Webinar

UCB Reinforcement Learning

In R Programming

Lecturer:

Sajad Heydari

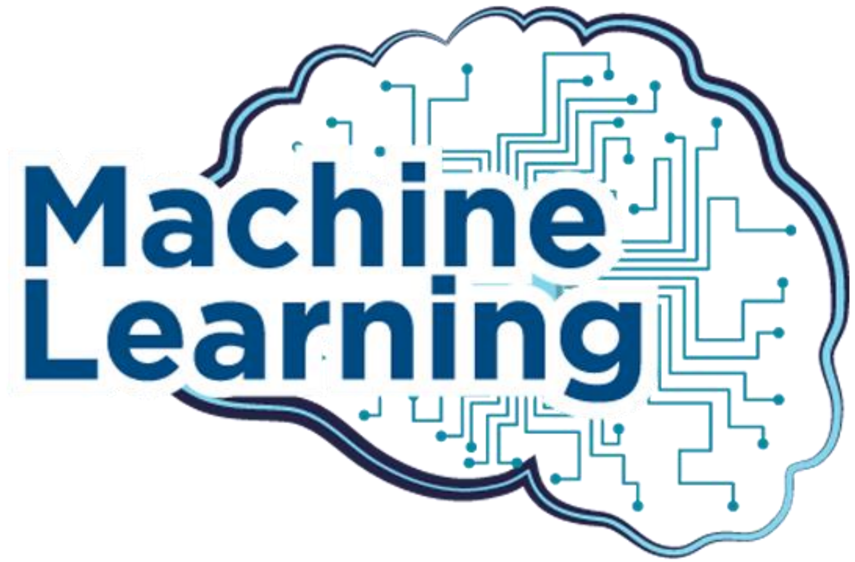
Aspiring Data Scientist

MSc. Industrial Engineering

Summer 2022

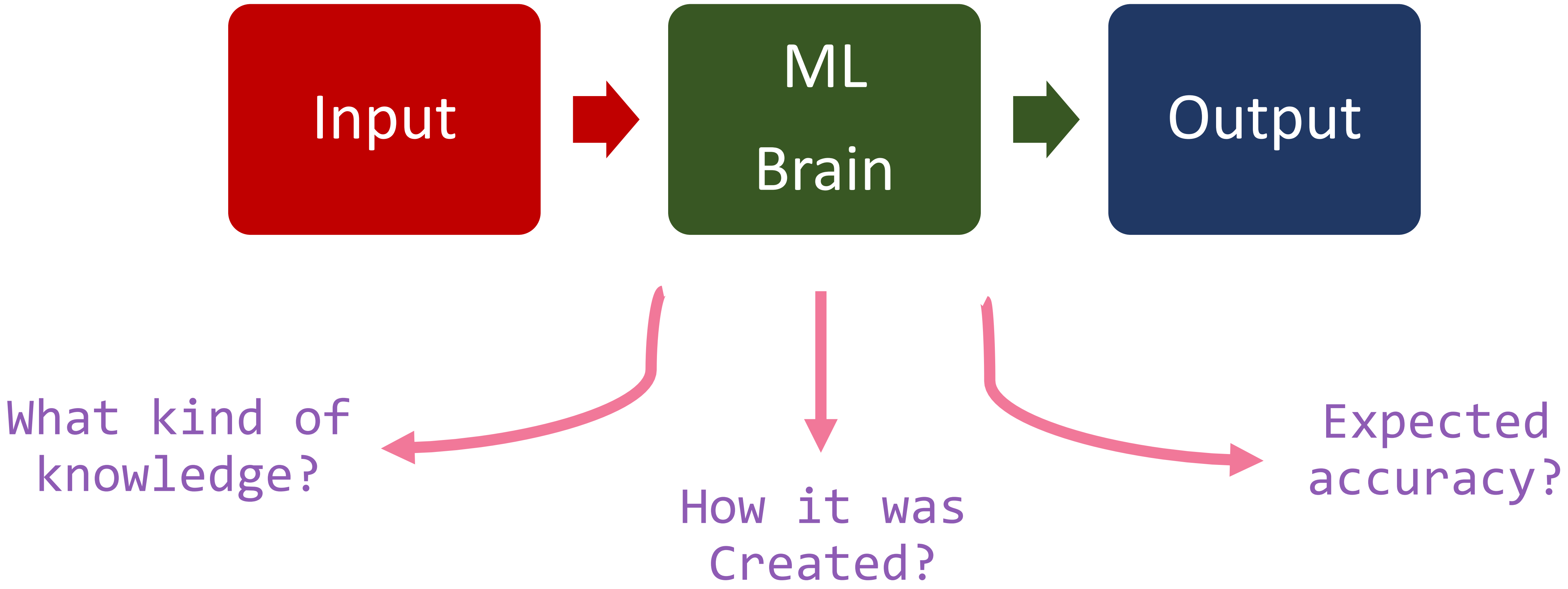
What is Machine Learning?

What is Machine Learning?



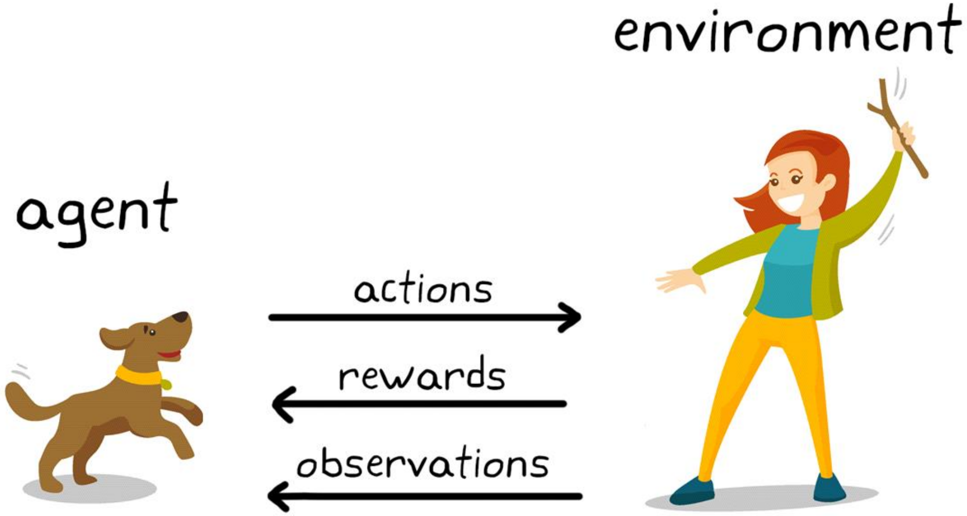
Machine learning is a branch of artificial intelligence and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Introduction to Machine Learning

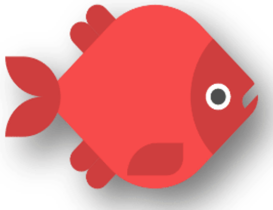


ML Sections

Supervised Learning



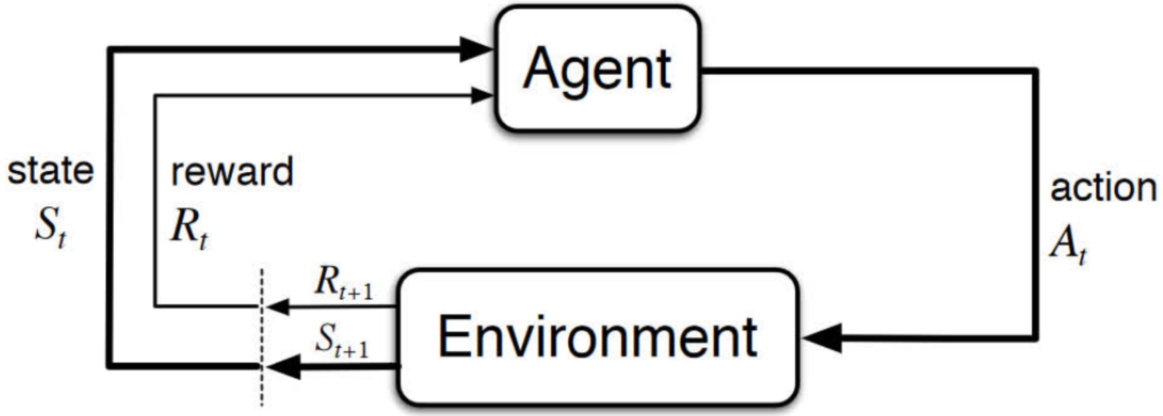
Reinforcement Learning



Unsupervised Learning

What is Reinforcement Learning?

What is Reinforcement Learning?



Reinforcement learning is a simulation-based optimization method which relies on interacting agents to find the optimal (or near-optimal) solution.

Let's break down this diagram into steps.

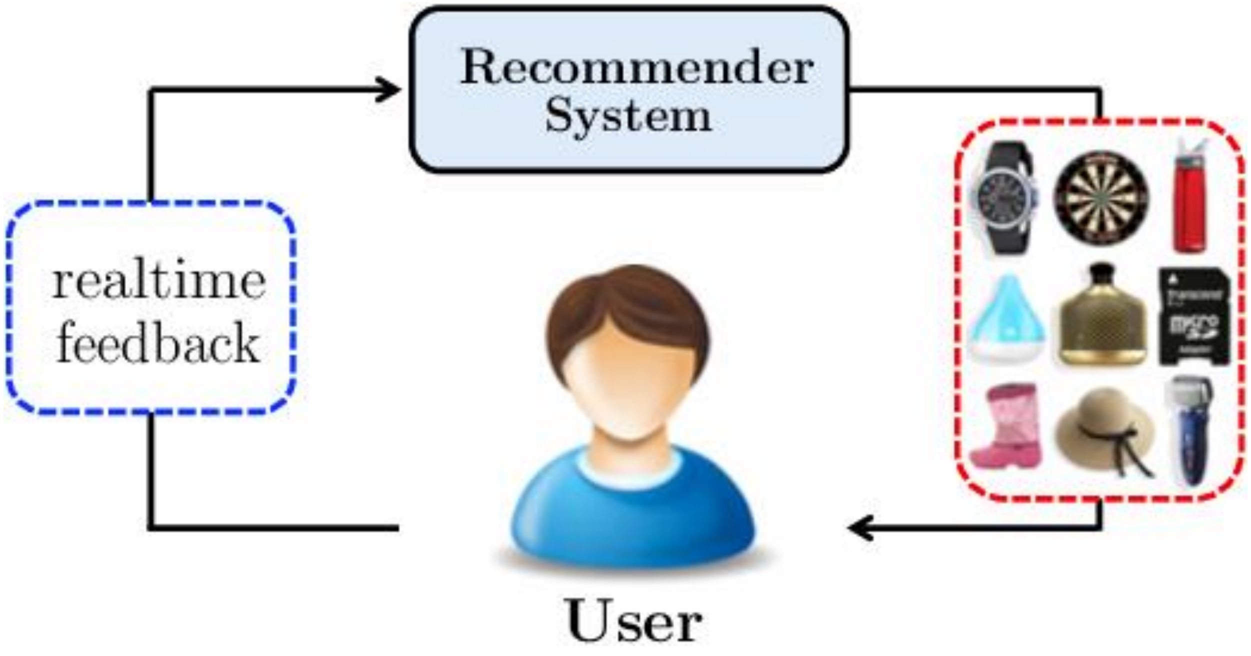
1. At time t, the environment is in state S_t.
2. The agent observes the current state and selects action A_t.
3. The environment transitions to state S_{t+1} and grants the agent reward R_{t+1}.
4. This process then starts over for the next time step, $t+1$.
    - Note, $t+1$ is no longer in the future, but is now the present. When we cross the dotted on the bottom left, the diagram shows $t+1$ transforming into the current time step t sc that S_{t+1} and R_{t+1} are now S_t and R_t.

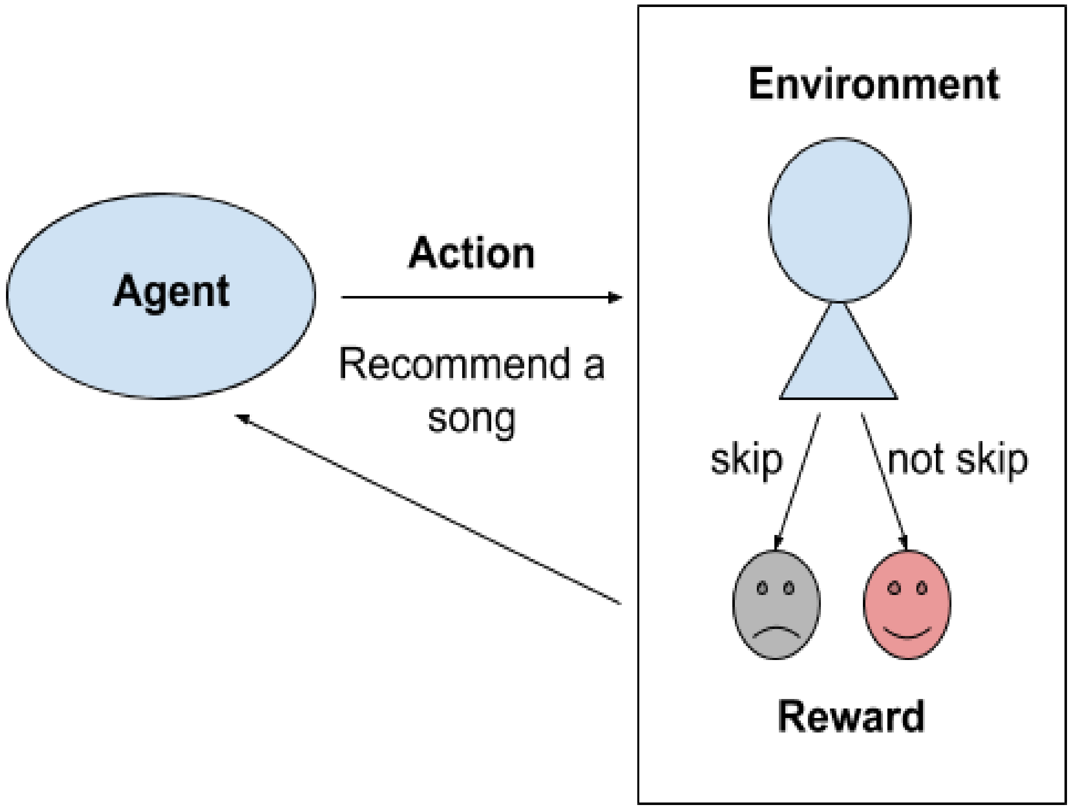
RL comprises two main elements, the agent (learner) and the learning environment

Interaction between Game theory and Reinforcement Learning

Interaction between Recommendation Systems and Reinforcement Learning

What is Reinforcement Learning?





Elements of RL

1. **Agent** — the learner and the decision maker.

2. **Environment** — where the agent learns and decides what actions to perform.

3. **Action** — a set of actions which the agent can perform.

4. **State** — the state of the agent in the environment.

5. **Reward** — for each action selected by the agent the environment provides a reward. Usually a scalar value.

Bellman Equation

Source: Martin L. Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994. ISBN 0471619779.

Dynamic Programming Iterations:

$$
\begin{aligned}
v_{k+1}(s) &\doteq \max_a \mathbb{E}[R_{t+1} + \gamma v_k(S_{t+1}) \mid S_t = s, A_t = a] \\
&= \max_a \sum_{s',r} p(s',r|s,a)\Big[r + \gamma v_k(s')\Big]
\end{aligned}
$$

Bellman Equation:

$$
\begin{aligned}
v_*(s) &= \max_a \mathbb{E}[R_{t+1} + \gamma v_*(S_{t+1}) \mid S_t = s, A_t = a] \\
&= \max_a \sum_{s',r} p(s',r|s,a)\Big[r + \gamma v_*(s')\Big]
\end{aligned}
$$

Exploitation Vs Exploration

Environments

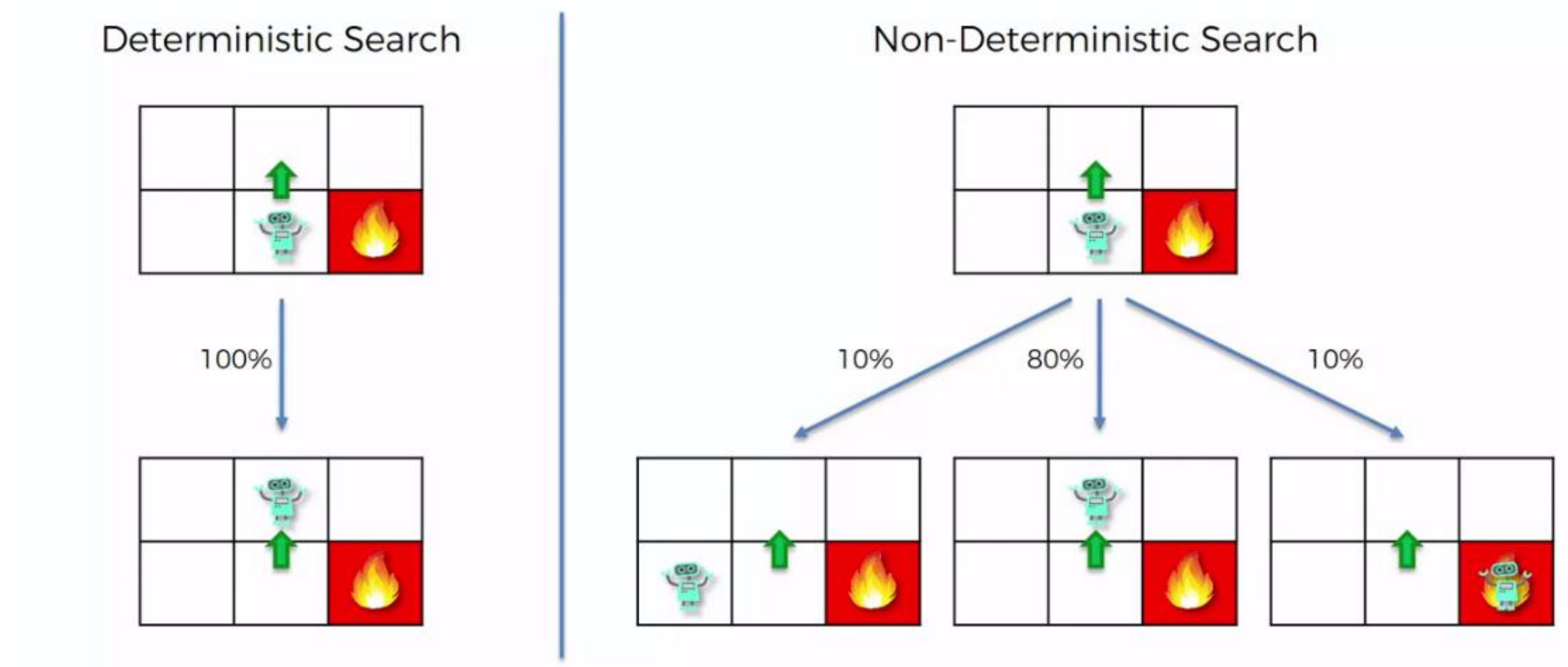
Environments in RL

- Deterministic
- Stochastic

Environments

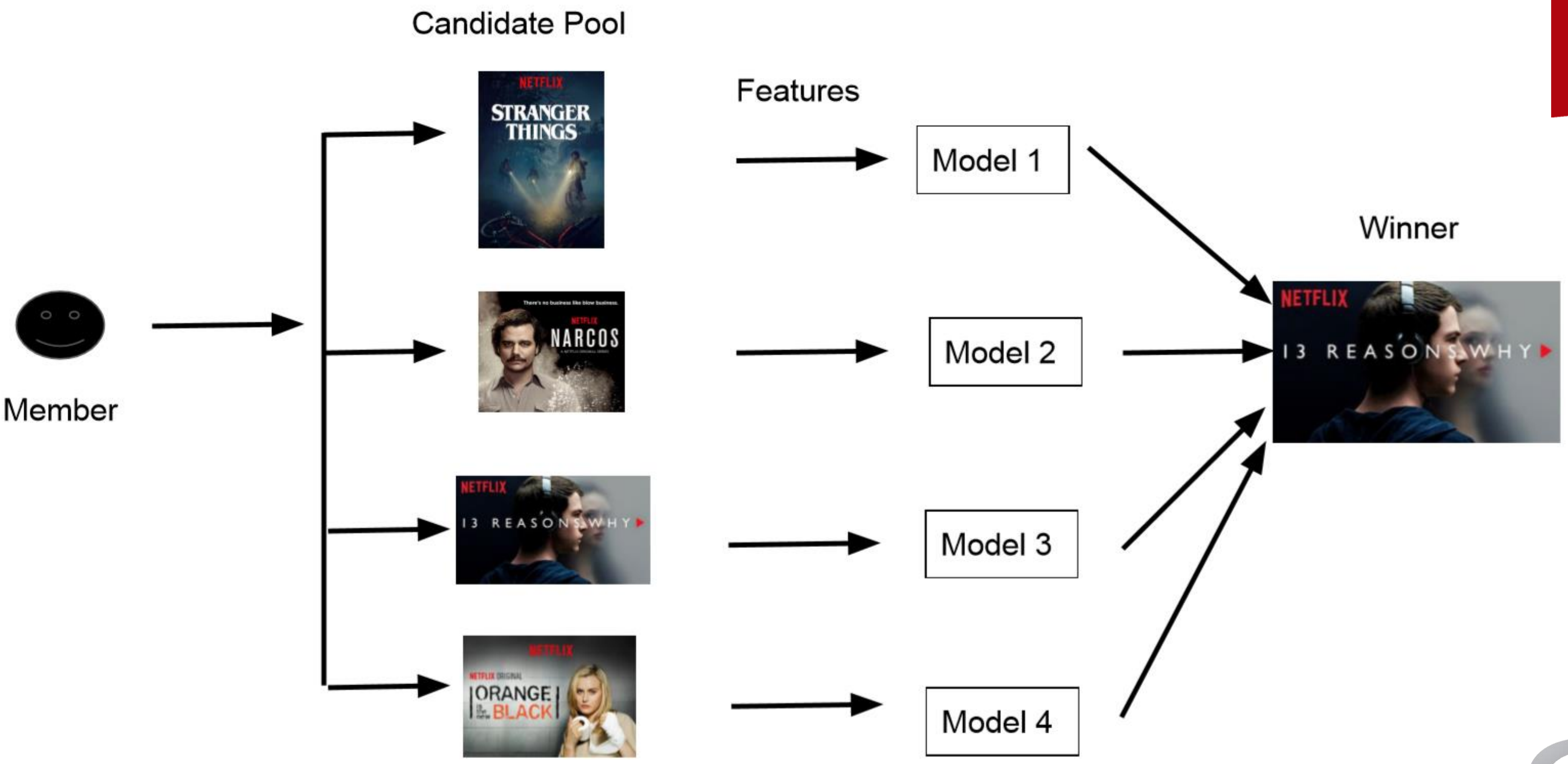
Exploitation

Exploration



RL Real Cases

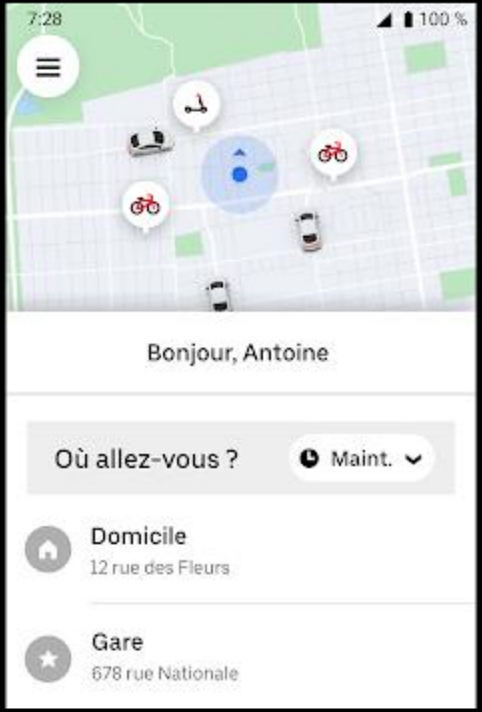
Case Netflix



ML Cases

Find Best Price

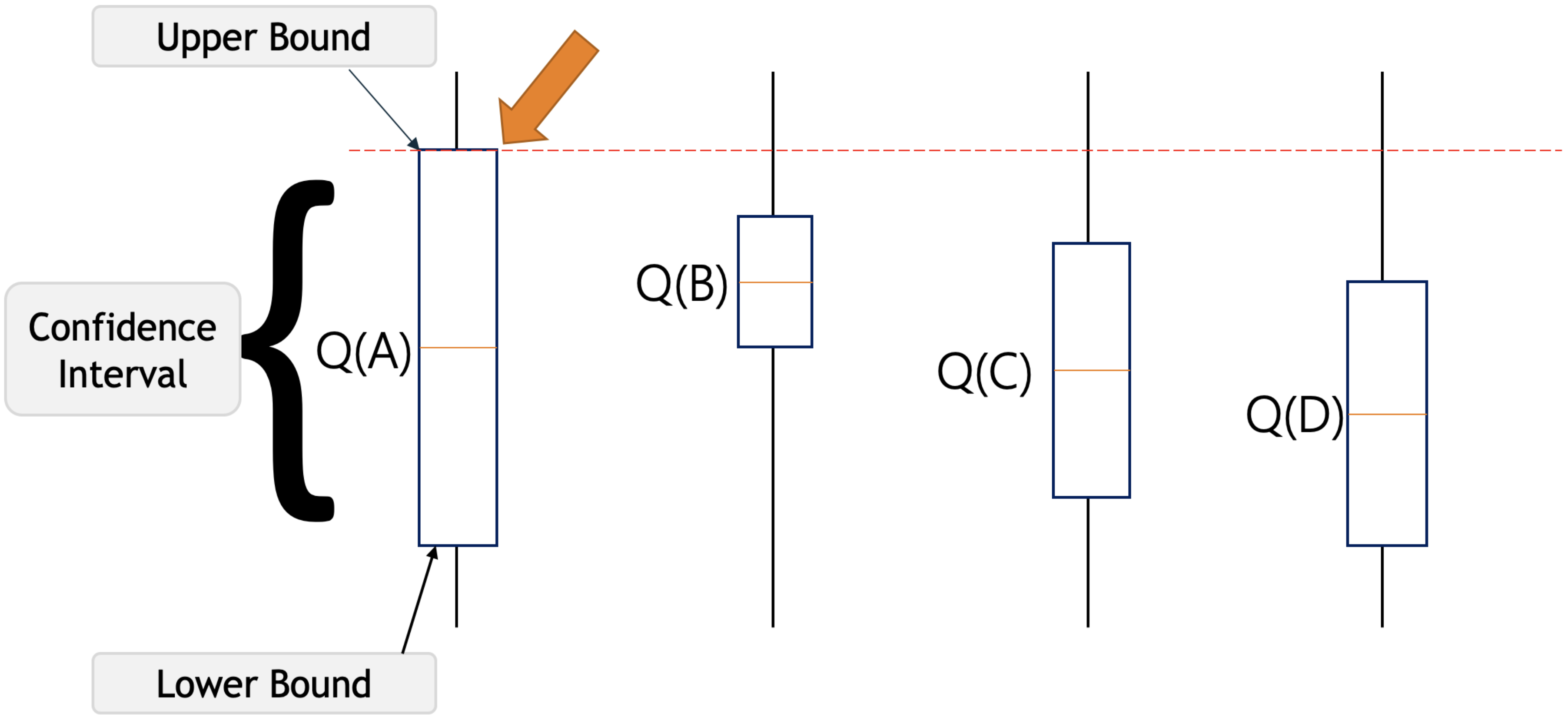
Optimize Waiting Time and Idle Vehicle



UBER

Upper Confidence Bound (UCB)

Upper Confidence Bound (UCB)



Upper Confidence Bound is a Reinforcement Learner

Upper Confidence Bound (UCB)

1. *At each round* ***n****, we consider two numbers for machine m.*
   -> $N_m(n)$ = number of times the machine m was selected up to round n.
   -> $R_m(n)$ = number of rewards of the machine m up to round n.

2. *From these two numbers we have to calculate,*
   a. The average reward of machine m up to round n, $r_m(n) = R_m(n) / N_m(n)$.
   b. The confidence interval $[\, r_m(n) - \Delta_m(n),\ r_m(n) + \Delta_m(n) \,]$ at round n with, $\Delta_m(n) = \text{sqrt}(\ 1.5 * \log(n) / N_m(n)\ )$

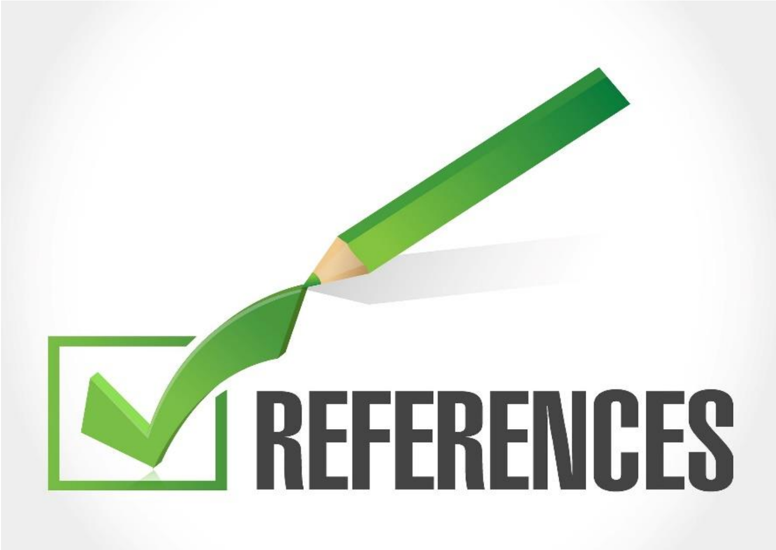
3. *We select the machine* ***m*** *that has the maximum UCB,* $(\, r_m(n) + \Delta_m(n) \,)$

References

Martin L. Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994. ISBN 0471619779.

Decision Theory: Single & Sequential Decisions. VE for Decision Networks .Alan Mackworth, UBC CS 322 – Decision Theory, 2013.

Sutton R. Barto, A. Reinforcement Learning: An Introduction. The MIT Press., Cambridge, Massachusetts, UK, 2nd edition, 2018.



Conclusion and Questions?


THANK YOU!