

# Pautas de elaboración y presentación del proyecto

## Elaboración del proyecto

Para el proyecto final de la asignatura se pide diseñar, entrenar y evaluar un agente que realice correctamente las tareas definidas para cada uno de los tres niveles del entorno del almacén. Es obligatorio **resolver los entornos en orden creciente de dificultad**, es decir, no se puede resolver directamente el entorno 3, primero hay que entrenar y evaluar el agente en el entorno 1, luego en el entorno 2, y finalmente en el 3. Esto no quiere decir que haya que usar distintos algoritmos. El **algoritmo puede ser el mismo**, por ejemplo, la DQN, pero muy probablemente los hiperparámetros, la duración del entrenamiento, y evidentemente, el diseño de la recompensa será distinto.

Por otro lado, queda **prohibido presentar la solución como un ensamblado de agentes**, es decir, usar dos agentes para resolver un mismo entorno, por ejemplo, un conjunto de pesos que define a un agente para llegar al objeto y cogerlo, y otro conjunto distinto de pesos que define a otro agente para llevar el objeto al área de entrega.

Por lo tanto, lo que se espera de cada grupo de cara a la elaboración del proyecto es:

- Diseño de la **representación de estado** (el tile-coding solo no es suficiente para los nuevos entornos). **Es necesario hacer ingeniería de variables, no puede usarse la imagen del entorno como entrada** para el agente (eso es parte de la asignatura de aprendizaje por refuerzo profundo).
- **Diseño de la recompensa** de acuerdo con la tarea que debe realizar el agente en cada entorno (los entornos 2 y 3 resuelven la misma tarea, sólo que en el caso 3 el agente debe generalizar su conocimiento para ser capaz de coger objetos cuya posición varía entre episodios).
- **Diseño, entrenamiento y evaluación del algoritmo de aprendizaje del agente** (se pueden usar agentes ya diseñados como los disponibles en la librería de Python **stable-baselines3**). Se **recomienda usar una aproximación no lineal** para las representaciones de las funciones de valor. Se **valorará positivamente** usar **conceptos vistos en clase como elegibility traces, experience replay, etc.** No es necesario usar algoritmos de *policy gradient*.

## Presentación del proyecto

El proyecto se evaluará principalmente en base a la **presentación grupal** que se realizará en la última sesión de clase. La presentación tendrá una **duración máxima de 8 minutos**, y luego habrá un **turno de preguntas individual** a cada miembro del grupo.

Las secciones que se evaluarán en esta presentación son las siguientes:

- **Introducción** (muy breve ya que el límite de tiempo es muy estricto)
- Breve explicación de los **resultados en la sesión 1** (MDP, hiperparámetros, gráficas del proceso de aprendizaje, y porcentaje de éxito en evaluación). Es opcional.
- **Metodología** seguida para la resolución de los entornos de la sesión 2.
- Definición del **MDP e hiperparámetros** empleados.
- Resultados del **proceso de entrenamiento** (métricas y gráficas).
- Resultados del **proceso de evaluación** (métricas y vídeos de funcionamiento).
- **Conclusiones** (principales resultados, principales dificultades encontradas, posibles mejoras para mejorar los resultados).

Ejemplos de métricas interesantes son: la recompensa media, la duración media de los episodios, el porcentaje de éxito y/o choques.

Además de la presentación, se deberá **entregar un informe** donde se hayan recogido los diseños y resultados de las dos sesiones de trabajo que se consultará en caso de que haya algún resultado inconsistente en la presentación.