

Guía docente

230817 - ARAP - Aprendizaje por Refuerzo y Aprendizaje Profundo

Última modificación: 24/03/2026

Unidad responsable:	Escuela Técnica Superior de Ingeniería de Telecomunicación de Barcelona		
Unidad que imparte:	739 - TSC - Departamento de Teoría de la Señal y Comunicaciones.		
Titulación:	GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE TELECOMUNICACIÓN (Plan 2015). (Asignatura optativa). GRADO EN CIENCIA E INGENIERÍA DE DATOS (Plan 2017). (Asignatura optativa). GRADO EN INGENIERÍA ELECTRÓNICA DE TELECOMUNICACIÓN (Plan 2018). (Asignatura optativa).		
Curso: 2026	Créditos ECTS: 6.0	Idiomas: Catalán, Inglés	

PROFESORADO

Profesorado responsable:	JOSE VIDAL MANZANO
Otros:	Primer quadrimestre: ANTONIO JESUS BONAFONTE CAVEZ - 11, 13 MARGARITA ASUNCION CABRERA BEAN - 11, 13 JOSE VIDAL MANZANO - 11, 13

CAPACIDADES PREVIAS

Algebra, Probabilidad y procesos, Señales y sistemas, Aprendizaje automático y profundo.

REQUISITOS

Para los estudiantes de GRETST Y GREELEC, el prerequisite es Introducción al aprendizaje profundo (IDL 230325). Para los estudiantes del GCED, el requisito es Aprendizaje Automático 2 (AA2 270222).

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Genéricas:

08 CRPE. CAPACIDAD PARA IDENTIFICAR, FORMULAR Y RESOLVER PROBLEMAS DE INGENIERÍA. Capacidad para plantear y resolver problemas de ingeniería en el ámbito TIC con iniciativa, toma de decisiones y creatividad. Desarrollar un método de análisis y solución de problemas sistemático y creativo

Transversales:

04 COE N2. COMUNICACIÓN EFICAZ ORAL Y ESCRITA - Nivel 2: Utilizar estrategias para preparar y llevar a cabo las presentaciones orales y redactar textos y documentos con un contenido coherente, una estructura y un estilo adecuados y un buen nivel ortográfico y gramatical.

06 URI N1. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN - Nivel 1: Identificar las propias necesidades de información y utilizar las colecciones, los espacios y los servicios disponibles para diseñar y ejecutar búsquedas simples adecuadas al ámbito temático.

06 URI N2. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN - Nivel 2: Después de identificar las diferentes partes de un documento académico y de organizar las referencias bibliográficas, diseñar y ejecutar una buena estrategia de búsqueda avanzada con recursos de información especializados, seleccionando la información pertinente teniendo en cuenta criterios de relevancia y calidad.

06 URI N3. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN - Nivel 3: Planificar y utilizar la información necesaria para un trabajo académico (por ejemplo, para el trabajo de fin de grado) a partir de una reflexión crítica sobre los recursos de información utilizados.

METODOLOGÍAS DOCENTES

Clases teóricas y laboratorio

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

Dominar los principios del aprendizaje por refuerzo como herramienta de inteligencia artificial basada en la interacción de la máquina con su entorno, con aplicaciones en tareas de control (ej. robótica, vehículos autónomos) y toma de decisiones (ej. optimización de recursos en redes de comunicaciones sin hilos).

Diseñar y entrenar redes neuronales profundas que capaces de aprender con poca o nula supervisión, tanto en tareas discriminativas como generativas, con especial énfasis en las aplicaciones multimedia (visión, lenguaje y audio).

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTE

Tipo	Horas	Porcentaje
Horas grupo pequeño	26,0	17.33
Horas grupo grande	26,0	17.33
Horas aprendizaje autónomo	98,0	65.33

Dedicación total: 150 h

CONTENIDOS

1. Introducción al aprendizaje por refuerzo

Descripción:

Describir con ejemplos los conceptos fundamentales y los problemas que seremos capaces de resolver.

Dedicación: 2h

Grupo grande/Teoría: 2h

2. El compromiso exploración-explotación

Descripción:

- El compromiso exploración-explotación
- Caso de estudio: La sala de máquinas tragaperras

Dedicación: 1h

Grupo grande/Teoría: 1h

3. Procesos de decisión de Markov (MDP)

Descripción:

- El interfaz agente-entorno
- Objetivos y recompensa
- Procesos de decisión de Markov
- Función de valor y optimalidad: la ecuación de Bellman

Dedicación: 2h

Grupo grande/Teoría: 2h

4. Programación dinámica

Descripción:

- Evaluación de políticas, mejoras e iteraciones
- Programación dinámica basada en MDP

Dedicación: 1h

Grupo grande/Teoría: 1h

5. Métodos de Monte-Carlo

Descripción:

- Aprendizaje sin modelos
- Métodos de Montecarlo de primera visita
- Métodos de Montecarlo de cada visita
- Exploración y explotación

Dedicación: 2h 30m

Grupo grande/Teoría: 2h 30m

6. Aprendizaje por diferencia temporal (TD)

Descripción:

- Aprendizaje sin modelos usando diferencias temporales
- SARSA y Q-learning con acciones discretas
- Juegos

Dedicación: 2h 30m

Grupo grande/Teoría: 2h 30m

7. Gradientes sobre las políticas

Descripción:

- Gradientes sobre las políticas
- Aproximación del valor de las funciones
- Métodos Actor-Crítico
- Funciones baseline

Dedicación: 2h 30m

Grupo grande/Teoría: 2h 30m

8. Aprendizaje profundo por refuerzo

Descripción:

Modelado de funciones Q y de las políticas con redes neuronales profundas

Dedicación: 2h 30m

Grupo grande/Teoría: 2h 30m

9. Aprendizaje profundo avanzado

Descripción:

- Modelos generativos: GANs, VAEs and Flows.
- Redes Neuronales Recurrentes (RNNs) avanzadas.
- Mecanismos de atención y Transformers.
- Redes Neuronales con Grafos.
- Aprendizaje auto-supervisado y meta-aprendizaje.
- Supercomputación para aprendizaje profundo.

Dedicación: 10h

Grupo grande/Teoría: 10h

Laboratoris de aprendizaje por refuerzo

Descripción:

Laboratorios en Matlab u/o Python que se distribuirán a lo largo de los temas

- Problema de máquina tragaperras
- MDP o robot reciclador
- MDP de gridworlds
- Aprendizaje-Q con tablas
- Aprendizaje con redes neuronales

Dedicación: 13h

Grupo pequeño/Laboratorio: 13h

Laboratorios de aprendizaje profundo

Descripción:

Laboratorios en PyTorch sobre:

- Aprendizaje-Q con redes neuronales
- Políticas de gradiente con redes neuronales (REINFORCE)
- Optimizadores para de redes neuronales profundas
- Modelos de atención
- Transferencia del conocimiento
- Redes Adversarias Generativas (GAN)
- Auto-codificadores variacionales (VAE)

Dedicación: 13h

Grupo pequeño/Laboratorio: 13h

SISTEMA DE CALIFICACIÓN

Cualquier acto de fraude académico, plagio o uso o mera tenencia al alcance de medios no autorizados en cualquier actividad de evaluación comportará la calificación de cero (0) en la prueba o entrega afectada. Además, de acuerdo con la normativa de la Universidad, la posible derivación de los hechos para la apertura de un expediente disciplinario implicará que la asignatura quede en el estado provisional de "pendiente de evaluación" hasta la resolución del expediente. La gestión de estas incidencias se lleva a cabo de acuerdo con el [Marco de actuación para la integridad académica en la evaluación de la UPC](#).

Exámenes y evaluación durante las sesiones de prácticas.

BIBLIOGRAFÍA

Básica:

- Sutton, R.S.; Barto, A.G. Reinforcement learning: an introduction [en línea]. 2nd ed. Cambridge, Mass.: MIT Press, 2018 [Consulta: 14/09/2020]. Disponible a: <https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=6260249>. ISBN 9780262039246.
- Goodfellow, I.; Bengio, Y.; Courville, A. Deep learning [en línea]. Cambridge, Massachusetts: The MIT Press, [2016] [Consulta: 19/10/2022]. Disponible a: <http://www.deeplearningbook.org/>. ISBN 978-0262035613.

RECURSOS

Otros recursos:

Contenidos actualizados en Atenea