Open | DSE

Introduction to Data Science with Python Lecture 5: Reinforcement Learning

Vladimir Osin

Data Scientist/Engineer

Signify Research (formerly known as Philips Lighting)

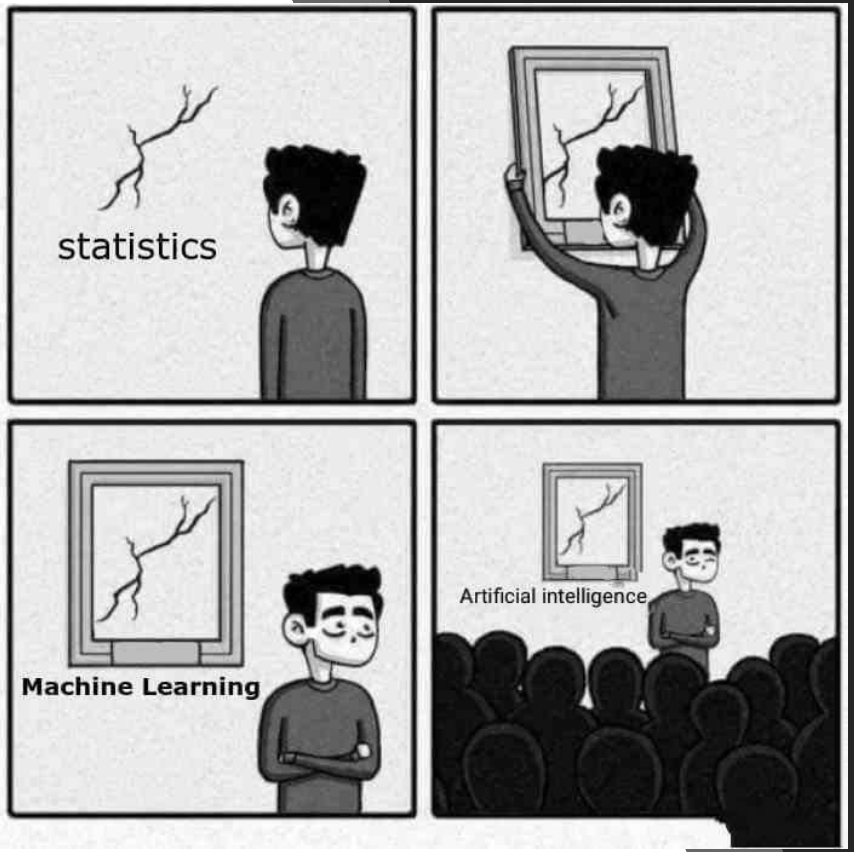
Outline

- Definition
- Examples
- Open AI GYM
- Practice

Reinforcement Learning

- **Reinforcement learning** (**RL**) is an area of machine learning concerned with how software agents ought to take *actions* in an *environment* to maximize some notion of cumulative *reward*.

Journey into ML world

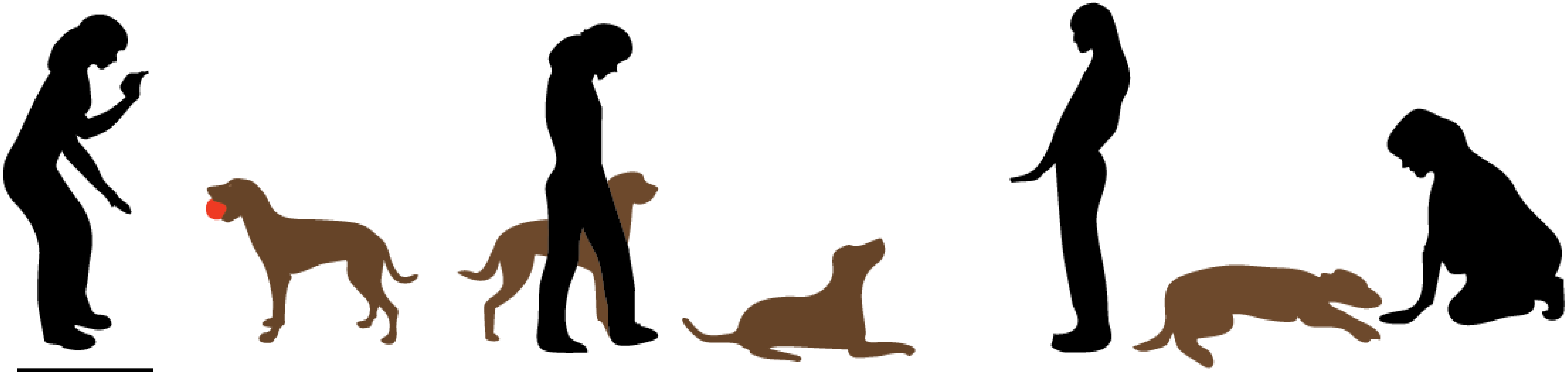


- Supervised Learning
  - Data: features and ground truth is available (x, y)
  - Goal: learn a function that maps x to y
  - Examples: classification, regression

- Unsupervised Learning
  - Data: just features and no ground truth
  - Goal: learn some underlying hidden structure of the data
  - Examples: clustering, dimensionality reduction

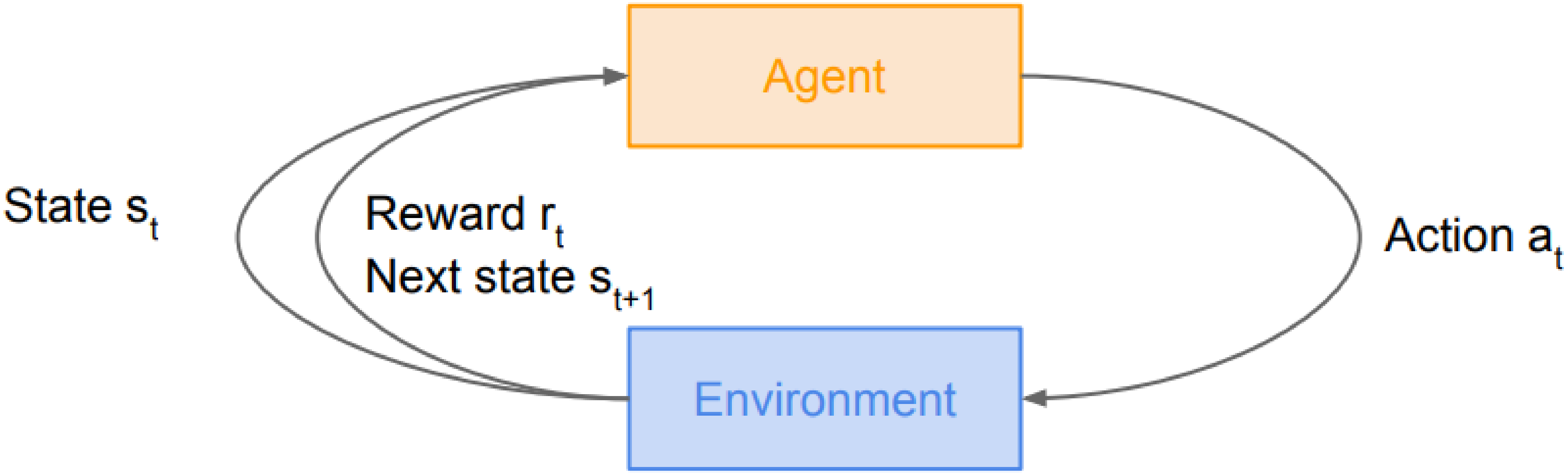
Open | DSE

Reinforcement Learning Idea



Open | DSE

Reinforcement Learning Idea



Agent

State s_t

Reward r_t

Next state s_{t+1}

Action a_t

Environment

Markov Decision Process

- Mathematical formulation of the RL problem
- **Markov property**: Current state completely characterises the state of the world

Defined by: $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathbb{P}, \gamma)$

$\mathcal{S}$: set of possible states
$\mathcal{A}$: set of possible actions
$\mathcal{R}$: distribution of reward given (state, action) pair
$\mathbb{P}$: transition probability i.e. distribution over next state given (state, action) pair
γ : discount factor

Markov Decision Process

- At time step t=0, environment samples initial state $s_0 \sim p(s_0)$
- Then, for t=0 until done:
  - Agent selects action a_t
  - Environment samples reward $r_t \sim R(. \mid s_t, a_t)$
  - Environment samples next state $s_{t+1} \sim P(. \mid s_t, a_t)$
  - Agent receives reward r_t and next state s_{t+1}

- A policy **π** is a function from S to A that specifies what action to take in each state
- **Objective**: find policy **π*** that maximizes cumulative discounted reward: $\sum_{t \geq 0}$

Examples

- https://deepmind.com/research/publications/playing-atari-deep-reinforcement-learning/
- http://news.berkeley.edu/2015/05/21/deep-learning-robot-masters-skills-via-trial-and-error/
- https://blog.google/technology/ai/alphago-machine-learning-game-go/
- https://storage.googleapis.com/deepmind-media/dqn/DQNNaturePaper.pdf

OpenAI

References

- CS231n Reinforcement Learning Lecture
- https://gym.openai.com/
- https://pytorch.org/tutorials/intermediate/reinforcement_q_learning.html