PA602

FACULTAD DE INGENIERIA

Facultad de Ingeniería
Comisión Académica de Posgrado

Formulario de aprobación de curso de posgrado/educación permanente

Asignatura: APRENDIZAJE POR RECOMPENSAS

(Si el nombre contiene siglas deberán ser aclaradas)

Modalidad:
(posgrado, educación permanente o ambas)

Posgrado [X]

Educación permanente

Profesor de la asignatura [1]: Dr Juan Bazerque, Gr.3, IIE.
(título, nombre, grado o cargo, instituto o institución)

Profesor Responsable Local [1]: No
(título, nombre, grado, instituto)

Otros docentes de la Facultad: Dr José Lezama, Gr.3, IIE
(título, nombre, grado, instituto)

Docentes fuera de Facultad: No
(título, nombre, cargo, institución, país)

[1] 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 y doctorado en Ingeniería Eléctrica,

Instituto o unidad: Instituto de Ingeniería Eléctrica

Departamento o área: Departamento de Control / Departamento de procesamiento de señales

Horas Presenciales: 50 horas presenciales.
(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 posgrados de ingeniería.

Cupos: mínimo 6 estudiantes, máximo 40 estudiantes
(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: El principal objetivo de este curso es llevar al estudiante al estado del arte en técnicas de reinforcement learning, basándonos especialmente en la referencia [1].

Conocimientos previos exigidos: Conocimientos básicos de álgebra y cálculo probabilidad y programación.

Conocimientos previos recomendados: Conocimientos de optimización y procesos estocásticos, y nociones básicas de control

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: La metodología será la impartición de clases magistrales por el profesorado de forma online o presenciales, que se complementarán con trabajos llevados a cabo por los estudiantes, con cuatro tareas de análisis y programación que se publicarán a lo largo del semestre.

Detalle de horas:

- Horas de clase (teórico): 42 horas
- Horas de clase (práctico): 0 horas
- Horas de clase (laboratorio): 0 horas
- Horas de consulta: 6 horas
- Horas de evaluación: 2
    - Subtotal de horas presenciales: 50 horas
- Horas de estudio: 30 horas
- Horas de resolución de ejercicios/prácticos: 40 horas
- Horas proyecto final/monografía: No aplica
    - Total de horas de dedicación del estudiante: 120

Forma de evaluación:
[Indique la forma de evaluación para estudiantes de posgrado, si corresponde]. Entrega de 4 tareas de análisis y programación durante el semestre, más un examen final, que será oral si la cantidad de estudiantes es menor a 10.

[Indique la forma de evaluación para estudiantes de educación permanente, si corresponde]. No corresponde

Temario:

[1] Introducción

Definición del problema de reinforcement learning
Aprendizaje por recompensas
Concepto de realimentación con el medio
Ejemplos

[2] Multi-arm bandits

Descripción
Función de valor
Concepto de regret
Métodos incrementales

[3] Markov Decision Processes

Estructura de un Markov Decision Process(MDP)
Objetivos, recompensas, episodios y políticas
Ecuaciones de Bellman
Control de un MDP
V-Function and Q-Function
Condiciones de optimalidad

[4] Approximación de funciones

Aprendizaje en espacios de dimensiones altas
Expansión paramétrica de la política
Aproximaciones lineales o no lineales
Redes neuronales

[5] Policy Gradient

Descenso por gradiente estocástcio
Policy Gradient Theorem
Garantías de convergencia
REINFORCE

[6] Improving the Policy Gradient

On-policy and Off-policy
Policy Gradient determinísitico
Métodos de reducción de la varianza
Bootstrapping y TD-Learning

[7] Actor-Critic Policy Gradient

Q-Function - estimación
Métodos SARSA y Q-Learning
Métodos Actor-Crítico
Off-Policy Actor-Critic

[8] Métodos Modernos

Trust Region Methods
Trust Region Policy Optimization
Proximal Policy Optimization

[9]Investigación en curso sobre Reinforcement Learning

Multi-task Reinforcement Learning
Safe Reinforcement Learning
Reinforcement Learning en robótica

Bibliografía:

[1] Sutton, R. S., and Barto A. G., *Reinforcement learning: An introduction.* MIT press, 2018. Available online: http://webdocs.cs.ualberta.ca/sutton/book/the-book.html

[2] Szepesvari, C., *Algorithms for Reinforcement Learning*. Morgan & Claypool, 2010. Available online: https://sites.ualberta.ca/~szepesva/RLBook.html

[3] Altman, E. *Constrained Markov Decision Process*, volume 7. CRC Press, 1998. ISBN 9780849303821.
URL http://www-sop.inria.fr/members/Eitan.Altman/PAPERS/h.pdf.

Datos del curso

Fecha de inicio y finalización: **1° semestre 2021**

Horario y Salón: A determinar, en forma remota o presencial y en horario a determinar en dos clases semanales de hora y media.

Arancel:
[Si la modalidad no corresponde indique "no corresponde". Si el curso contempla otorgar becas, indíquelo]

Arancel para estudiantes inscriptos en la modalidad posgrado: 0
Arancel para estudiantes inscriptos en la modalidad educación permanente: no corresponde