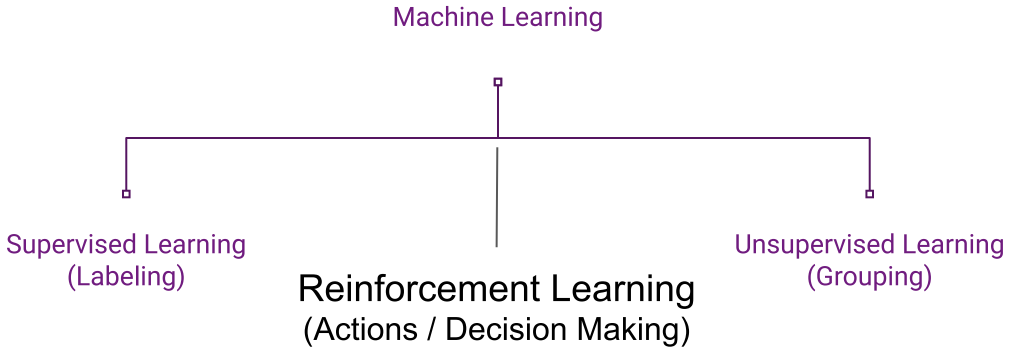
Reinforcement Learning

Jacob Wan (BMI6015)

How does it fit into grand scheme of things.

When to use what? Why do we need RL?



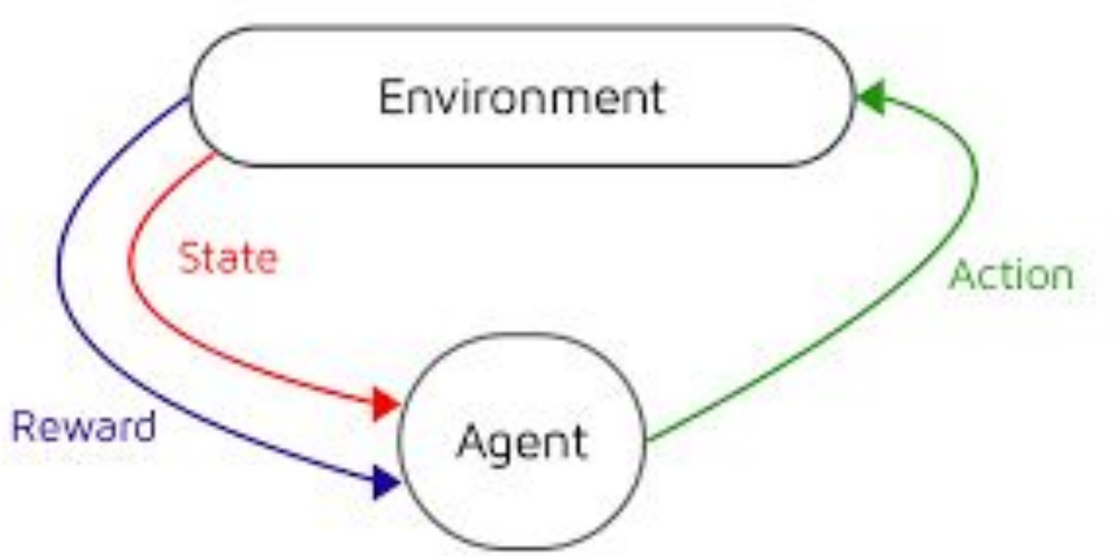
Machine Learning

Supervised Learning
(Labeling)

Reinforcement Learning
(Actions / Decision Making)

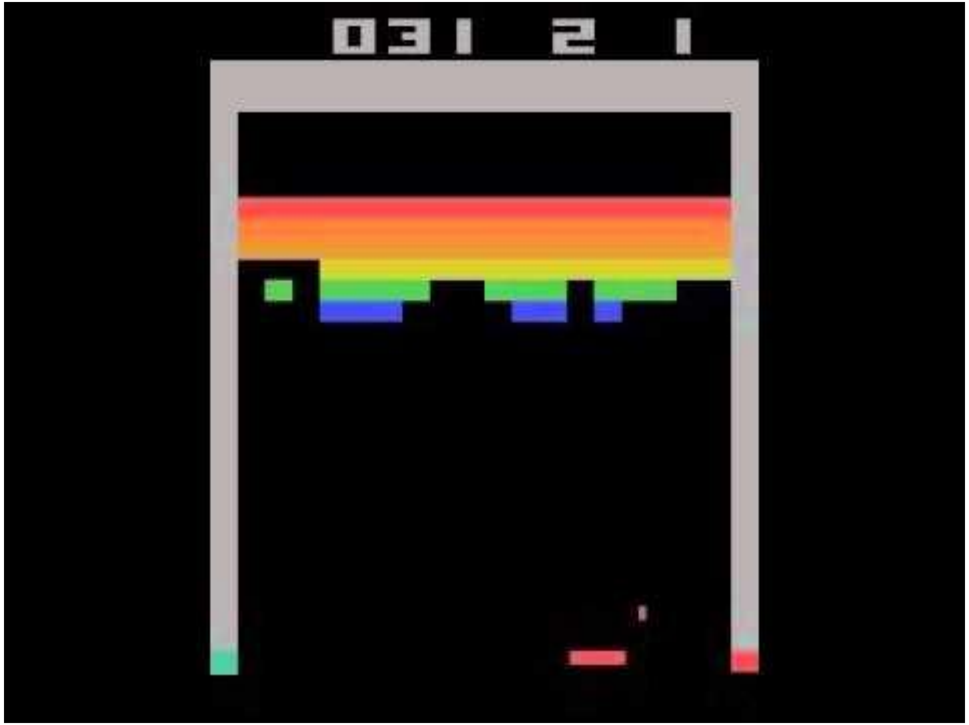
Unsupervised Learning
(Grouping)

a picture's worth. ..



src;intel.ai https://www.intel.ai/openai/#gs.1yc0yx

A video's worth. ...



Components

Overall Goal (Reward Function)

- What are we trying achieve? (Example: To be rich)

State (Reward Shaping)

- How to we guide the agent to maximize the reward:
  - For example, bigger paycheck means more money in the future.

Action (Policy Gradient)

- Work harder? Too hard? Work smarter?, etc.

Reward Received (Feedback Frequency)

- How often do we get pay raise?

Bellman Equation: = Max the benefits

Reward function: (how do I go further?)

$$V(x) = \max_{a \in \Gamma(x)} \{F(x, a) + \beta V(T(x, a))\}.$$

The State's value (where have I been?)

THERE

Challenges

Credit Assignment Problem (all decisions are either good or bad)

Are all the decisions lead us to more money?

Sparse Reward settings (how frequent do you get the reward)

Pay raise, once a year? too sparse?

Alignment Problem (are we rewarding the right thing?)

Money = root of all evil ?

3 Types of Reinforcement Learning

1. Model-Based

Find out how the world works, and act best based on that model of the world. For example in a game, the rules of the game is the model.

2. Value-Based

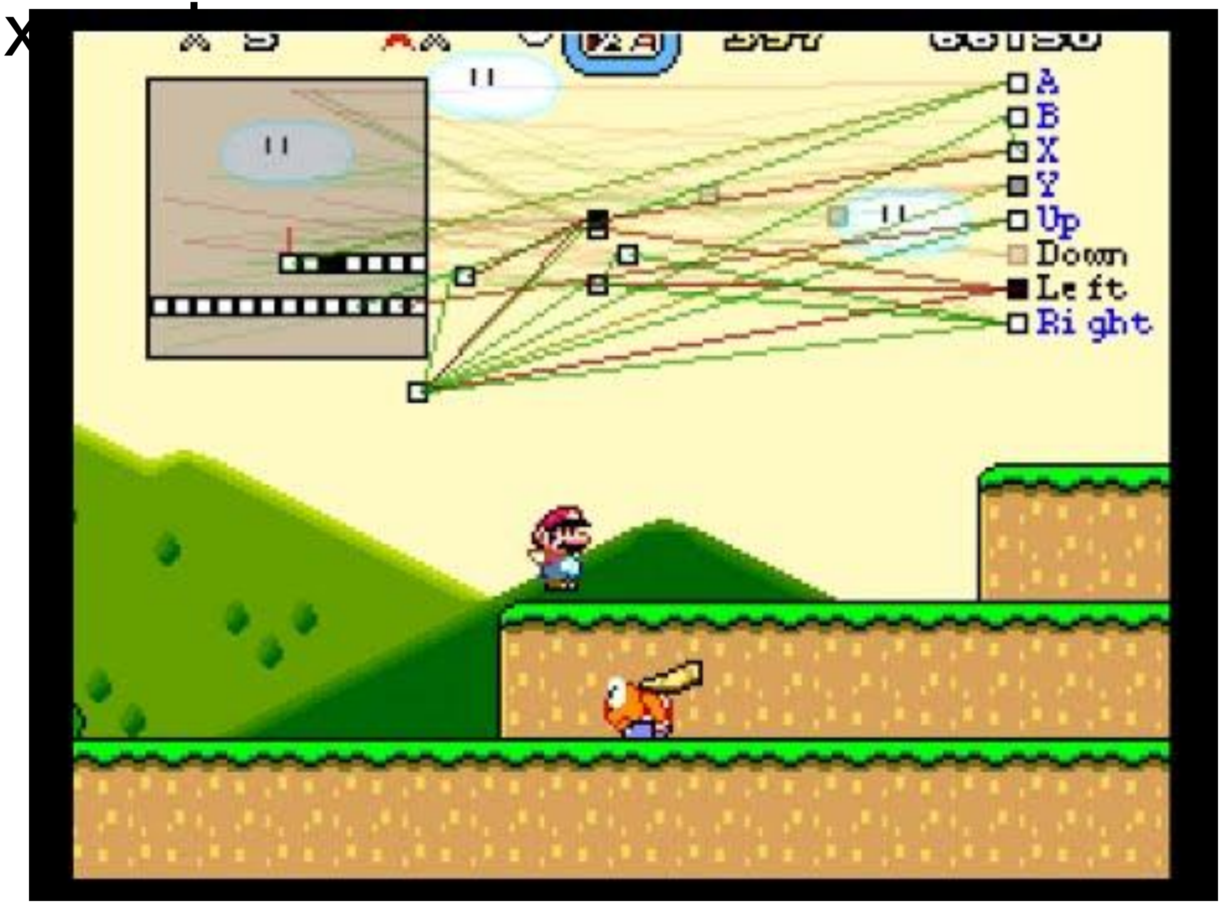
Assume which action are more valuable given a state, do those actions, observe the consequences and then update the values. Doesn't know much about the world

3. Policy-based

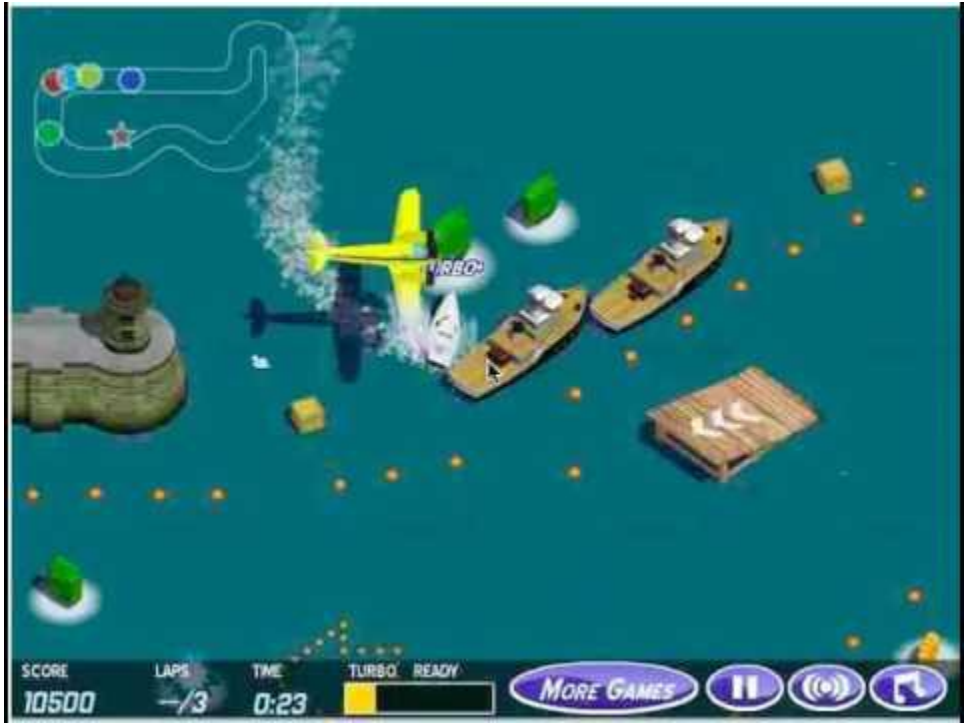
Apply the mapping policy (sequence of steps) regardless of values of each state

Popular Ex

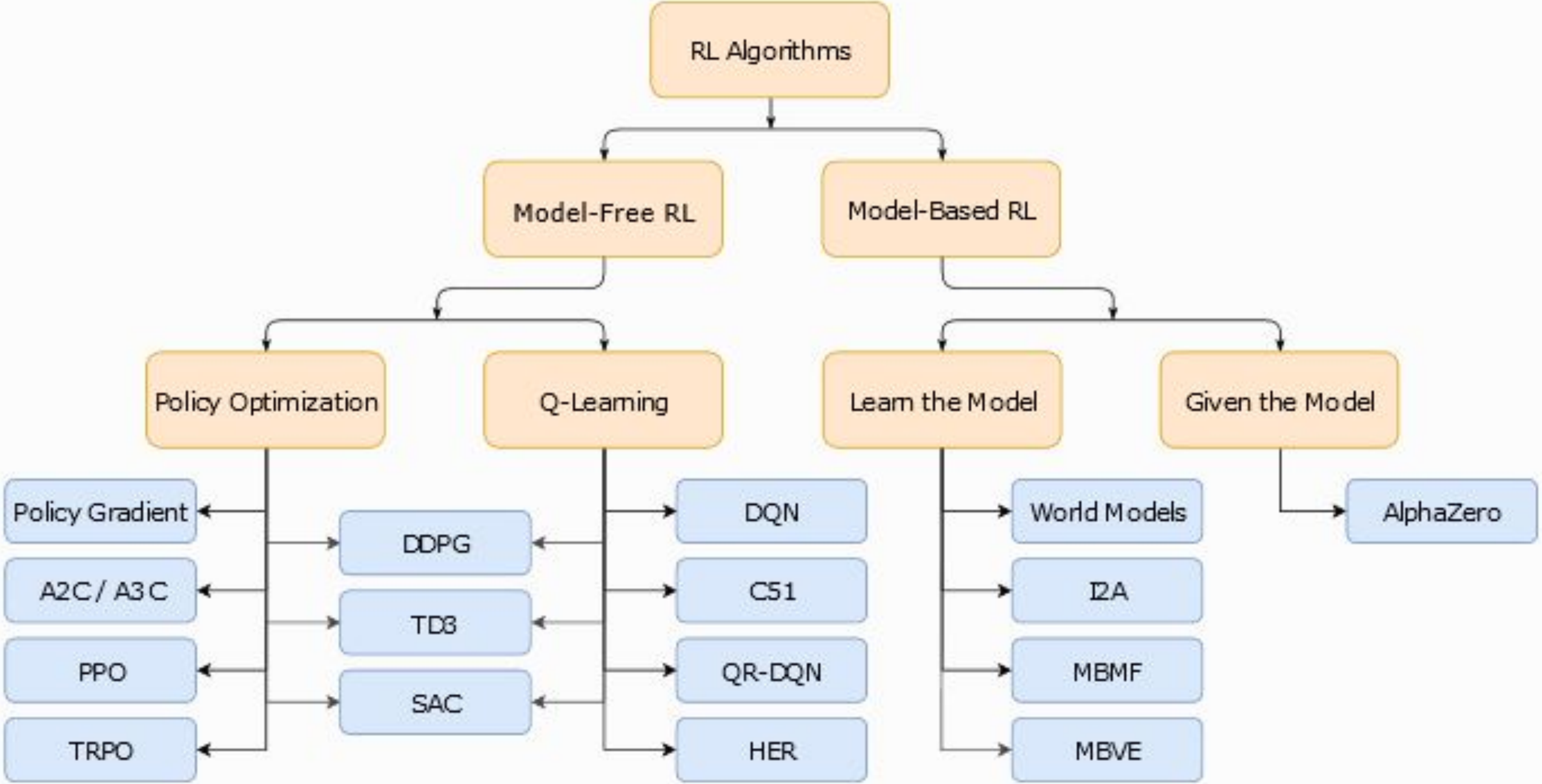
Games



Games with Faulty reward function

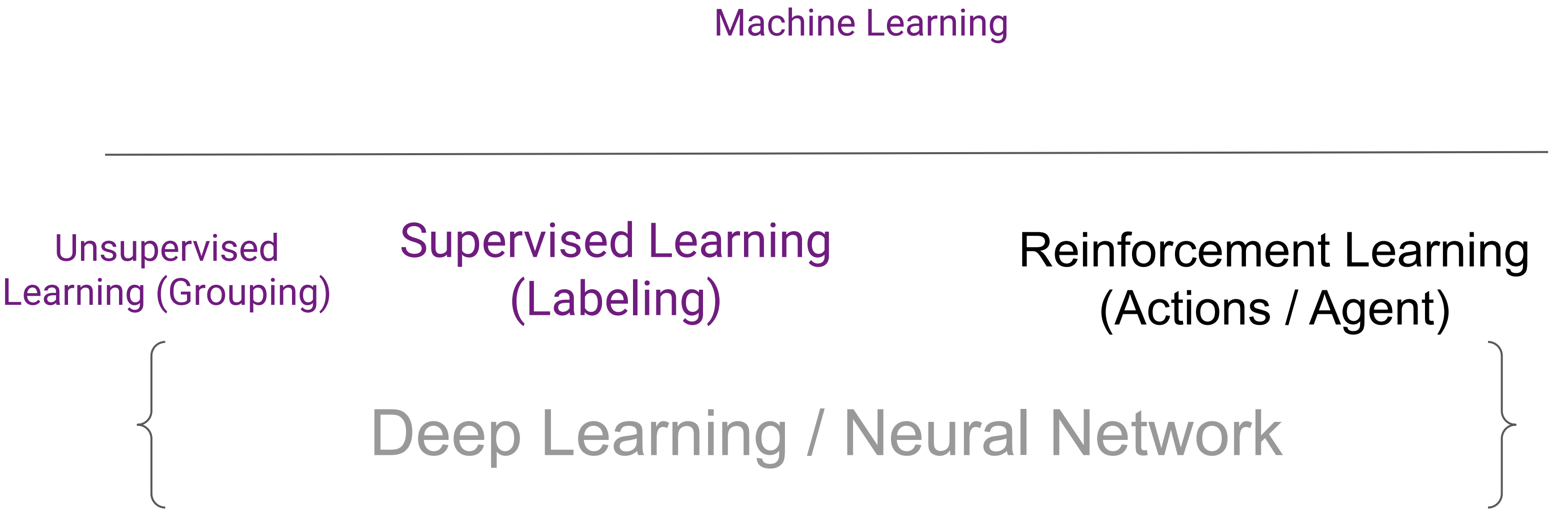


A Taxonomy of RL Algorithms



A non-exhaustive, but useful taxonomy of algorithms in modern RL. Citations below.

Continuum of integration



Machine Learning

Unsupervised Learning (Grouping)

Supervised Learning (Labeling)

Reinforcement Learning (Actions / Agent)

Deep Learning / Neural Network

How does it applies to Biomedical Informatics?

A Reinforcement Learning–Based Method for Management of Type 1 Diabetes: Exploratory Study

Mahsa Oroojeni Mohammad Javad[1], PhD ; Stephen Olusegun Agboola[2,3], MPH, MD ; Kamal Jethwani[2], MPH, MD ; Abe Zeid[4], PhD ; Sagar Kamarthi[4], PhD

[1]Department of Information Technology and Analytics, Kogod School of Business, American University, Washington, DC, United States

[2]Department of Dermatology, Harvard Medical School, Boston, MA, United States

[3]Partners HealthCare, Boston, MA, United States

[4]Mechanical and Industrial Engineering Department, College of Engineering, Northeastern University, Boston, MA, United States

https://diabetes.jmir.org/2019/3/e12905/#table4

A Reinforcement Learning–Based Method for Management of Type 1 Diabetes: Exploratory Study

Not the first paper, but the most recent (top results from Pubmed)

Environment: Patient (Patient's Data)

Agent's Action: Recommends Insulin Dosage Level

Reward: HbA1c Delta

Performance / Error checking: Compare with actual physician prescribed dosage

Method

Training / Validation Set = 87 / 60

Algorithm = Q-Learning (State Value Table)

State Representation = HbA1c Level + BMI + Alcohol usage Lvl.+ Activity Lvl.

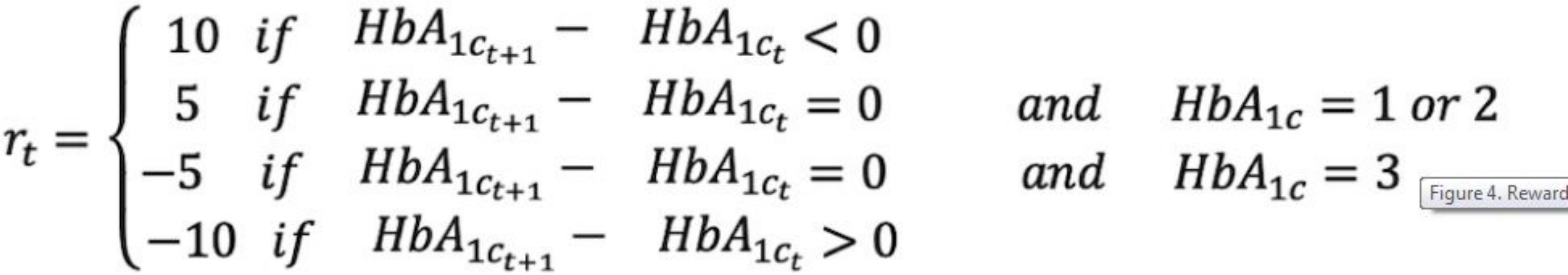
Action = Dosage Level recommendation

Reward = ***if*** the patient actually taken that dosage *and* HbA1c is lower = good.

Training = Agent randomly chooses the dosage, find *State* that comes after

Iteration = Each Patient visit is an epoch

Reward Function:



$$r_t = \begin{cases} 10 \quad if \quad HbA_{1c_{t+1}} - \quad HbA_{1c_t} < 0 & \\ 5 \quad if \quad HbA_{1c_{t+1}} - \quad HbA_{1c_t} = 0 & and \quad HbA_{1c} = 1 \; or \; 2 \\ -5 \quad if \quad HbA_{1c_{t+1}} - \quad HbA_{1c_t} = 0 & and \quad HbA_{1c} = 3 \\ -10 \quad if \quad HbA_{1c_{t+1}} - \quad HbA_{1c_t} > 0 & \end{cases}$$

Results

53 / 60 cases are similar* physician-prescribed

(*agent recommends an interval rather than a hard number)

Other Biomedical Informaticsapplications

Imaging:

https://ieeexplore.ieee.org/abstract/document/4221426

https://ieeexplore.ieee.org/abstract/document/1716136

Diagnosis: Survey of chronic illness and predict outcome

https://link.springer.com/article/10.1007/s10994-010-5229-0

Decision Support: Dynamic Treatment adjustment

https://www.ncbi.nlm.nih.gov/pubmed/31475215

Medical Record

http://www.kamishima.net/archive/2013-ws-ecmlpkdd-print.pdf

How do I use it?

OpenAI Gym

OpenAI Baselines

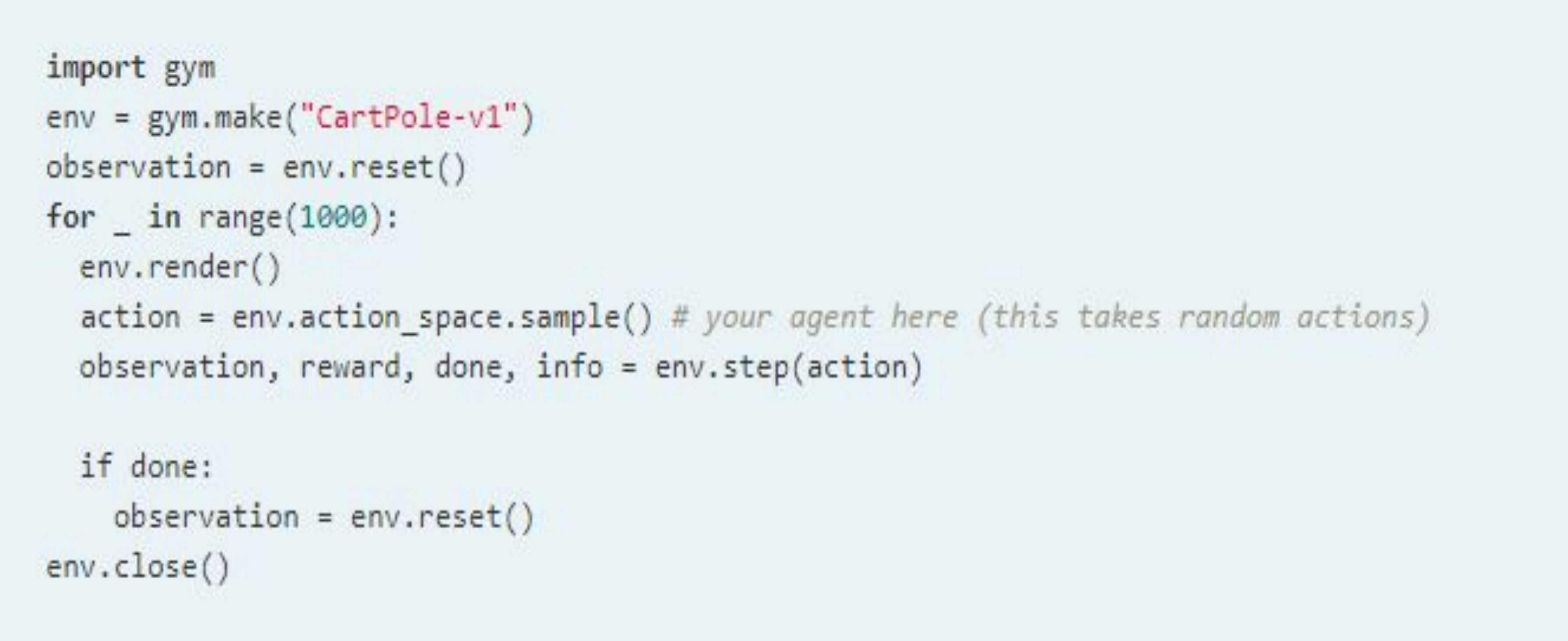
OpenAI Retro

Rllib

TF Agents

https://www.quora.com/What-are-some-of-the-good-Reinforcement-Learning-libraries

Gym from OpenAI



```
import gym
env = gym.make("CartPole-v1")
observation = env.reset()
for _ in range(1000):
  env.render()
  action = env.action_space.sample() # your agent here (this takes random actions)
  observation, reward, done, info = env.step(action)

  if done:
    observation = env.reset()
env.close()
```

More topics:

Many other favor of RL algorithm:

Multi-Agent implementation

Parallel Processing Implementation: Communications / Time / Protocol / Privacy concern.

Optimization challenges, etc

Resources:

The Book

https://towardsdatascience.com/what-are-the-types-of-machine-learning-e2b9e5d1756f

http://www2.econ.iastate.edu/tesfatsi/RLUsersGuide.ICAC2005.pdf

Other resources

https://towardsdatascience.com/simple-reinforcement-learning-q-learning-fcddc4b6fe56

https://www.learndatasci.com/tutorials/reinforcement-q-learning-scratch-python-openai-gym/

the whole book :

http://www.incompleteideas.net/book/RLbook2018trimmed.pdf