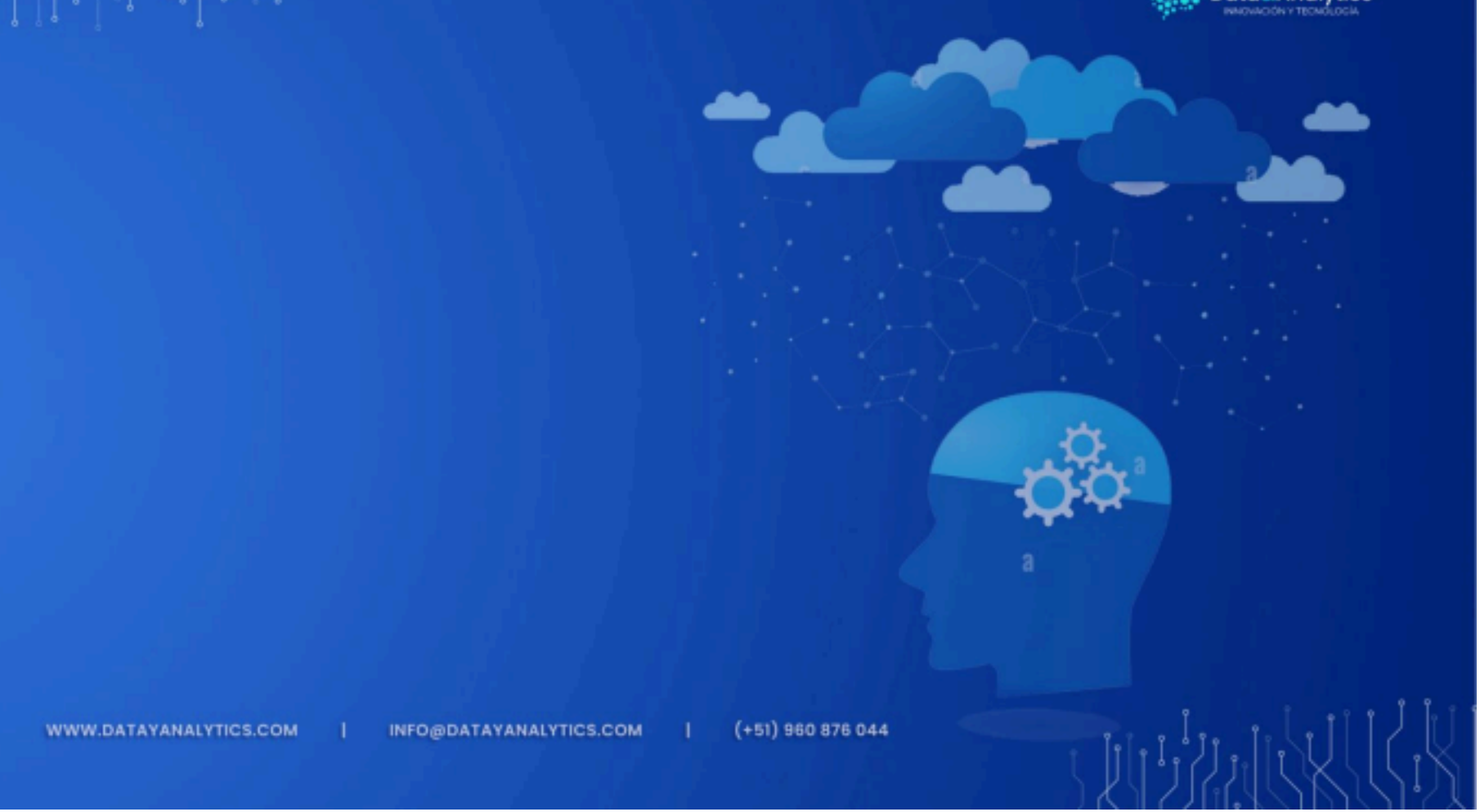
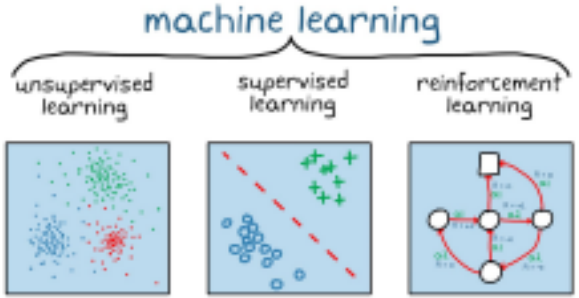
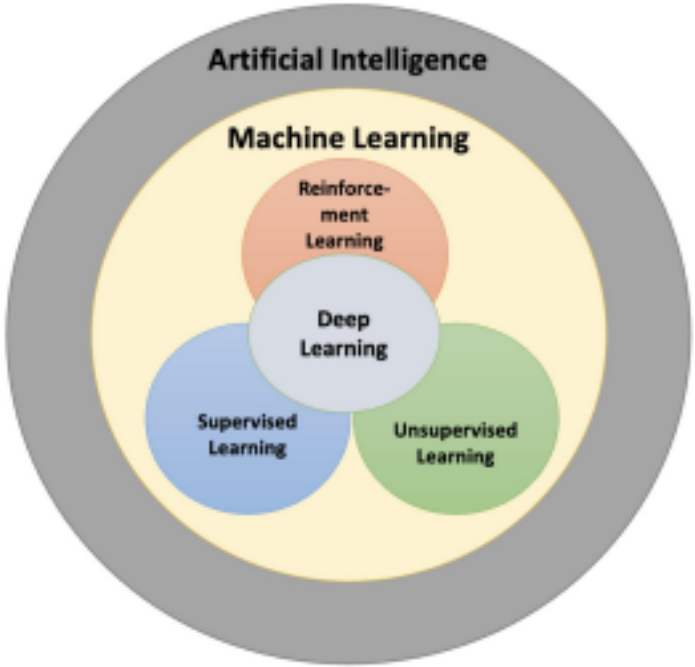


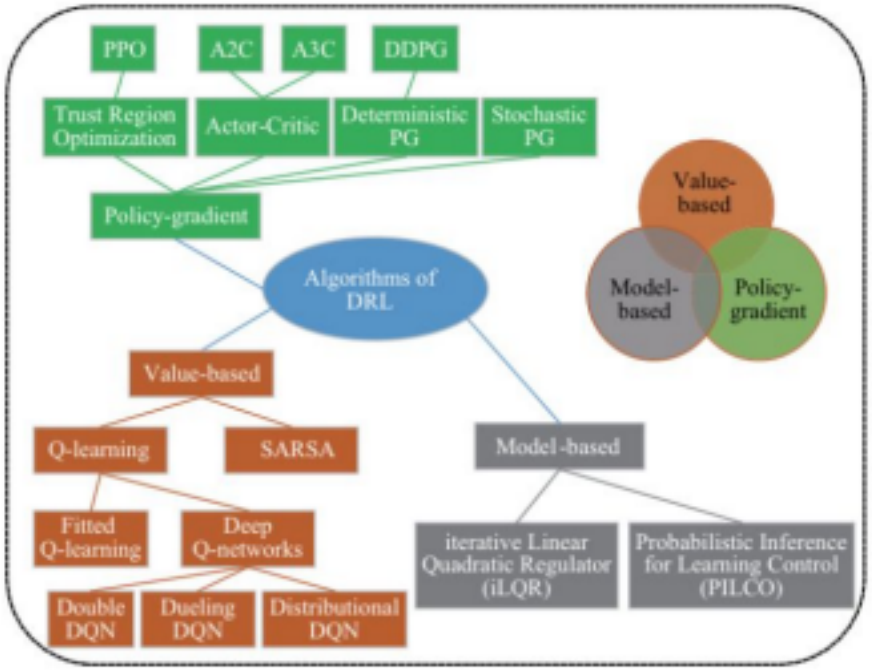
WWW.DATAYANALYTICS.COM | INFO@DATAYANALYTICS.COM | (+51) 960 876 044


machine learning
unsupervised learning
supervised learning
reinforcement learning

CIENCIA DE DATOS | INTELIGENCIA DE NEGOCIOS | BIG DATA | MACHINE LEARNING | INTELIGENCIA ARTIFICIAL | INNOVACION Y TECNOLOGÍA

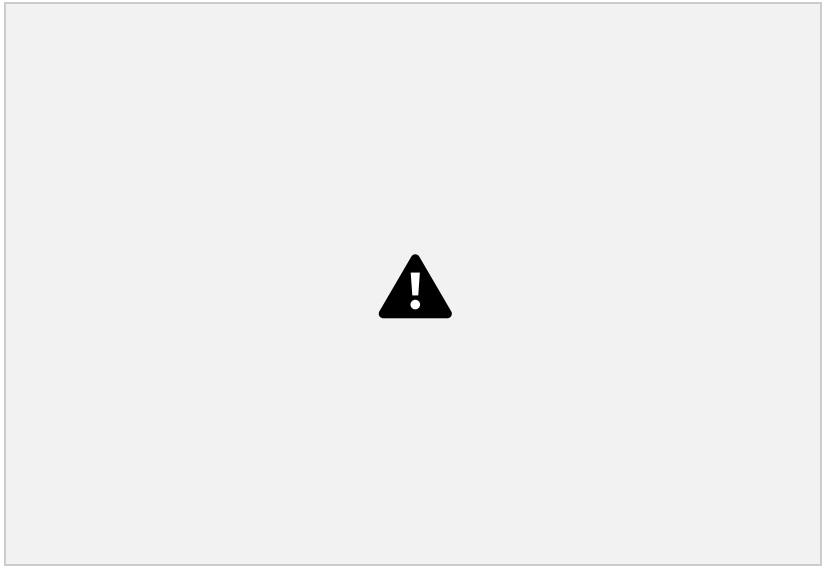

Artificial Intelligence
Machine Learning
Reinforce-
ment
Learning
Deep
Learning
Supervised
Learning
Unsupervised
Learning

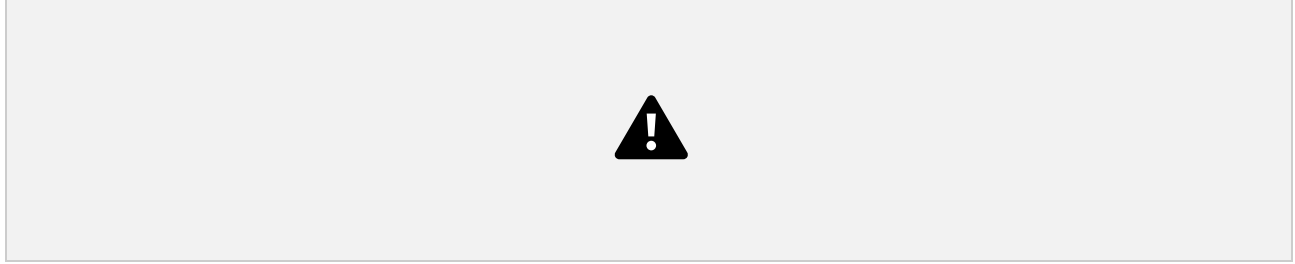

AlphaStar
2019
@Przemek Tokar /
Shutterstock.com
BY GOOGLE DEEPMIND


PPO
A2C
A3C
DDPG
Trust Region Optimization
Actor-Critic
Deterministic PG
Stochastic PG
Policy-gradient
Algorithms of DRL
Value-based
Model-based
Policy-gradient
Value-based
Q-learning
SARSA
Model-based
Fitted Q-learning
Deep Q-networks
iterative Linear Quadratic Regulator (iLQR)
Probabilistic Inference for Learning Control (PILCO)
Double DQN
Dueling DQN
Distributional DQN

Basados en política

Donde



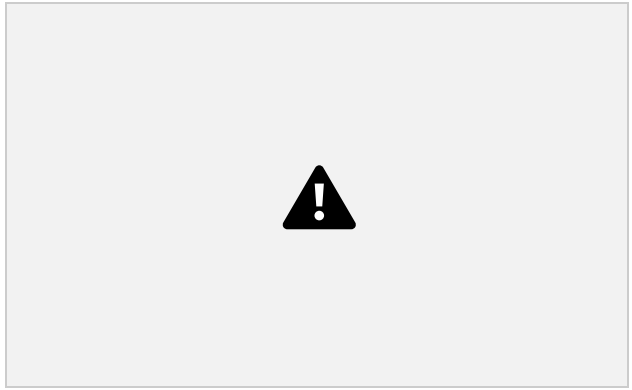


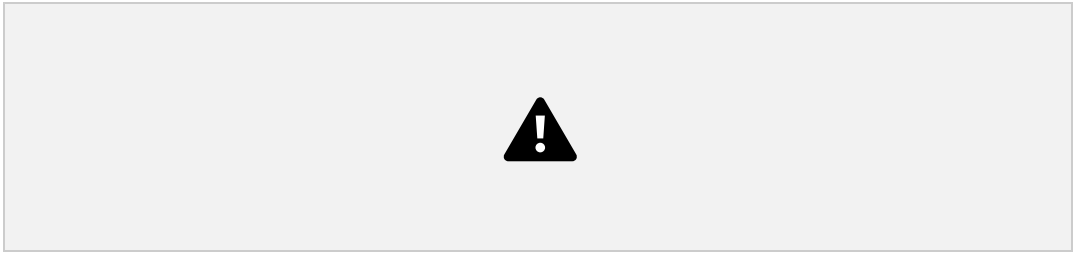
Basados en política
Estados y acciones continuas

Basados en valor
Estados y acciones discretos

(Por valor)

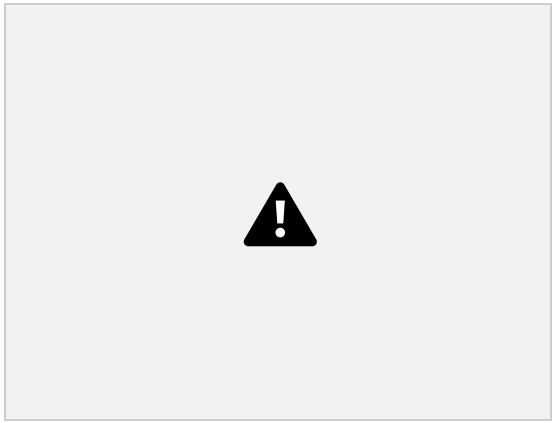
Propiedad de Markov





Propiedad de Markov

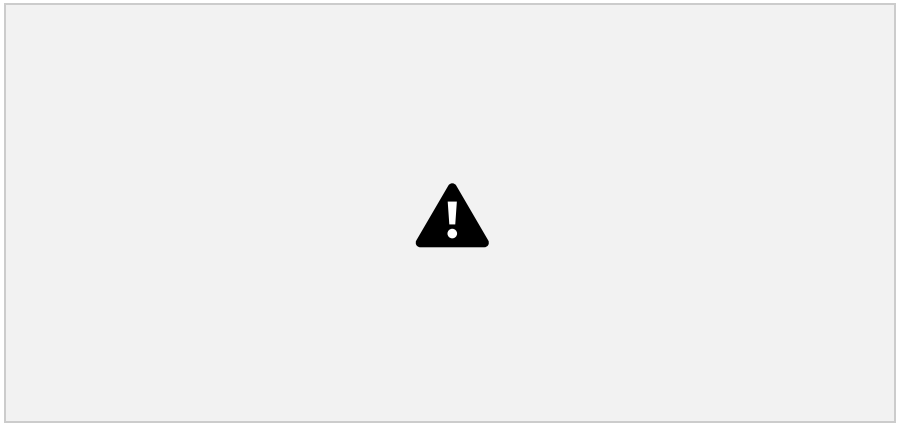
Distribu~~ción inicial~~

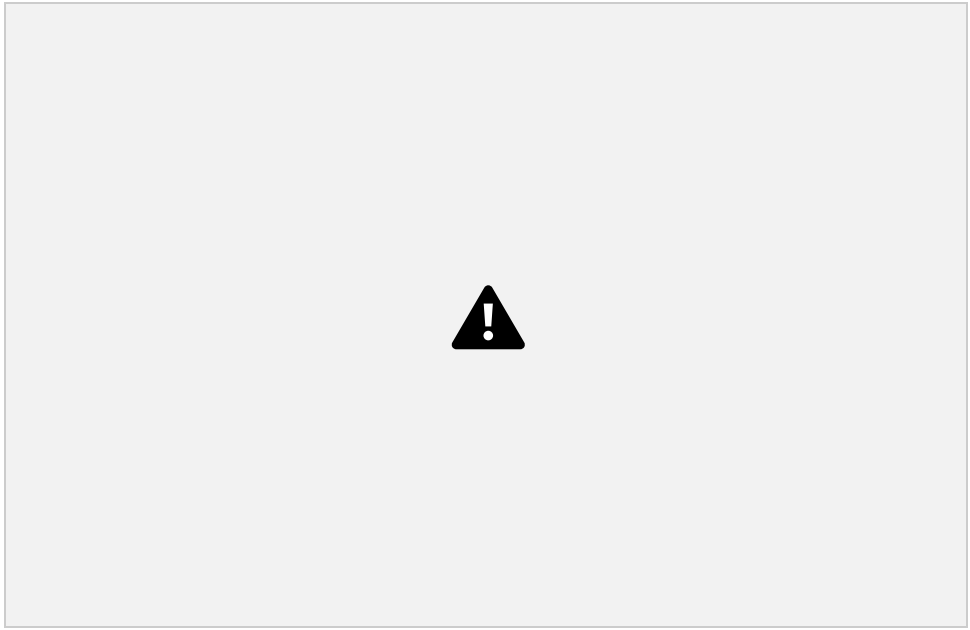


Propiedad de Markov

Distribución ~~inicial~~

Cadena de Markov homogénea

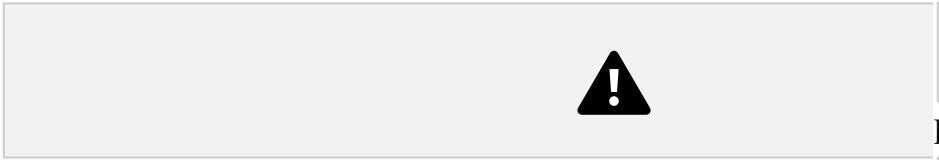




P(a|s)

P(s,s’)

Cadena de Markov



homogénea

Probabilidad de transición sujeta a una acción

Recompensas a un paso

Función de recompensas

Política

Retorno

Función valor estado acción

Ecuación de Bellman

Continua

Recompensas a un paso

Resultados de la función valor estado acción

Política aleatoria

Probabilidades de transición

Política óptima

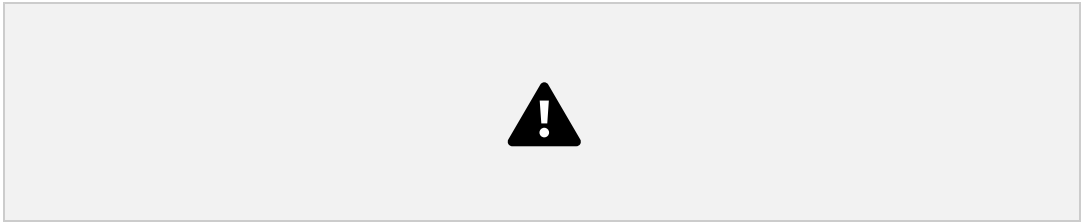
Ecuaciones de Bellman

Predicción y Control

Predicción: Dado una política, se desea predecir el valor de la función valor.

mejorar esa
política a través de la experiencia.

Control: Dado una política, se desea



Predicción

Evaluación iterativa de políticas

Predicción Control Iteración de política

Dada una política inicial arbitraria, se evalúa su función valor usando el método de Evaluación iterativa de políticas , y posteriormente se mejorará la política usando la política **greedy** o la política **epsilon-greedy.**

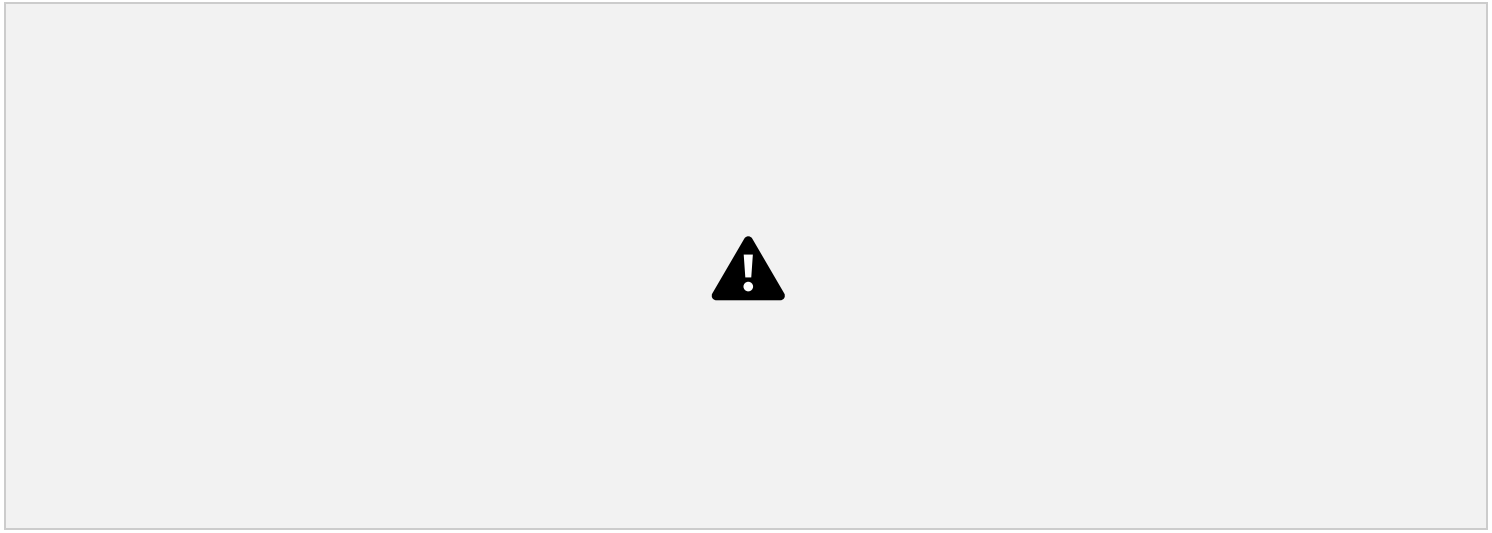
Evaluación iterativa de políticas

Política codiciosa **Política epsilon-greedy**

Predicción Control Iteración de política

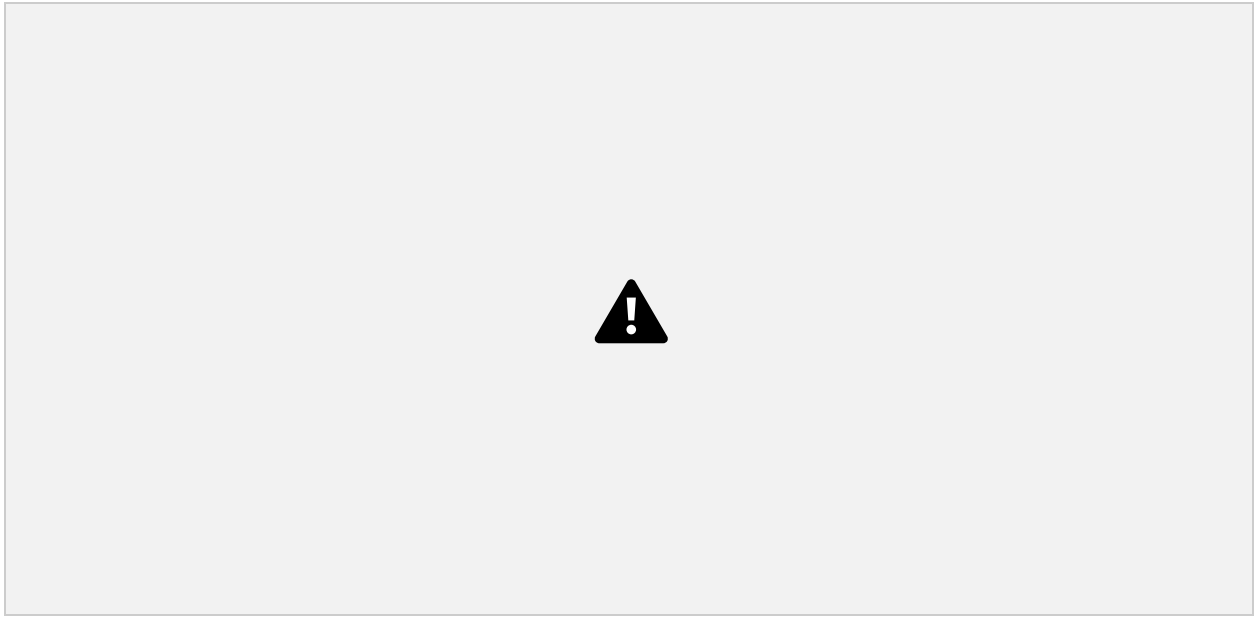
Dada una política inicial arbitraria, se evalúa su función valor usando el método de Evaluación iterativa de políticas , y posteriormente se mejorará la política usando la política **greedy** o la política **epsilon-greedy.**

Evaluación iterativa de políticas
Iteración valor



Evaluación iterativa de políticas

Iteración de política



RL - Montecarlo

Episodios

1

Para cada par estado acción

Usando una política, para cada episodio se simula por montecarlo la secuencia. Episodio 1

Se calcula el valor de para cada t.

del episodio.

2

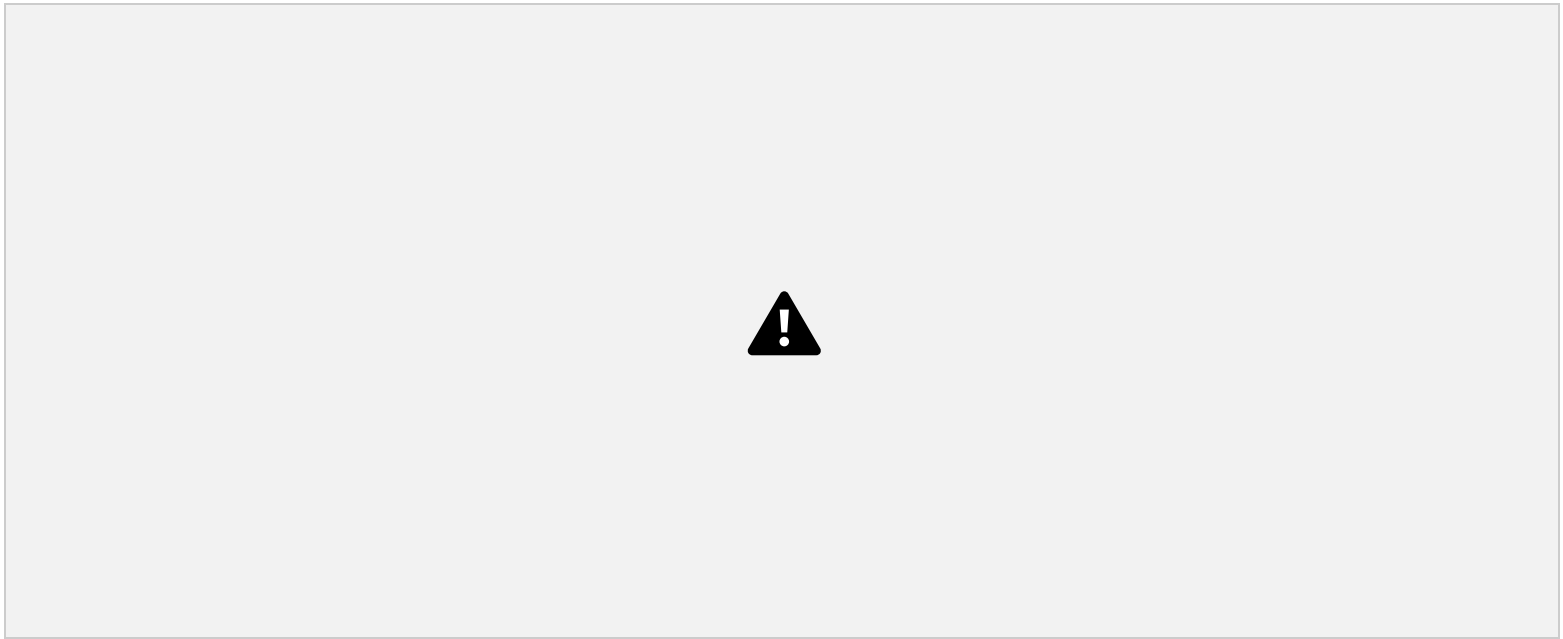
Episodio 2

N Teorema central del límite

Episodio N

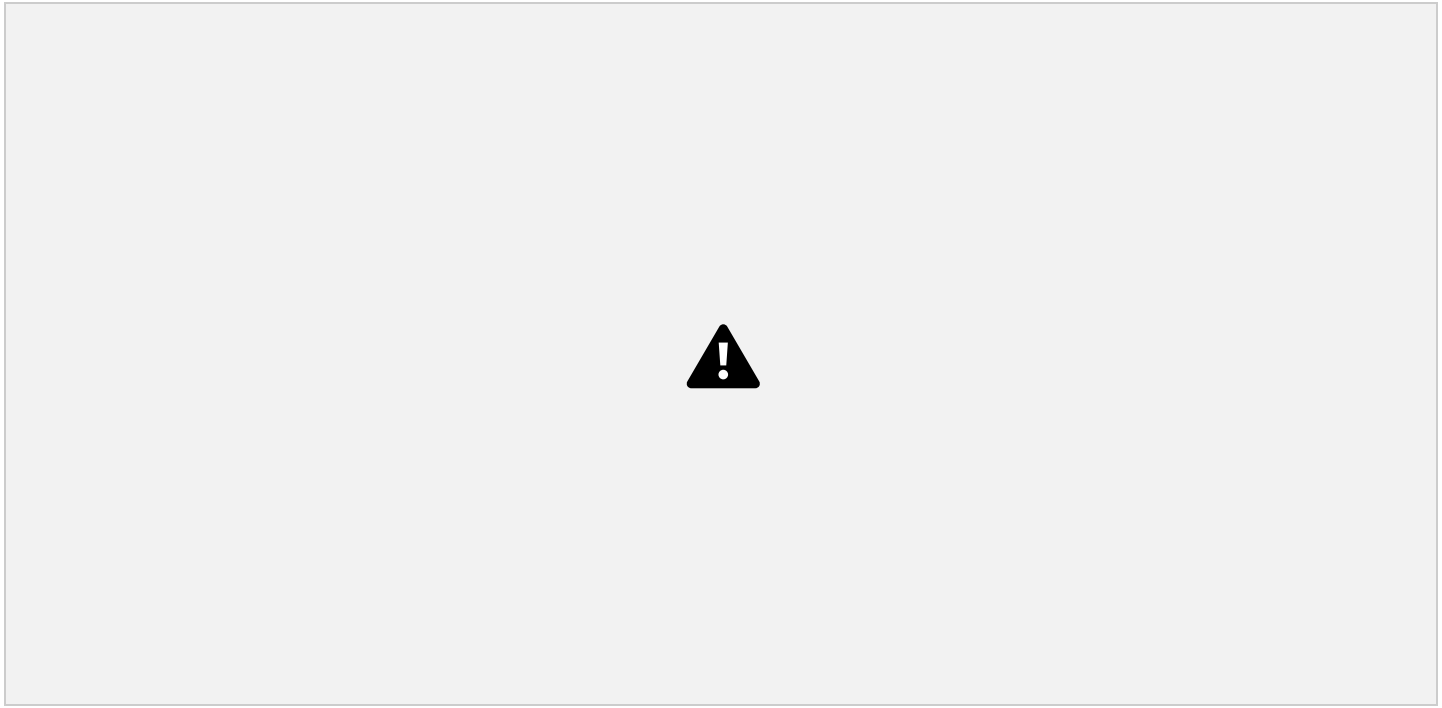
RL - Montecarlo

Predicción



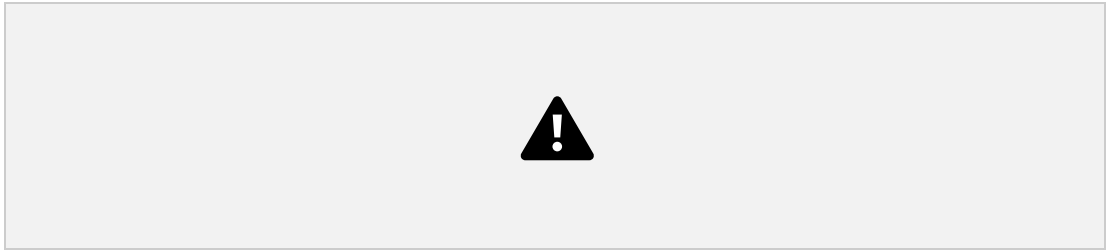
RL - Montecarlo

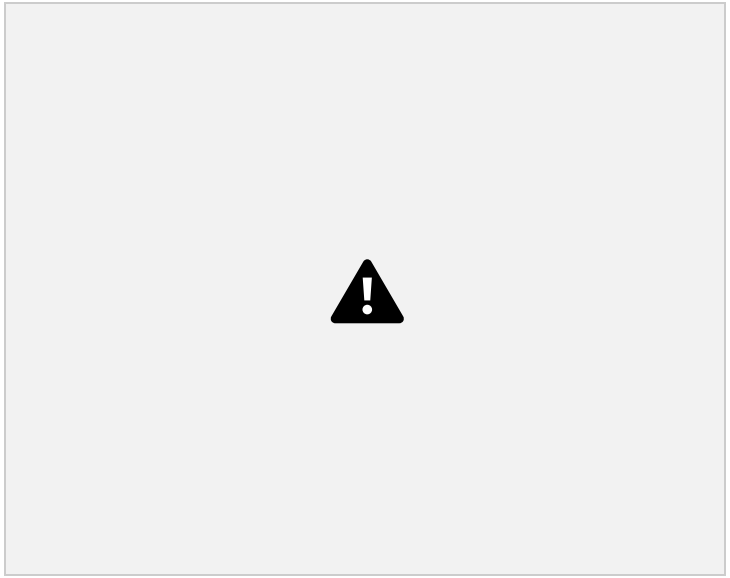
Control



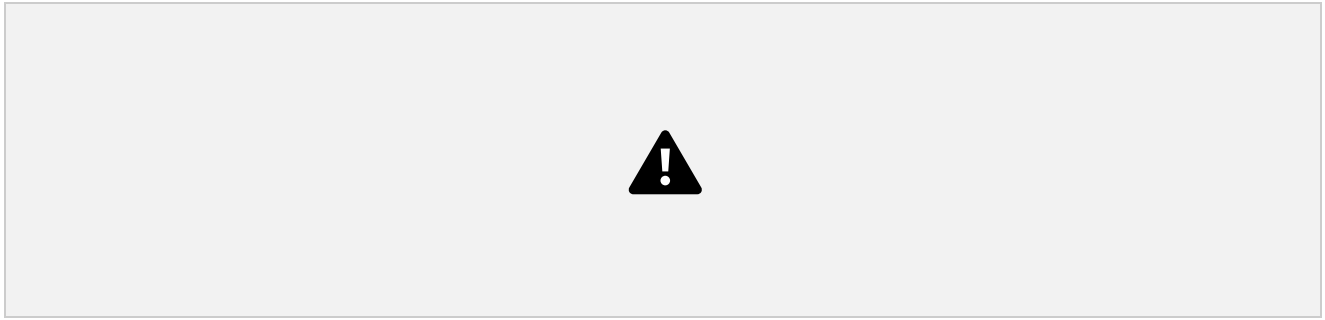
RL - Montecarlo

Control

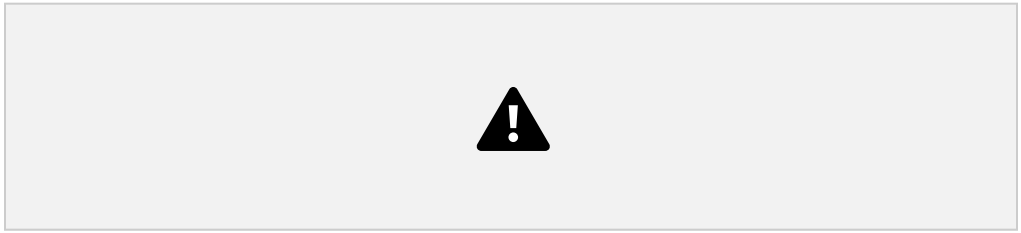




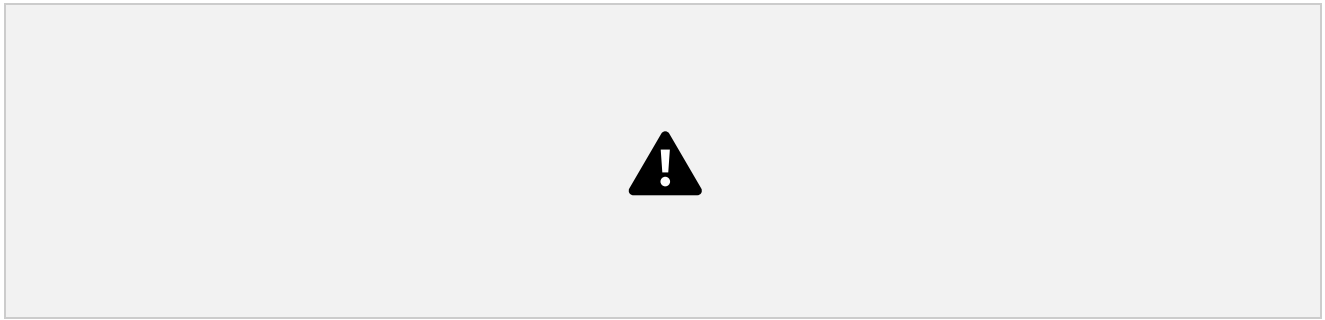
Método montecarlo en forma secuencial



Diferencia temporal (TD)



Predicción



Control

SARSA (On-policy) Q-Learning Comportamiento (Off-policy)

y Aprendizaje

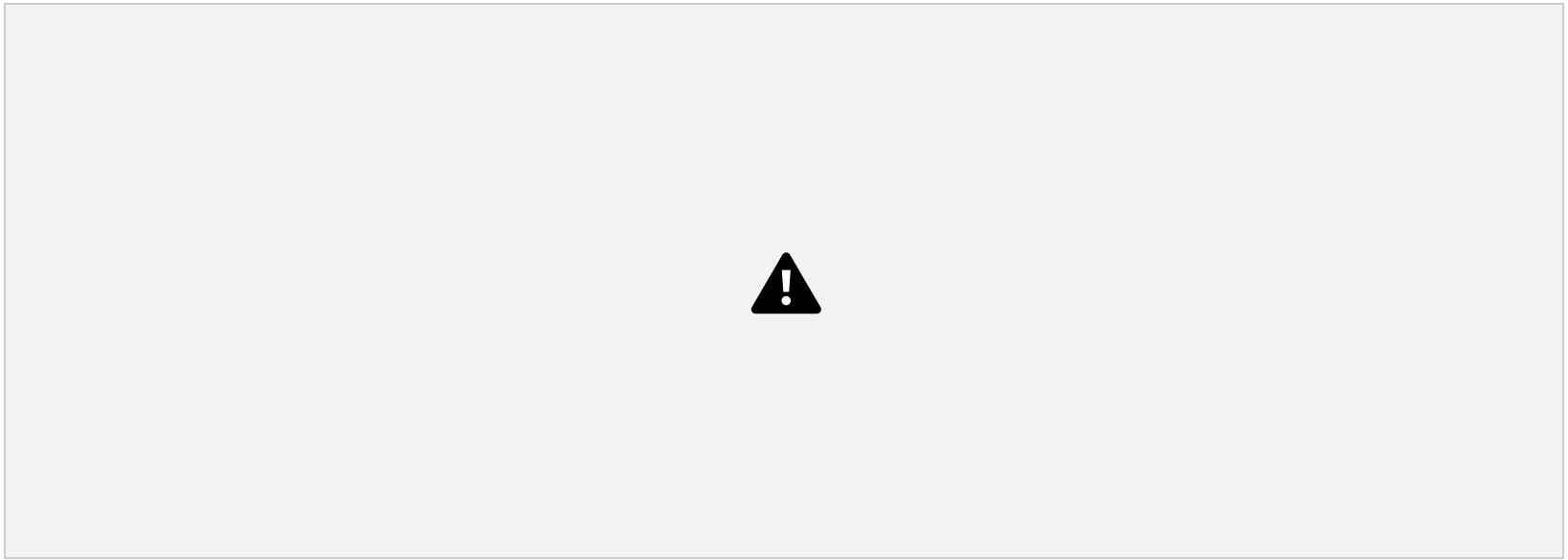
s’

ra’

s a

Control

SARSA



Control

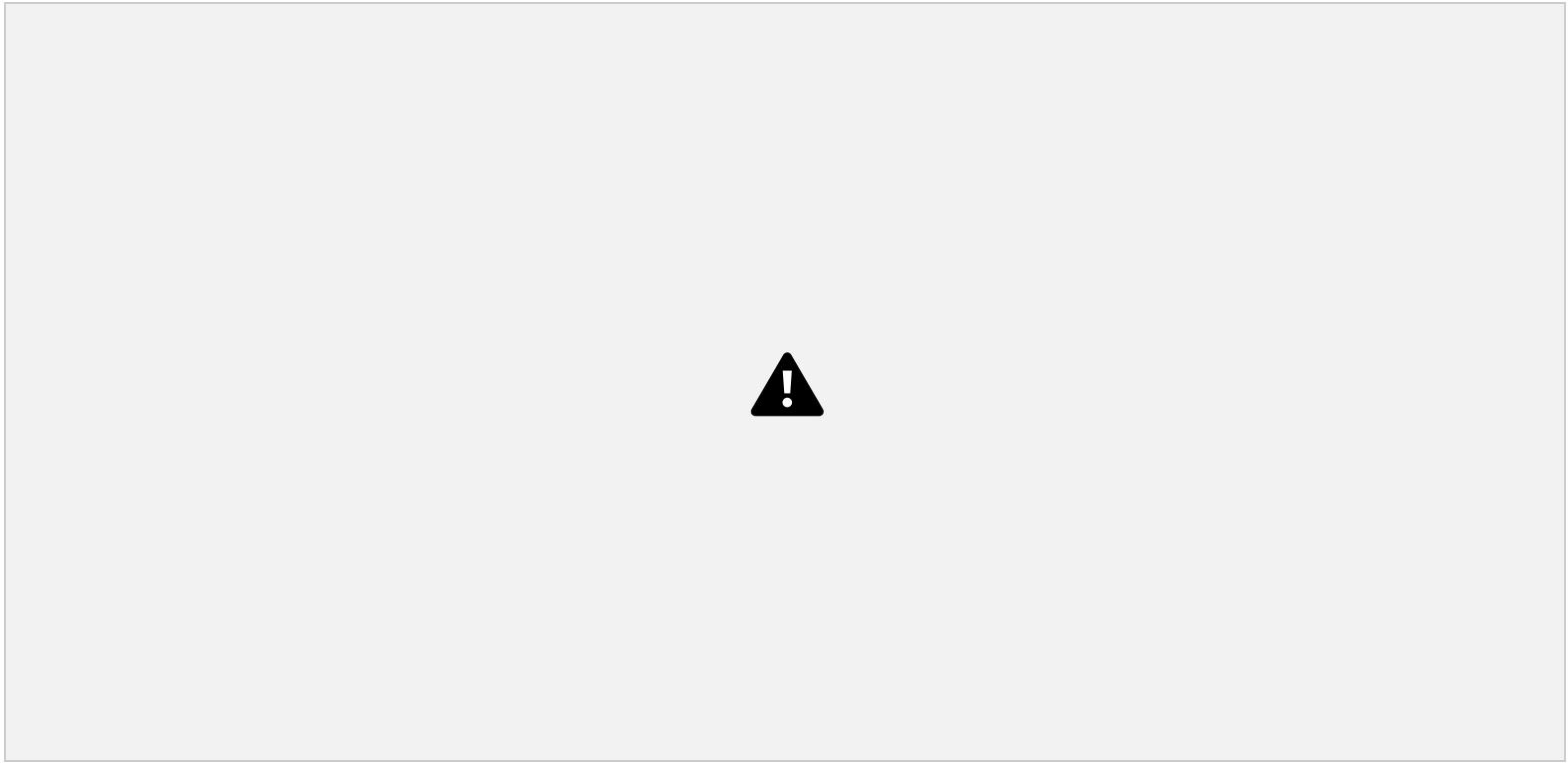
Q-Learning

Montecarlo

Diferencia
temporal

Montecarlo

Predicción



Montecarlo
Control

Diferencia temporal

SARSA

Diferencia temporal

Q-Learning