

# Aprendizaje Reforzado e Inferencia Variacional

Julián Martínez, Sebastian Grynberg, Nahuel Soprano-Loto.

**Duración.** 12 semanas.

**Carga horaria.** 48 horas: 4 horas semanales, una vez por semana.

**Modalidad de enseñanza.** Clases teórico-prácticas. Se presentarán problemas, la teoría para abordarlos, y los algoritmos para resolverlos que serán implementados en el lenguaje Python.

Se priorizará el enfoque *probabilístico* a los efectos de comprender los elementos teóricos subyacentes a las técnicas utilizadas.

El lenguaje Python es de carácter libre y utilizado en diversas áreas del conocimiento. Este lenguaje es el de referencia en lo que respecta al Aprendizaje Reforzado.

**Modalidad de evaluación.** Entrega periódica de trabajos prácticos y trabajo final el cual deberá ser acompañado de una breve presentación.

**Objetivos.** Comprender las bases y las distintas componentes de los algoritmos de Aprendizaje Reforzado. Se espera que el estudiante desarrolle capacidades para:

- formular los problemas de aplicación en el contexto de los métodos expuestos, implementando computacionalmente posibles soluciones,
- evaluar mediante distintas métricas el desempeño de los modelos y algoritmos propuestos.

**Requisitos previos.** Probabilidad y Estadística, Álgebra lineal y experiencia en programación en algún lenguaje.

## Programa

1. Introducción. Definición del problema. Enumeración de aplicaciones. Relación con otros campos. Software utilizado.
2. Procesos de Decisión Markovianos. Señal de Recompensa. Políticas y Funciones de Valor. Métodos On-Policy vs Métodos Off-Policy. Aprendizaje Reforzado Model-based vs. Aprendizaje Reforzado Model-free.
3. Exploración y Explotación. Problema del Bandido Multibrazo. Métodos de solución. Aplicación a Procesos de Decisión Markovianos.
4. Programación Dinámica: Ecuación de Bellman. Iteración de Políticas y Funciones de Valor.
5. Métodos Tabulares: Predicción y Control Monte-Carlo. Método de Diferencias Temporales. Importance Sampling. Q-aprendizaje.
6. Revisión breve de Redes Neuronales: Backpropagation, Redes Fully Connected, Capas Convolucionales, Regularización. Método de Descenso por Gradiente Estocástico.

7. Integración de Aprendizaje y Planificación. Arquitecturas Integradas.
8. Métodos de Solución Aproximados: Aproximación de funciones. Métodos de Política Gradiente. Actor-Critic.
9. Casos de Estudio.

## Plantel Docente

Los profesores responsables para el dictado de esta materia serán:

- Prof. Julián Martínez. Phd in Mathematics (área Probabilidad y Mecánica Estadística).
- Prof. Sebastián Grynberg. Dr. en Ciencias Matemáticas (área Probabilidad y Mecánica Estadística).
- Jtp. Nahuel Soprano-Loto. Dr. en Ciencias Matemáticas (área Probabilidad y Mecánica Estadística).

El 1er cuatrimestre del 2020 esta asignatura será dictada por el Prof. Julián Martínez.

## Bibliografía

- [Ber16] Dimitri P Bertsekas. Dynamic programming and optimal control 4th edition. 2016.
- [Pow07] Warren B Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.
- [Put14] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
- [SB15] Richard S Sutton and Andrew G Barto. Reinforcement learning an introduction—second edition, in progress (draft), 2015.
- [Sze09] Csaba Szepesvári. Algorithms for reinforcement learning. *Morgan and Claypool*, 2009.