métodos de transferencia del aprendizaje.
- Aleatorización del dominio.
- Adaptación del dominio.
- Redes neuronales progresivas.
- Casos de estudio y aplicaciones.

Tema 9: Tendencias en el aprendizaje por refuerzo

- Aprendizaje por refuerzo a partir de retroalimentación humana.
- Meta aprendizaje por refuerzo.
- Aprendizaje por refuerzo jerárquico y multi agente.

Laboratorio

Práctica 1: Aprendizaje de la representación del estado

En esta práctica se realizarán diversas tareas relacionadas con el aprendizaje de representación del estado, implementando distintos entornos y algoritmos, con el fin de obtener una comparativa entre los diferentes métodos introducidos en las sesiones teóricas.

Práctica 2: Aproximación de la función de valor y de valor de la acción

En esta práctica se implementarán distintos algoritmos de los propuestos en las sesiones teóricas para estimar las funciones de valor y valor de la acción en entornos cuya representación de estado deba ser aprendida, con el fin de entrenar distintos agentes y comparar su desempeño.

Práctica 3: Aproximación de la política

En esta práctica se implementarán distintos algoritmos de los propuestos en las sesiones teóricas para aproximar directamente la política en entornos cuya representación de estado deba ser aprendida, con el fin de entrenar distintos agentes y comparar su desempeño.

Proyecto final

Como proyecto final del laboratorio se propondrán dos alternativas:

1. La elección de un entorno virtual de la librería de Gymnasium y el entrenamiento de un agente mediante algún algoritmo de aprendizaje por refuerzo que se considere estado del arte y que no haya sido visto en clase (por lo que el alumno tendrá que hacer un trabajo de investigación previo y deberá estar justificada la elección).
2. El diseño de un entorno virtual propio, basado en la estructura que proporciona cualquier entorno de la librería Gymnasium, y el entrenamiento de un agente mediante algún algoritmo de aprendizaje por refuerzo de los que se hayan visto en las sesiones teóricas pero que no se hayan empleado en las prácticas anteriores (el uso de un algoritmo de las prácticas anteriores que no esté justificado repercutirá negativamente en la nota).

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Calificaciones

Convocatoria ordinaria

El peso de cada una de las actividades de evaluación es el siguiente:

- Teoría (50%)
  - Pruebas de clase: 5%
  - Prueba intersemestral: 10%
  - Examen final: 35%
- Laboratorio (50%)
  - Prácticas: 15% (10% informe de la práctica y 5% la prueba inicial)
  - Proyecto final: 30%
  - Presentación del proyecto final: 5%

Convocatoria extraordinaria

En la convocatoria extraordinaria se realizarán el examen de teoría y/o laboratorio suspensos, conservándose el resto de calificaciones para obtener la nota final.

- La calificación final de la parte teórica se obtendrá de la misma forma que en la convocatoria ordinaria, y atendiendo por tanto a las mismas restricciones.
- En el caso de laboratorio, se sustituirá las notas del proyecto final y la presentación del mismo por la nota obtenida en el examen extraordinario de laboratorio.

Normativa sobre el uso de la IA

En las actividades de laboratorio (incluyendo la elaboración de informes) y en las actividades grupales se permite el uso de IA bajo las siguientes restricciones:

- La IA puede utilizarse para la preparación de las tareas, la investigación inicial necesaria y como ayuda en el desarrollo del código, pero la evaluación tendrá en cuenta la demostración de la capacidad del alumno de, partiendo de ese punto, refinar o elaborar propuestas propias y argumentarlas correctamente.

- En la elaboración de informes la IA puede ser usada para mejorar la redacción o pedir retroalimentación, pero el alumno debe evaluar y modificar de forma crítica los resultados ofrecidos por la IA, demostrando la comprensión de los conceptos tratados.
- En cualquier caso, el uso de IA debe estar citado correctamente en las prácticas y/o informes.

El uso total e indiscriminado de la IA por parte del alumno para elaborar las prácticas e informes queda prohibido. De ser descubierto se puntuará como un 0 en la práctica completa.

En las pruebas de clase y el examen final correspondientes a las sesiones de teoría, el uso de la IA está completamente prohibido. De ser descubierto se puntuará como un 0 en el global de las pruebas de clase o en el examen final.

BIBLIOGRAFÍA Y RECURSOS