Overview of Machine Learning:

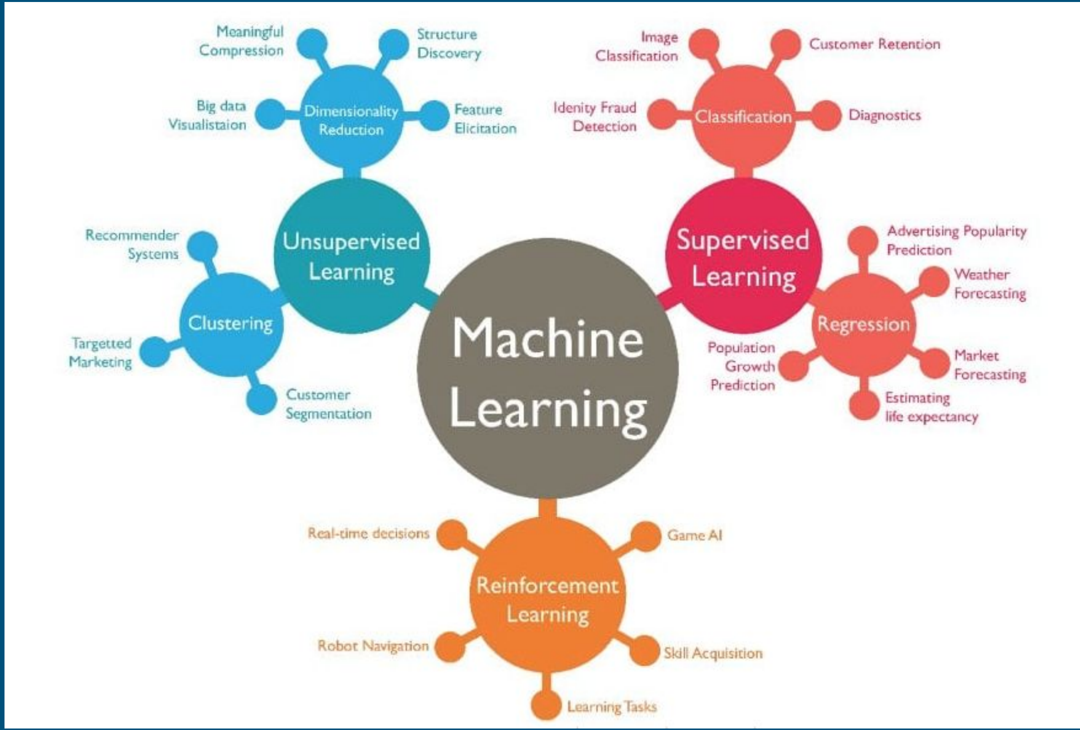
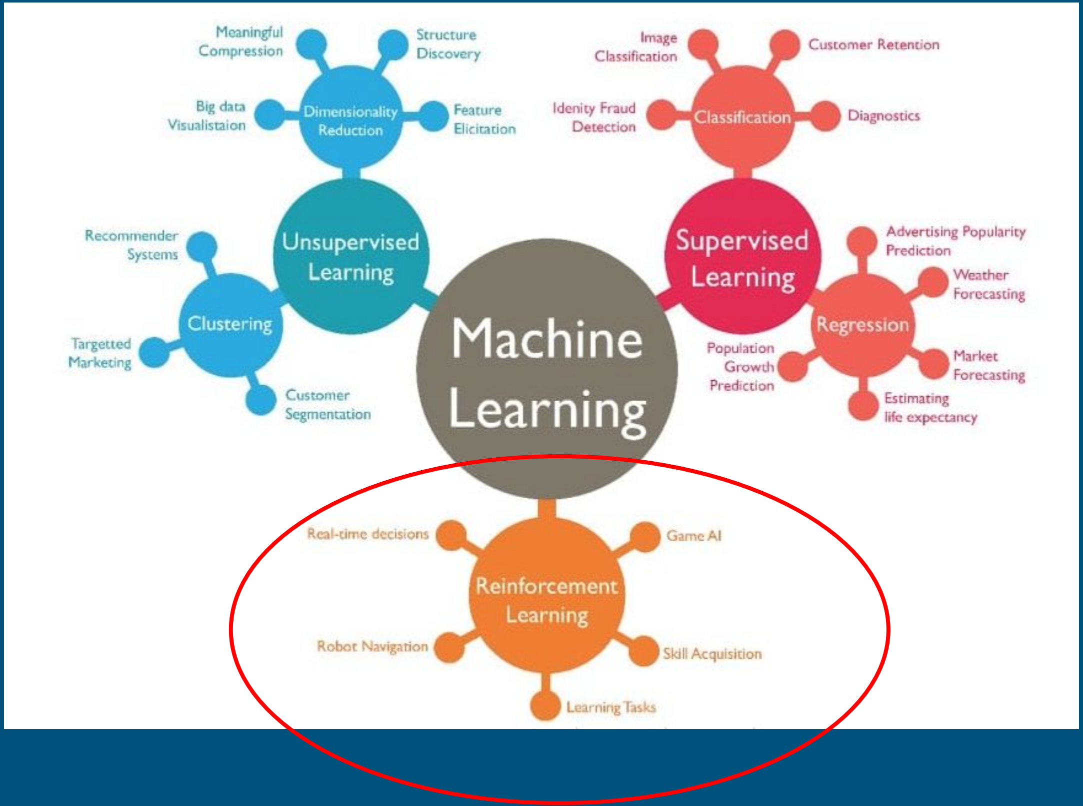


Image source:
https://datasciencedojo.com/blog/machine-learning-101/#

Overview of Machine Learning:

Reinforcement Learning



Reinforcement Learning Basics

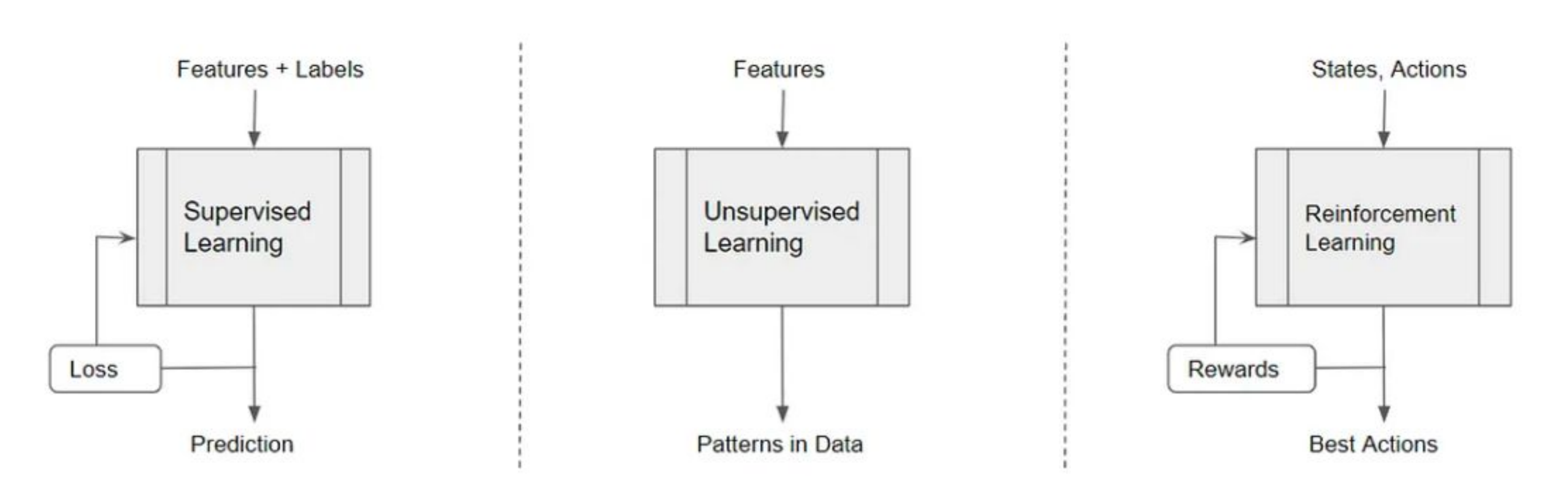
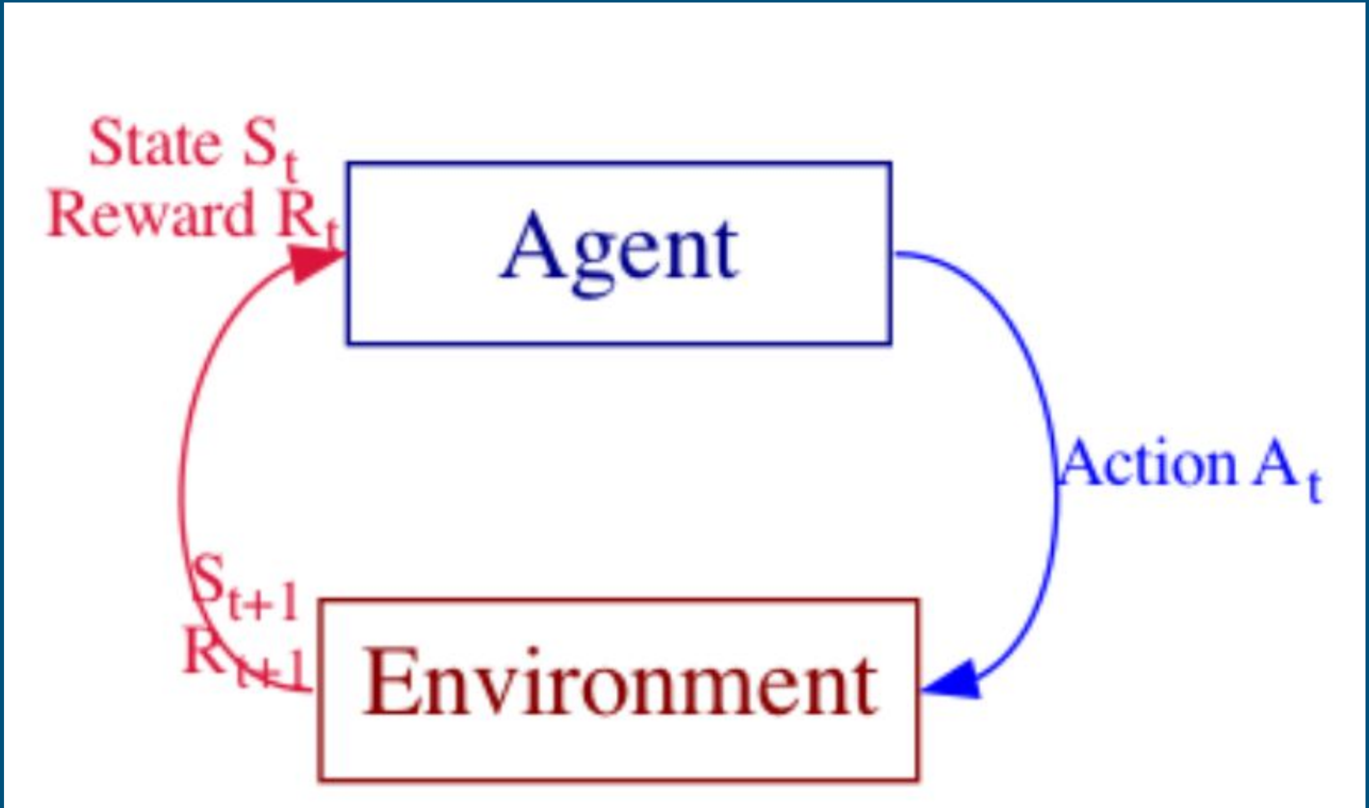


Image source:
https://towardsdatascience.com/reinforcement-learning-made-simple-part-1-intro-to-basic-concepts-and-terminology-1d2a87aa060

Reinforcement Learning Basics



Reinforcement Learning Basics

Environment:Physical world in which the agent operates
State — Current situation of the agent
Reward — Feedback from the environment
Policy — Method to map agent's state to actions
Agent: Decision Maker that take action

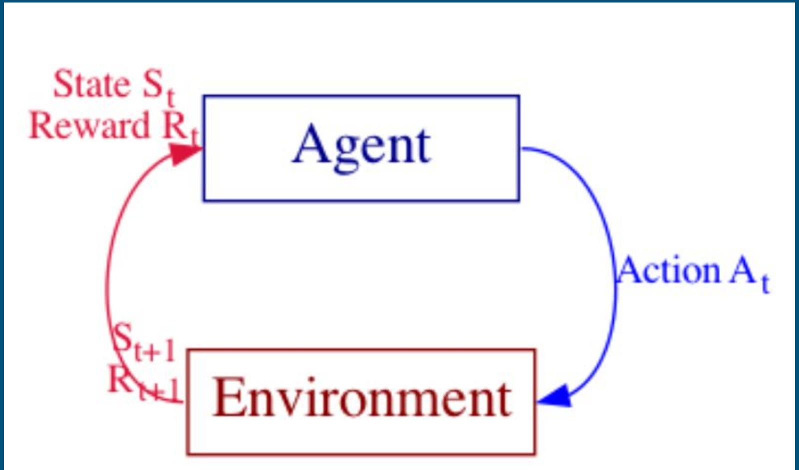


Image source:
"Foundation of Reinforcement Learning", Rao Jelvis

Reinforcement Learning Basics

Environment:Physical world in which the agent operates
State — Current situation of the agent
Reward — Feedback from the environment
Policy — Method to map agent's state to actions
Agent: Decision Maker that take action

Image source:
https://towardsdatascience.com/reinforcement-learning-made-simple-part-1-intro-to-basic-concepts-and-terminology-1d2a87aa060

Reinforcement Learning: Main Definitions

Episodes: As Series of Atomic Experiences

State, Actions, Reward, New State, Action...

$$(S_0, A_0, R_1, S_1, A_1, R_2, \cdots, S_{T-1}, A_{T-1}, R_T, S_T)$$

Reinforcement Learning: Main Definitions

Policy: Mapping the State to Action

$$\pi(a|s) = P(A_t = a|S_t = s)$$

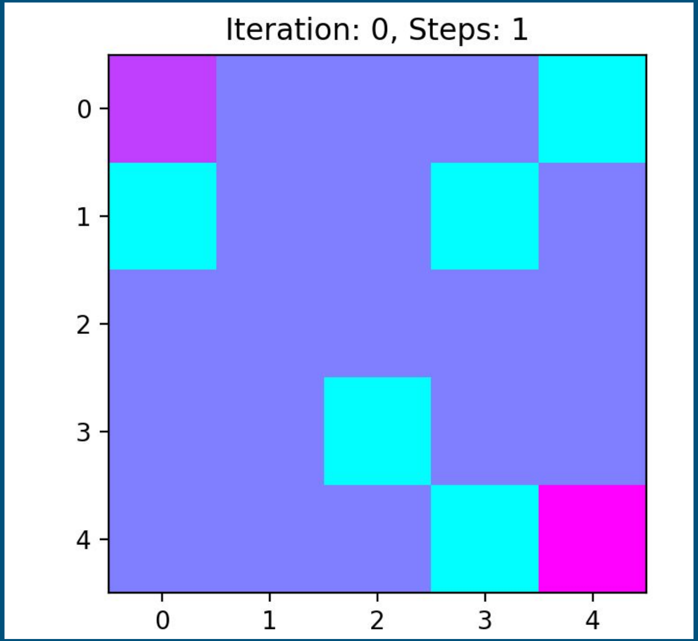
Reinforcement Learning: Main Definitions

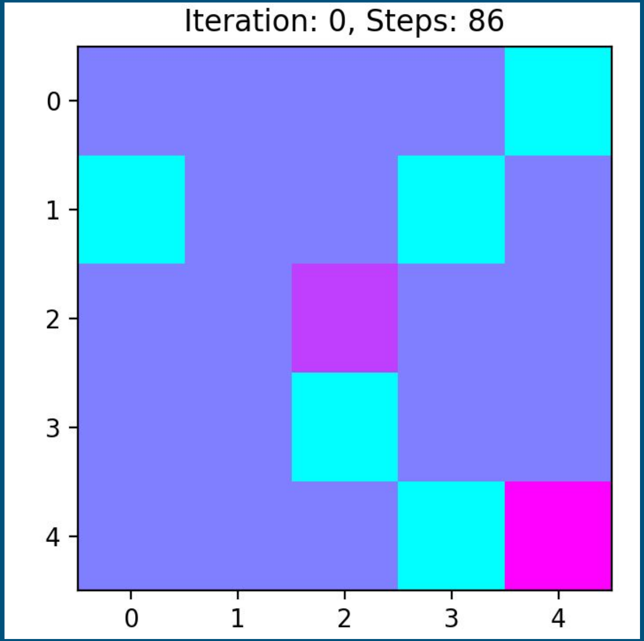
The Learning Eq: Q-learning

$$Q^{new}(S_t, A_t) \leftarrow (1 - \underbrace{\alpha}_{\text{learning rate}}) \cdot \underbrace{Q(S_t, A_t)}_{\text{current value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \bigg(\overbrace{\underbrace{R_{t+1}}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(S_{t+1}, a)}_{\text{estimate of optimal future value}}}^{} \bigg)$$

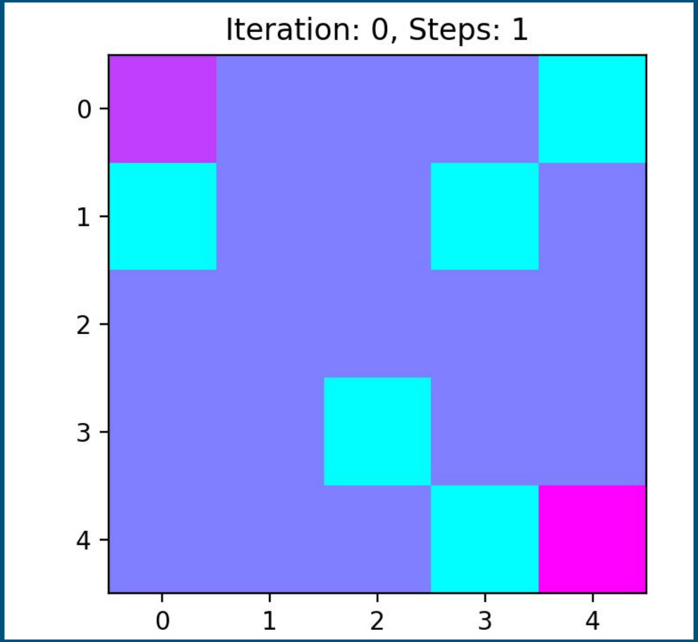
$$\underbrace{R_{t+1} + \gamma \cdot \max_a Q(S_{t+1}, a)}_{\text{new value (temporal difference target)}}$$

RL Example:
Coding Example:





RL Example:
Coding Example:



Reward Structure:

Reward Arriving at goal: **+1**
Rewards At each step: **-0.1**
Reward Hitting Obstacles: **-1**

Available Actions:

0: Up, 1: Right, 2: Down, 3: Left

Reinforcement Learning for Reservoir Management

Deep reinforcement learning for optimal well control in subsurface systems with uncertain geology

Yusuf Nasir *, Louis J. Durlofsky

Department of Energy Science and Engineering, Stanford University, Stanford, CA 94305, USA

A R T I C L E I N F O

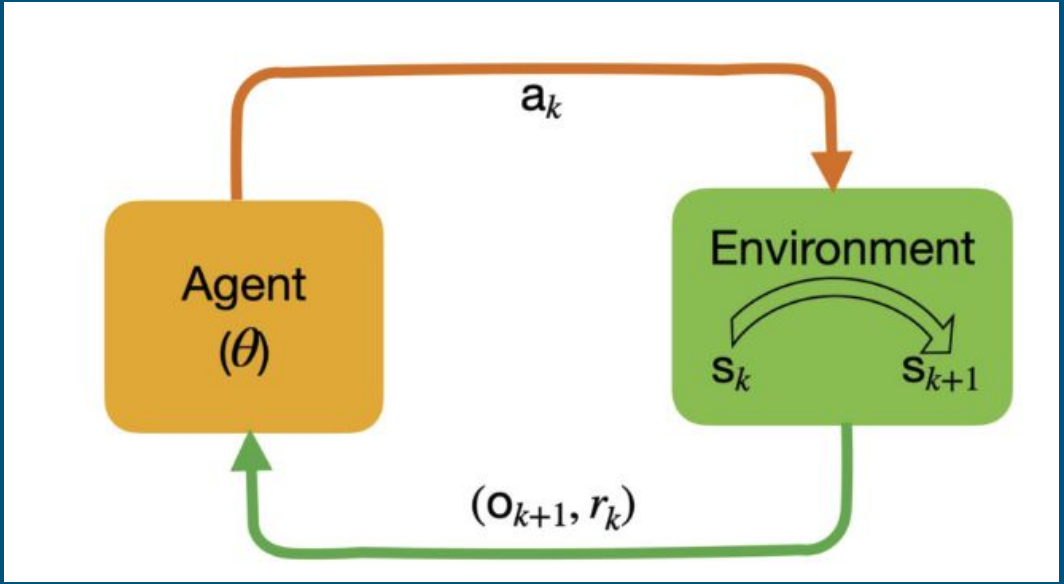
Article history:
Received 23 March 2022
Accepted 14 January 2023
Available online 20 January 2023

Keywords:
Deep reinforcement learning
Closed-loop modeling
Stochastic optimal control
Reservoir simulation
Transformers
Proximal policy optimization

A B S T R A C T

A general control policy framework based on deep reinforcement learning (DRL) is introduced for closed-loop decision making in subsurface flow settings. Traditional closed-loop modeling workflows in this context involve the repeated application of data assimilation/history matching and robust optimization steps. Data assimilation can be particularly challenging in cases where both the geological style (scenario) and individual model realizations are uncertain. The closed-loop reservoir management (CLRM) problem is formulated here as a partially observable Markov decision process (POMDP), with the associated optimization problem solved using a proximal policy optimization algorithm. This provides a control policy that instantaneously maps flow data observed at wells (as are available in practice) to optimal well pressure settings. The policy is represented by a temporal convolution and gated transformer blocks. Training is performed in a preprocessing step with an ensemble of prior geological models, which can be drawn from multiple geological scenarios. Example cases involving the production of oil via water injection, with both 2D and 3D geological models, are presented. The DRL-based methodology is shown to result in an increase in NPV of 15% (for the 2D cases) and 33% (3D cases) relative to robust optimization over prior models, and to an improvement of 2 - 7% in NPV relative to traditional CLRM. The solutions from the control policy are found to be comparable to those from deterministic optimization, in which the geological model is assumed to be known, even when multiple geological scenarios are considered. The control policy approach results in a 76% decrease in computational cost relative to traditional CLRM with the algorithms and parameter settings considered in this work.

Reinforcement Learning for Reservoir Management



Reinforcement Learning for Reservoir Management

5. Concluding remarks

The new framework was applied to 2D and 3D example cases. In the 2D case, binary channelized geological models, corresponding to realizations drawn from a single geological scenario, were considered. The training of the control policy required 135,000 total flow simulations, which is equivalent to 500 sequential simulations in a fully parallelized setting. This represents only 24% of the simulations required for traditional CLRM (using the algorithms and parameter values considered in this study). The DRL-based approach was shown to provide solutions close to those from deterministic optimization of individual geological realizations. This is a significant finding, as deterministic optimization is not possible in practice because geological uncertainty is always present. Our results clearly demonstrated the advantages of the control policy approach relative to both robust optimization over prior geological models and to the traditional CLRM approach. Specifically, the control policy approach provided average improvement of 14.7% in NPV relative to robust (prior) optimization, and to an increase of 2 - 7% compared to traditional CLRM.