

**Formulario de aprobación de curso de posgrado/educación permanente**

**Asignatura:** Taller de Aprendizaje por Refuerzo  
(Si el nombre contiene siglas deberán ser aclaradas)

<b>Modalidad:</b> (posgrado, educación permanente o ambas)	<b>Posgrado</b> <input checked="" type="checkbox"/>
	<b>Educación permanente</b> <input checked="" type="checkbox"/>

**Profesor de la asignatura**<sup>1</sup>: Isabel Amigo (gr3, IIE)  
(título, nombre, grado o cargo, instituto o institución)

**Profesor Responsable Local**<sup>1</sup>:  
(título, nombre, grado, instituto)

**Otros docentes de la Facultad:** Juan Bazerque (gr3,IIE)  
(título, nombre, grado, instituto)

**Docentes fuera de Facultad:**  
(título, nombre, cargo, institución, país)

<sup>1</sup> Agregar CV si el curso se dicta por primera vez.  
(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

[Si es curso de posgrado]

**Programa(s) de posgrado:** Maestría en Ingeniería Matemática (IngeMat), Maestría en Ciencia de Datos (MCDA) y Aprendizaje Automático, Maestría y Doctorado en Ingeniería Eléctrica.

**Instituto o unidad:** IIE

**Departamento o área:**

**Horas Presenciales:** 50

(se deberán discriminar las horas en el ítem Metodología de enseñanza)

**Nº de Créditos:** 8

[Exclusivamente para curso de posgrado]  
(de acuerdo a la definición de la UdelaR, un crédito equivale a 15 horas de dedicación del estudiante según se detalla en el ítem Metodología de enseñanza)

**Público objetivo:** Estudiantes de posgrado (y eventualmente estudiantes avanzados de grado) con formación básica en álgebra lineal y probabilidad y cuyos intereses estén relacionados con el modelado y control de sistemas complejos desde una perspectiva de aprendizaje automático. El curso espera llegar a estudiantes de programas de la Facultad de Ingeniería (IE, IngeMat, MCDA) pero también a estudiantes de posgrado por fuera de la facultad como la maestría en Informática del PEDECIBA.

**Cupos:** No tiene cupos

(si corresponde, se indicará el número de plazas, mínimo y máximo y los criterios de selección. Asimismo, se adjuntará en nota aparte los fundamentos de los cupos propuestos. Si no existe indicación particular para el cupo máximo, el criterio general será el orden de inscripción, hasta completar el cupo asignado)

**Objetivos:** Aprender qué son los procesos de decisión markoviana (MDP) y cómo se utilizan para modelar problemas de aprendizaje por refuerzo. Aprender los métodos básicos de resolución de MDPs y algoritmos de resolución de problemas de Aprendizaje por Refuerzo. Familiarizarse con la programación de dichos algoritmos y con la utilización de librerías de software modernas para su resolución.

**Conocimientos previos exigidos:** Álgebra de matrices. Fundamentos de probabilidad y estadística. Programación a nivel básico.

**Conocimientos previos recomendados:** Programación en Python científico.

**Metodología de enseñanza:**

(comprende una descripción de la metodología de enseñanza y de las horas dedicadas por el estudiante a la asignatura, distribuidas en horas presenciales -de clase práctica, teórico, laboratorio, consulta, etc.- y no presenciales de trabajo personal del estudiante)

Descripción de la metodología:

El curso consta de una clase teórica presencial durante 12 semanas además de 5 laboratorios asociados a los temas del curso, y la lectura y presentación por parte de los estudiantes de un artículo en la materia.

Los laboratorios incluyen un taller práctico en Python que busca ilustrar algunos de los conceptos vistos en clase. Éstos serán realizados por los estudiantes y servirán sobre todo para auto-estudio y preparación del examen final. El objetivo es que luego de realizar el laboratorio el estudiante sea capaz de modelar y programar problemas bajo la forma de un MDP y como un problema de RL, y resolver los mismos a través de la programación de algoritmos y de la utilización de bibliotecas.

Detalle de horas:

- Horas de clase (teórico): 30
- Horas de clase (práctico): 0
- Horas de clase (laboratorio): 0
- Horas de consulta: 16
- Horas de evaluación: 4
  - Subtotal de horas presenciales: 50
- Horas de estudio: 20
- Horas de resolución de ejercicios/prácticos: 40
- Horas proyecto final/monografía: 10
  - Total de horas de dedicación del estudiante: 120

**Forma de evaluación:**

Examen final presencial.

Entrega de ejercicios prácticos y/o en el lenguaje de programación Python junto con un examen final y la lectura y presentación oral de un artículo de investigación en el área.

Una vez concluido el componente instructivo del curso y realizada la entrega de ejercicios, los estudiantes deberán seleccionar un artículo de investigación relacionado con el material tratado en clase, y preparar una presentación oral para presentar ante grupo y a los docentes.

Luego se realizará un examen final en forma presencial.

---

**Temario:** 1. Introducción y motivación

2. MDPs

- a) Value function
- b) Ecuaciones de Bellman
- c) Value iteration
- d) policy iteration

3. Aprendizaje por refuerzo

4. Algoritmos de resolución

- a) TD(0)
- b) SARSA
- c) Q-learning

5. Aproximación de función

6. Gradiente estocástico

6. Deep RL

7. Algoritmos Actor-Crítico

8- Proximal Policy Optimization.

---

**Bibliografía:**

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

[1] Richard S. Sutton and Andrew G. Barto *Reinforcement Learning: An Introduction*, Second Edition, MIT Press, Cambridge, MA, 2018.

[2] Warren B. Powell, *Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions*, Wiley, 2022.

[3] Csaba Szepesvári, *Algorithms for Reinforcement Learning*, Morgan and Claypool Publishers, 2010.

[4] Dimitri Bertsekas, *Reinforcement Learning and Optimal Control*, Athena Scientific, 2019.

[5] Miguel Morales, Miguel Morales *Grokking Deep Reinforcement Learning*, Manning, 2020.

[6] Maxim Lapan, *Deep Reinforcement Learning Hands-On (2a edición)*, Packt Publishing, 2020.

---

**Datos del curso**

---

**Fecha de inicio y finalización:** 1/04/2026 - 30/06/2026

**Horario y Salón:** a determinar con los estudiantes

**Arancel:**

[Si la modalidad no corresponde indique “no corresponde”. Si el curso contempla otorgar becas, indíquelo]

**Arancel para estudiantes inscriptos en la modalidad posgrado: no corresponde**

**Arancel para estudiantes inscriptos en la modalidad educación permanente: 4000 UI**

**El curso contempla otorgar becas a solicitud de los estudiantes.**

---