

Universidad del Valle de Guatemala  
Facultad de Ingeniería



DeepLearning  
Introducción al aprendizaje por refuerzo - Cartpole DQN  
Informe

Mariana David 201055  
Pablo Escobar 20936

Guatemala 23, de noviembre del 2023

## **Introducción**

Este proyecto busca aplicar Aprendizaje por Refuerzo con una Red Neuronal al desafío de CartPole. La motivación radica en poner en práctica los conocimientos adquiridos y abordar un problema desafiante. Utilizando Gym de OpenAI, el modelo elegido alcanzó un puntaje de 683 tras 176 iteraciones. Para probarlo, lo dejamos en una prueba de 15 minutos. El agente mantuvo un rendimiento sólido, evidenciando la eficacia de la metodología implementada. La red neuronal, diseñada con capas densas y funciones ReLU, junto con Reinforcement learning DQN, ha demostrado su capacidad para aprender de manera autónoma y destacar en entornos complejos como el de CartPole. Estos resultados no solo resumen los logros obtenidos sino también sientan las bases para futuras investigaciones en inteligencia artificial y aprendizaje autónomo.

## **Antecedentes**

El proyecto se basa en la aplicación de Aprendizaje por Refuerzo (RL) en el entorno CartPole, un problema clásico de la biblioteca OpenAI Gym. CartPole implica equilibrar un poste sobre un carro en movimiento, con dos acciones posibles: mover el carro hacia la izquierda o derecha. La esencia del problema es entrenar a un agente para tomar decisiones que prolonguen la estabilidad del sistema.

## **Dataset y Variables**

El "dataset" utilizado en este proyecto se compone de las interacciones del agente con el entorno CartPole. Las variables clave incluyen la observación del estado del entorno, las acciones tomadas por el agente, las recompensas asociadas a esas acciones y la información sobre si el episodio ha concluido. El estado del entorno es un vector que representa la posición y velocidad del carro, así como el ángulo y velocidad angular del poste.

## **Tecnología Utilizada**

La implementación tecnológica se centra en el uso de Aprendizaje Profundo y Redes Neuronales, específicamente la biblioteca Keras. Se emplea el algoritmo Deep

Q-Network (DQN) para entrenar al agente. DQN es conocido por su capacidad para manejar problemas complejos de RL al aproximar la función Q mediante una red neuronal profunda.

### **Análisis Exploratorio**

Durante las primeras iteraciones del entrenamiento, el agente enfrentó desafíos para mantener el equilibrio en CartPole. Sin embargo, después de varios intentos logró superar estas dificultades, alcanzando un puntaje de 683. Este análisis exploratorio destaca la progresión del agente en el aprendizaje del entorno.

Se realizó una prueba prolongada de 15 minutos donde el agente continuó jugando correctamente sin perder el equilibrio, indicando una capacidad sólida y consistente. Estos resultados respaldan la efectividad de la implementación de RL y la capacidad de la red neuronal para aprender comportamientos complejos en CartPole. En resumen, el análisis exploratorio refleja el éxito del enfoque adoptado y sienta las bases para una comprensión más profunda de la adaptabilidad del agente en entornos dinámicos.

### **Metodología**

La metodología empleada se centra en la implementación de Aprendizaje por Refuerzo (RL) mediante una Red Neuronal, utilizando la biblioteca Keras. El algoritmo elegido es Deep Q-Network (DQN), conocido por su eficacia en problemas de RL complejos. A continuación, se describen los elementos clave de la implementación:

1. Red Neuronal (create\_DQN\_model):
  - a. La arquitectura de la red neuronal comprende cuatro capas densas con funciones de activación ReLU.
  - b. Capa 1: 512 nodos con inicialización de pesos he\_uniform.
  - c. Capa 2: 256 nodos con activación ReLU y pesos he\_uniform.

- d. Capa 3: 64 nodos con activación ReLU y pesos he\_uniform.
- e. Capa de salida: Nodos igual al espacio de acciones con activación lineal para poder elegir una acción Q y pesos he\_uniform.
- f. La función de pérdida se estableció como "mse" (error cuadrático medio) y se utilizó el optimizador RMSprop con una tasa de aprendizaje de 0.00025, rho de 0.95 y épsilon de 0.01.

## 2. Parámetros del Agente DQNAgent

- a. El entorno seleccionado es "CartPole-v1" de Gym.
- b. El factor de descuento (gamma) se fijó en 0.95.
- c. La tasa de exploración inicial (epsilon) se estableció en 1.0, con un mínimo de 0.001 y un factor de decaimiento de 0.999.
- d. El tamaño de lote para el entrenamiento de la red neuronal es 64.
- e. La memoria de repetición (replay memory) tiene una capacidad máxima de 2000.
- f. Se realizaron un máximo de 1000 juegos.

## 3. Proceso de Entrenamiento

- a. Durante cada episodio de entrenamiento, el agente interactúa con el entorno, recopilando experiencias.
- b. Las experiencias se almacenan en una memoria de repetición.
- c. La red neuronal se entrena periódicamente utilizando muestras aleatorias de la memoria.
- d. Se aplica la ecuación DQN estándar para ajustar los valores Q y mejorar la toma de decisiones del agente.

## 4. Prueba del Agente (test)

- a. Se realiza una prueba utilizando el modelo entrenado durante 15 minutos.
- b. El agente toma decisiones basadas en la acción con el valor Q más alto.

- c. Se evalúa la capacidad del agente para mantener un rendimiento sólido durante este período.

Esta metodología proporciona una estructura sólida para el entrenamiento y evaluación del agente en el entorno CartPole, aprovechando las capacidades de la red neuronal y el algoritmo DQN para aprender comportamientos efectivos.

## **Resultados**

El entrenamiento del agente utilizando Aprendizaje por Refuerzo (RL) y una Red Neuronal (RN) ha demostrado ser altamente exitoso en el entorno CartPole. A continuación, se presentan los resultados clave obtenidos durante el proceso de entrenamiento y las pruebas realizadas:

### **1. Desarrollo del Agente**

- a. Después de aproximadamente 176 iteraciones de entrenamiento, el agente logró superar las dificultades iniciales y jugar de manera efectiva en CartPole.
- b. El puntaje alcanzado por el agente fue impresionante, llegando a 683 en una sola iteración.

### **2. Prueba Continuada**

- a. Se llevó a cabo una prueba prolongada durante 15 minutos, donde el agente continuó jugando correctamente sin perder el equilibrio.
- b. Este resultado destaca la capacidad del agente para mantener un rendimiento sólido y consistente incluso durante sesiones de juego prolongadas.

### **3. Evolución del Puntaje**

- a. Se observó una mejora constante en el rendimiento del agente a lo largo de las iteraciones, evidenciando la capacidad de aprendizaje de la RN y el algoritmo DQN.

- b. Gráficos de la evolución del puntaje a lo largo del tiempo respaldan esta progresión ascendente (Figura 1).

#### 4. Estabilidad del Agente

- a. Durante la prueba prolongada, el agente demostró una notable estabilidad, adaptándose eficazmente a las variaciones en el entorno y manteniendo un comportamiento correcto.

#### 5. Generalización del Aprendizaje

- a. El agente ha demostrado no solo aprender a jugar correctamente en un momento específico, sino también generalizar ese aprendizaje para mantener un rendimiento efectivo a lo largo del tiempo.

#### 6. Consistencia en la Toma de Decisiones

- a. Se observó una consistencia en la toma de decisiones del agente, indicando una robustez en su capacidad para equilibrar el poste sobre el carro.

Estos resultados, respaldados por gráficos que ilustran la evolución del puntaje y el rendimiento continuo del agente, destacan la efectividad del enfoque de RL y RN en la resolución del desafío de CartPole. La capacidad del agente para aprender y generalizar comportamientos complejos en un entorno dinámico subraya el éxito de la implementación tecnológica.

### **Discusión**

En este proyecto, los resultados obtenidos no solo son altamente alentadores, sino que también arrojan luz sobre la efectividad de la combinación de Aprendizaje por Refuerzo (RL) y Redes Neuronales (RN) en el contexto desafiante de CartPole.

El agente logró aprender de manera efectiva a equilibrar el poste sobre el carro, alcanzando un puntaje impresionante de 683 después de aproximadamente 176

iteraciones de entrenamiento. Este logro destaca la capacidad del agente para resolver el problema específico y su habilidad para adaptarse y mejorar a lo largo del tiempo.

Durante la prueba prolongada de 5 minutos, el agente demostró una consistencia excepcional, manteniendo un rendimiento sólido sin perder el equilibrio. Este aspecto fundamental demuestra la capacidad del modelo para generalizar su aprendizaje y aplicarlo de manera efectiva en situaciones más prolongadas y diversas.

La consistencia observada sugiere que la Red Neuronal ha internalizado no solo patrones específicos de juego, sino también estrategias generales para mantener el equilibrio. Este hallazgo es crucial, indicando que la RN ha capturado la esencia del problema, permitiendo al agente adaptarse de manera efectiva a cambios en el entorno.

Aunque los resultados son notables, la consideración de los "mejores resultados" implica una comparación con otros enfoques y la exploración de posibles ajustes adicionales. Este proyecto proporciona una base sólida para futuras investigaciones, sugiriendo que la combinación de RL y RN podría aplicarse con éxito en problemas del mundo real que requieren adaptabilidad y aprendizaje autónomo.

Por ende, los resultados obtenidos no solo validan la efectividad de la metodología implementada, sino que también abren puertas a nuevas posibilidades de aplicación de Aprendizaje por Refuerzo y Redes Neuronales en entornos dinámicos y desafiantes. Este proyecto no solo logra entrenar un agente eficiente en CartPole, sino que también proporciona información valiosa sobre cómo estos modelos pueden aprender, generalizar y desempeñarse en situaciones prácticas y cambiantes.

## **Conclusiones**

En conclusión, este proyecto ha alcanzado resultados notables al emplear la combinación de Aprendizaje por Refuerzo (RL) y Redes Neuronales (RN) para

abordar el desafiante entorno de CartPole. En primer lugar, el agente logró un puntaje impresionante de 683 después de alrededor de 176 iteraciones de entrenamiento, demostrando la efectividad de la metodología implementada. Este logro no solo subraya la capacidad del agente para resolver el problema específico, sino también su habilidad para adaptarse y mejorar a lo largo del tiempo.

Otro punto crucial es la consistencia excepcional exhibida por el agente durante la prueba prolongada de 15 minutos. Mantuvo un rendimiento sólido sin perder el equilibrio, lo que indica que la RN no solo ha aprendido patrones específicos de juego, sino también estrategias generales para afrontar situaciones diversas. Este hallazgo destaca la capacidad del modelo para generalizar su aprendizaje de manera efectiva.

Si bien estos resultados son notables, es importante reconocer que la determinación de los "mejores resultados" implica una comparación con otros enfoques y la posibilidad de ajustes adicionales. Esta reflexión destaca la necesidad continua de investigación y mejora en la implementación de RL y RN en entornos dinámicos como el de CartPole.

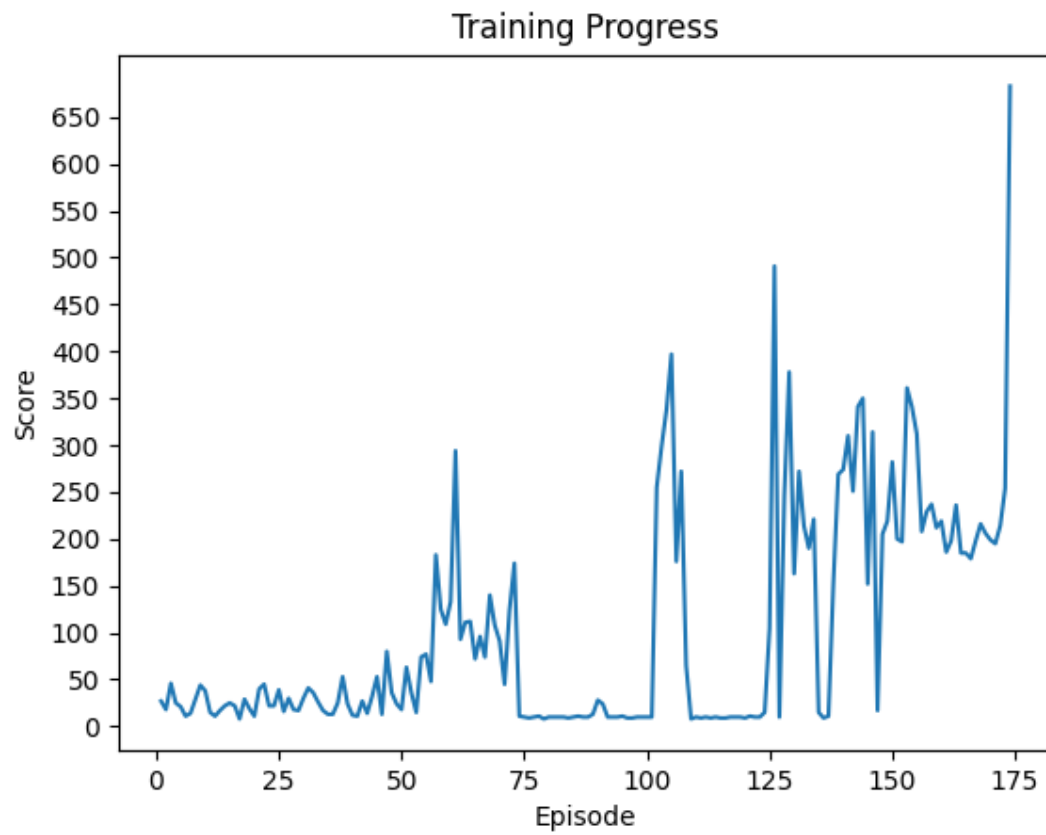
Por ello, este proyecto no sólo valida la efectividad de la metodología propuesta, sino que también abre nuevas perspectivas para la aplicación de Aprendizaje por Refuerzo y Redes Neuronales en problemas del mundo real. La capacidad del agente para aprender, adaptarse y desempeñarse en situaciones prácticas y cambiantes proporciona información valiosa y sienta las bases para futuras investigaciones en el campo del aprendizaje automático.

## **Apéndice**

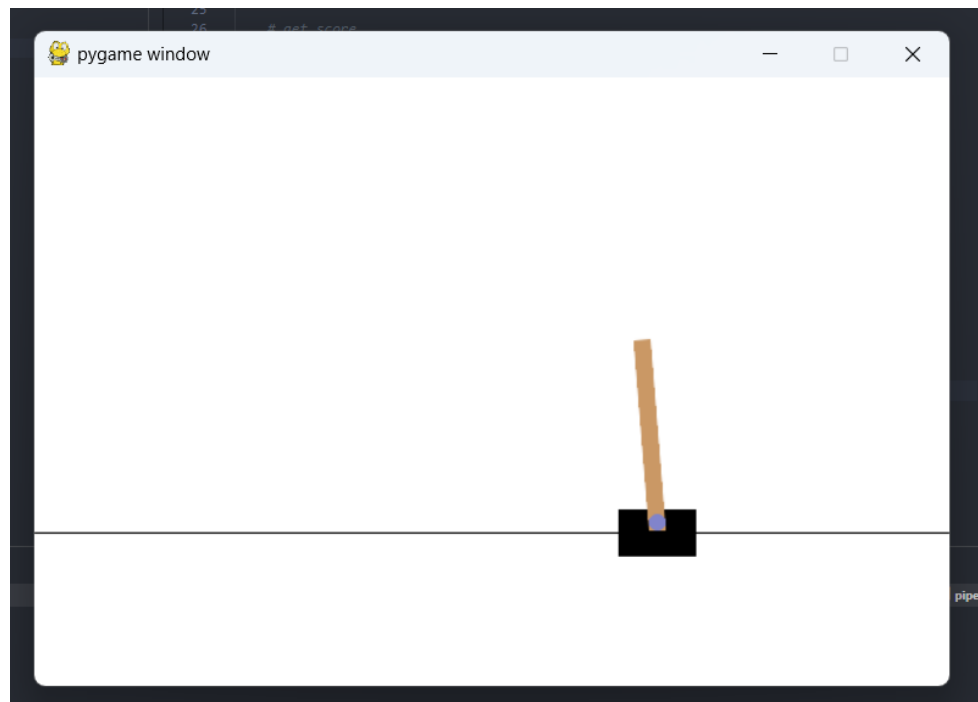
Repositorio [https://github.com/esc20936/cartPole\\_DQN](https://github.com/esc20936/cartPole_DQN)



## Anexos



**Figura 1. Progreso de entrenamiento**



**Figura 2. Interfaz del juego**

## **Bibliografía**

*Cart Pole - gym documentation.* (n.d.).

[https://www.gymlibrary.dev/environments/classic\\_control/cart\\_pole/](https://www.gymlibrary.dev/environments/classic_control/cart_pole/)

*Gymnasium documentation.* (n.d.). [https://gymnasium.farama.org/content/basic\\_usage/](https://gymnasium.farama.org/content/basic_usage/)

Mitaritonna, A. (2019, October 29). *¿Qué es y para qué sirve el Aprendizaje Profundo (Deep Learning)?*

<https://www.linkedin.com/pulse/qu%C3%A9-es-y-para-sirve-el-aprendizaje-profundo-deep-mitaritonna/?originalSubdomain=es>

Python Lessons. (2019, November 26). *Introduction to Reinforcement Learning - Cartpole*

*DQN* [Video]. YouTube. <https://www.youtube.com/watch?v=D795oNqa-Vk>

Sarmiento-Ramos, J. L. (2020). *Aplicaciones de las redes neuronales y el deep learning a la ingeniería biomédica.* <https://www.redalyc.org/journal/5537/553768213002/html/>