21AI638 Reinforcement Learning
21AM711 Reinforcement Learning

Course Policies

Syllabus

- **Unit 1**
- Introduction to Machine Learning and its various types, Motivation and Introduction to Reinforcement Learning, Multi arm Bandits ; Markov Decision Process, Value functions; Dynamic programming : Policy evaluation and improvement, Value iteration and Policy iteration algorithms
- **Unit 2**
- Value prediction problems : Temporal difference learning in finite state spaces Algorithms for large state spaces Control : Closed loop interactive learning, online and active learning in bandits, Q learning in finite MDPs, Q learning with function approximation
- **Unit 3**
- On policy approximation of action values : Value Prediction with Function Approximation, GradientDescent Methods, Policy approximation : Actor critic methods, Monte Carlo Methods : Monte carlo prediction, estimation of action values, off policy prediction via importance sampling

TEXTBOOKS/REFERENCES

1. Sutton and Barto, Reinforcement Learning: An Introduction, The MIT Press Cambridge, Massachusetts London, England, 2015
2. Csaba Szepesvari, Algorithms for Reinforcement Learning, Morgan & Claypool, United States, 2010

Course Outcomes (CO)

COs	Description
CO1	Understand the relevance of Reinforcement Learning and how does it complement other ML techniques.
CO2	Understand various RL algorithms.
CO3	Formulate a problem as a Reinforcement Learning problem and solve it
CO4	Implement RL algorithms using Python

21AM711 Reinforcement Learning 2-0-2-3

Evaluation Pattern - 70:30

Midterm Exam - 20%

Lab Assignments – 25%

Project – 25%

End Semester Exam - 30%

21AI638 Reinforcement Learning 3-0-2-4

- Evaluation Pattern - 70:30
- Midterm Exam - 20%
- Lab Assignments – 25%
- Project – 25%
- End Semester Exam - 30%

Overview of Machine Learning

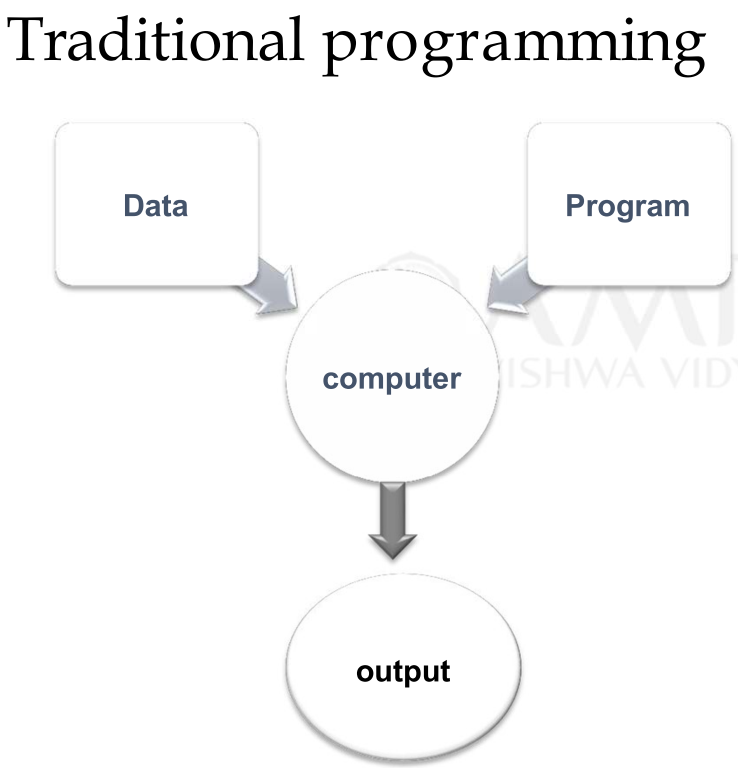
- Overview of ML
- Applications
- Types of ML

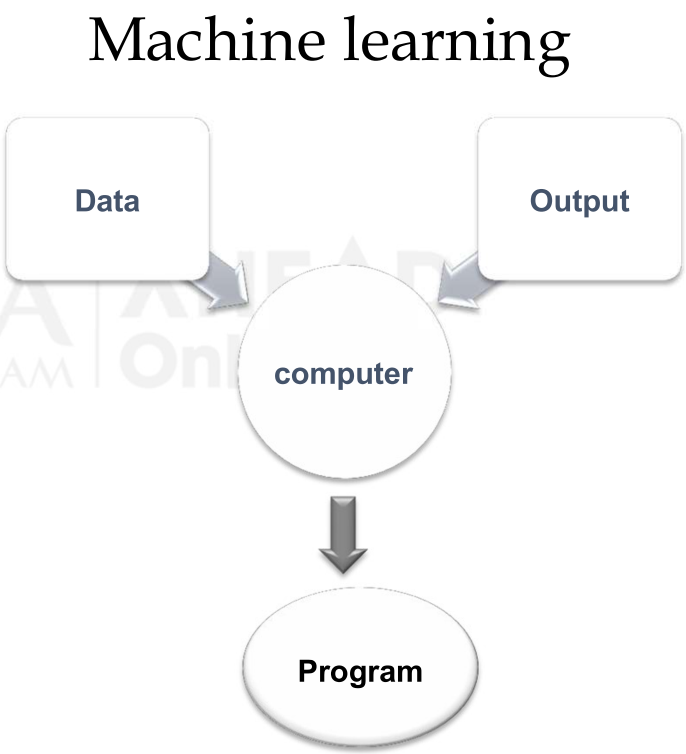
Machine Learning

- Artificial intelligence: Makes a computer system to mimic human intellect.
- Machine learning: Learn from data or experiences without being explicitly programmed.
- Machine learning models map inputs to the outputs of the given dataset
- "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" : Tom Mitchell
- E.g. Facebook photo tagging

Mitchell, Tom, and Machine Learning McGraw-Hill. "Edition." (1997).

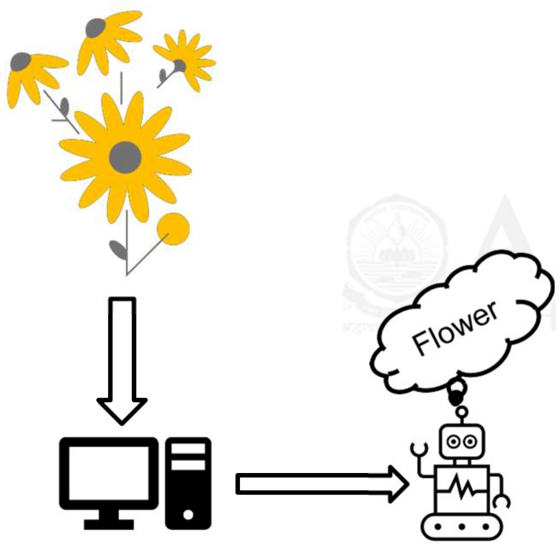
Traditional Programming vs Machine Learning



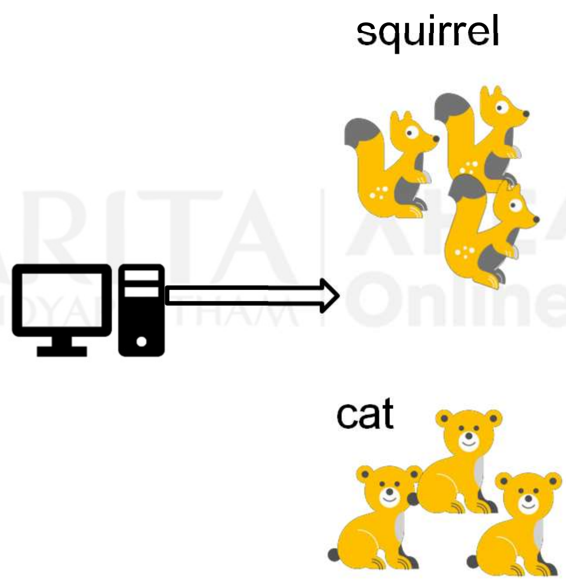


Types of machine learning

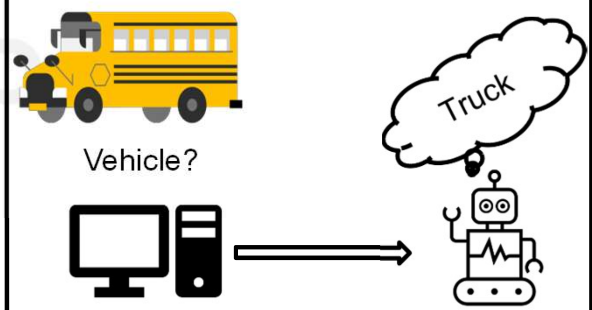
1.Supervise Learning
uses labeled data



2. Unsupervised Learning
uses unlabeled data

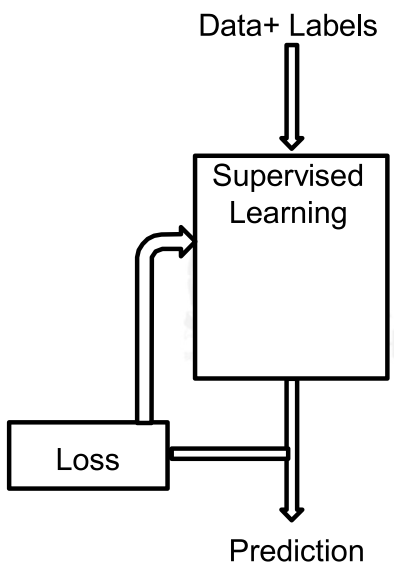


3. Reinforcement Learning
uses an agent and an environment to produce actions and rewards

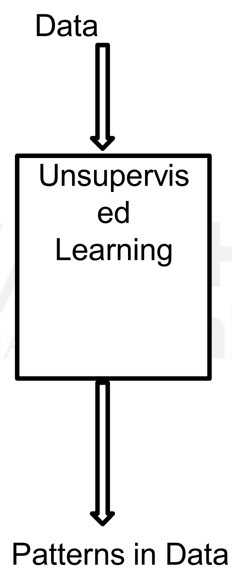


Types of machine learning

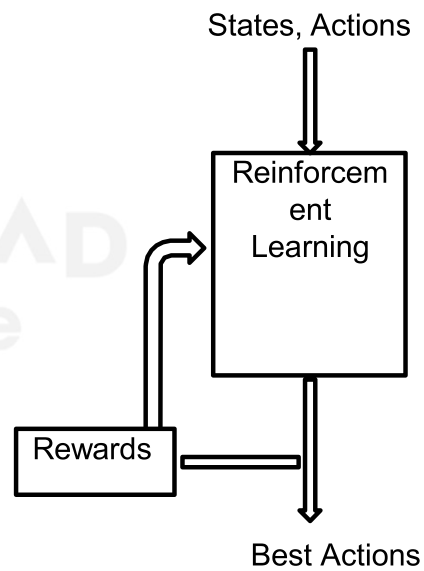
Supervised Learning



Unsupervised Learning



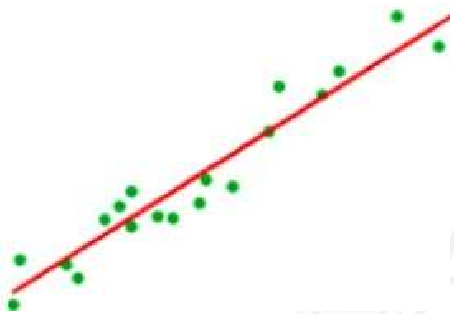
Reinforcement Learning



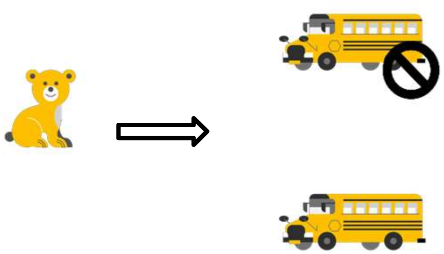
Type of Problems

1.Supervise Learning

A. Regression

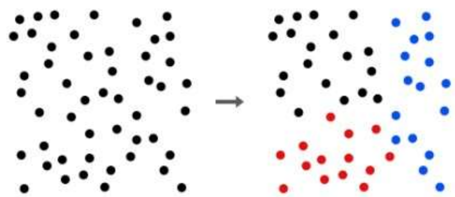


B. Classification



2. Unsupervised Learning

A. Clustering



B. Association

If customer purchased an item "A" ⟹ Then recommend item "B"

3. Reinforcement Learning

A. Reward Based

Applications of Machine Learning

- Recommendation Systems
- Healthcare
- Online Advertising
- Stock Market Trading
- Speech Recognition
- Virtual Personal Assistants
- Autonomous Vehicles
- Sentiment Analysis
- Smart IoT

Take away

- What is machine learning?
- How ML different from traditional programming?
- Types of ML algorithms
- Applications of ML

Introduction to Reinforcement Learning

Objectives

- What is RL?
- Why and where we use RL?
- How RL different from other ML methods?

What is Reinforcement Learning?

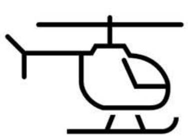
Learn to make good sequence of decisions

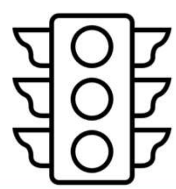
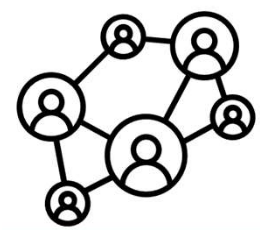
Trial-and-error Learning

Sequential Decision Making

Sequential Decision Making

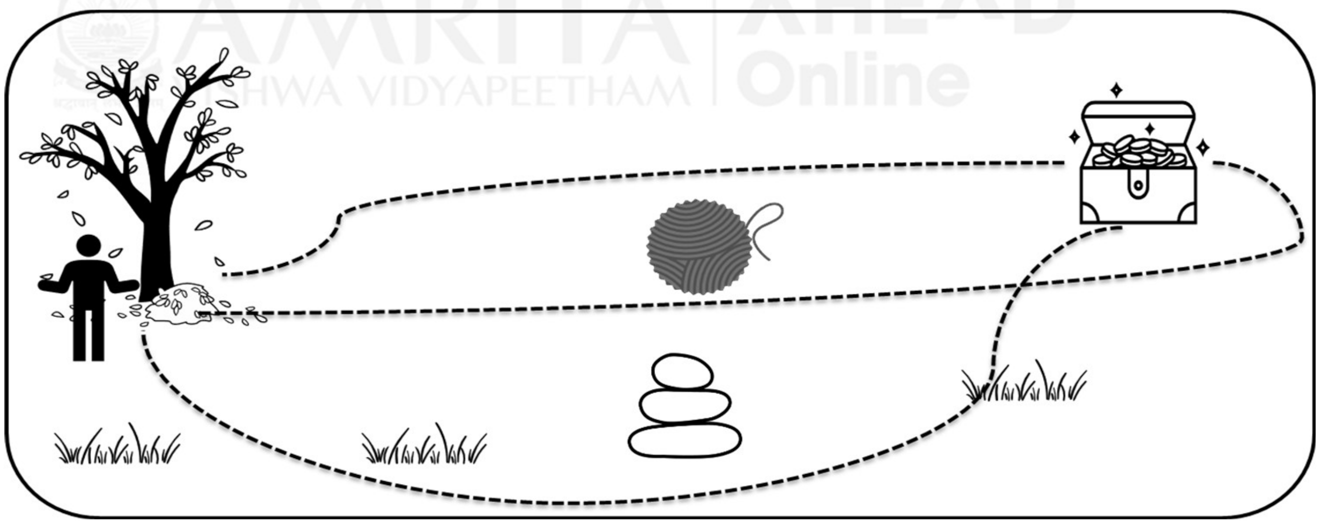
- Series of decisions over time
- Decision outcomes may depend on environmental factors
- Final goal depends on many interactive decisions and their random consequences
- Examples:
  - § Traffic signal control
  - § Communication Network Packet Routing
  - ▪ Autonomous Vehicles



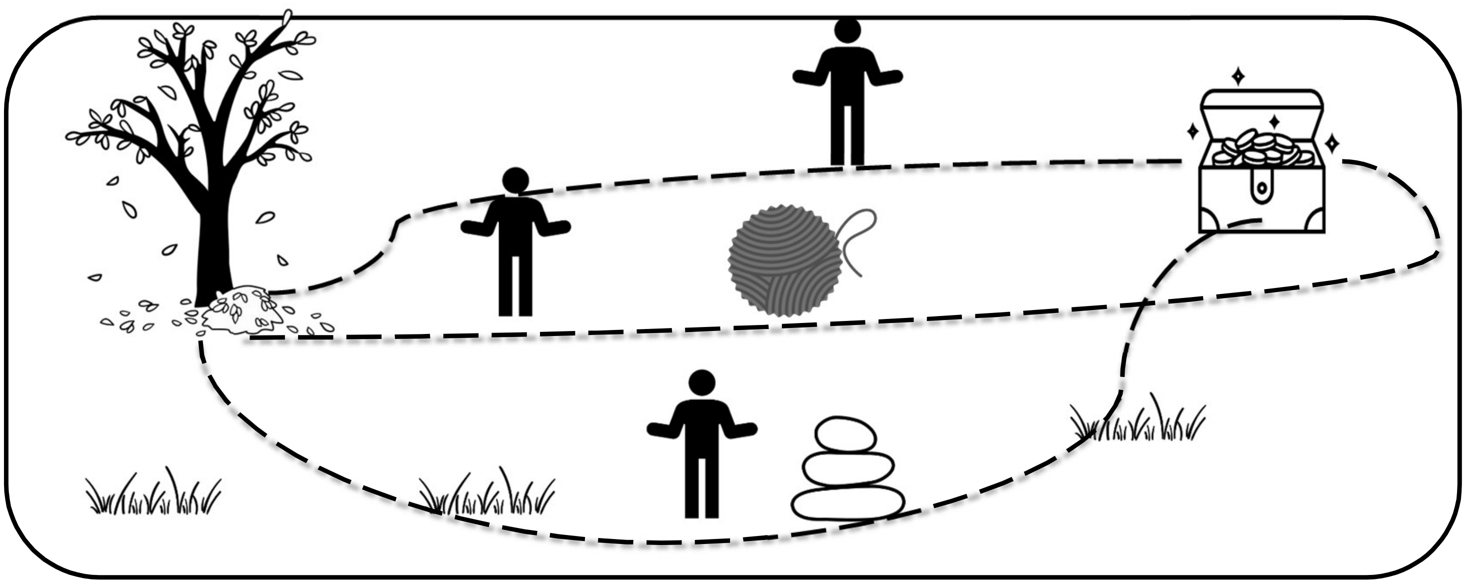
What is RL?

- Science of decision making
- Discover the sequence of actions – trial and error
- Learns the optimal behavior through interactions with the environment
- Actions receive a reward or penalty from the environment



What is RL?

- After many attempts the robot learns the best path



How RL different from SL and USL?

- There is no supervisor to guide the training
- Not required to train with a large (labeled or unlabeled) dataset.
- Data is provided dynamically via feedback from the real-world environment with which you are interacting.
- Make decisions over a sequence of time-steps
- Work in dynamic and uncertain environments

Comparison of SL, USL and RL

Supervised Learning	Unsupervised Learning	Reinforcement Learning
Labelled data with target	Unlabelled data without target	Input data not predefined: learns from environment using rewards and penalty
External Supervision	No supervision	Feedback signals
Learn pattern in data and its labels	Learn to group data	Compute best reward to reach goal from start state
Map input data to known labels	Find similar features in data and understand patterns	Maximize rewards following trail and error approach
Model training prior to testing	Model training prior to testing	Model training and testing simultaneously
E.g. Regression and classification problems	E.g. Association mining and clustering	E.g. Reward based problems planning, control

Real World Applications of RL

Smart Vehicles	• Self Driving Cars • Autonomous Helicopters
Games	• Atari • Alpha Go
Robotics	• Navigation • Surveillance
Healthcare	• Manage Critical Diseases • Adaptive Treatment Plans
Finance	• Stock market • Portfolio optimizations
Smart Ads	• Personalized Ads • Recommendation Systems
Chatbots	• Siri • Alexa

Take away

- Trial and error learning
- Difference from other ML techniques
- Real-world applications of RL

History and Characteristics of RL

Objectives

- History of RL
- RL Characteristics
- Challenges

History of Reinforcement Learning

- Originates from trial-and-error learning & optimal control
- 1911: Thorndike described trial-and-error learning with the "Law of effect" - reinforce the satisfying activities and deter from distracting activities
- 1927 : Pavlov's experiment with dogs - dog learned to associate sound with the presentation of food
- 1933 : Thomas Ross built a machine to navigate through a maze
- 1961: Minsky addressed the Credit Assignment Problem
- 1950: Optimization methods to find solutions in control problems

History of Reinforcement Learning

1911
- Thorndike psychology of animal learning.
- Law of Effect.

1950s-1960s
- Richard Bellman developed the optimal return function
- Introduced Markov Decision Processes

1960s-1970s
- Minsky's credit assignment

1970s-1980s
- Learning automata.

1990s
- Temporal difference learning

2000s - 2020s
- Deep Reinforcement Learning and Deep Q-learning
- Google DeepMind AlphaGo
- Atari games, Autonomous driving, Robotics

Characteristics of RL

- Trial-and-error search
- Sequential decision making- time plays an important role
- Delayed rewards
- Environment is stochastic

Challenges in RL

- Exploration and exploitation
- Reward design
- Partial observability
- Scalability
- Stochastic environment

Takeaways

- History
- Characteristics
- Challenges

Elements of Reinforcement Learning

Objectives

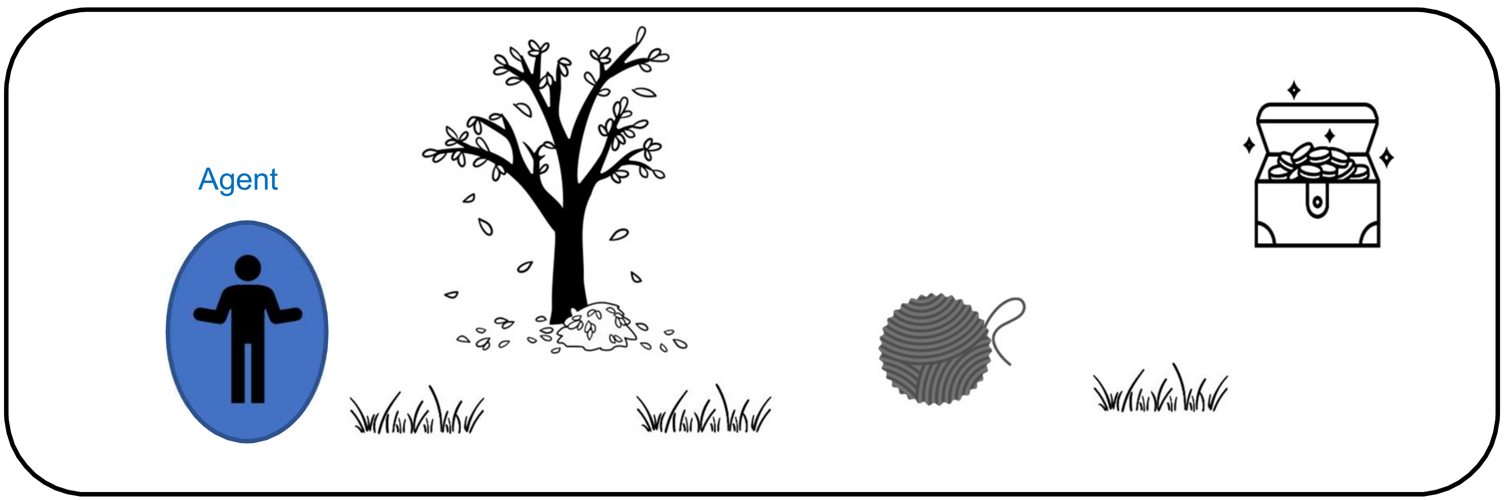
- Components of RL
- How RL works?

Elements of RL

1. Agent
2. Environment
3. State
4. Action
5. Reward
6. Value Function
7. Policy

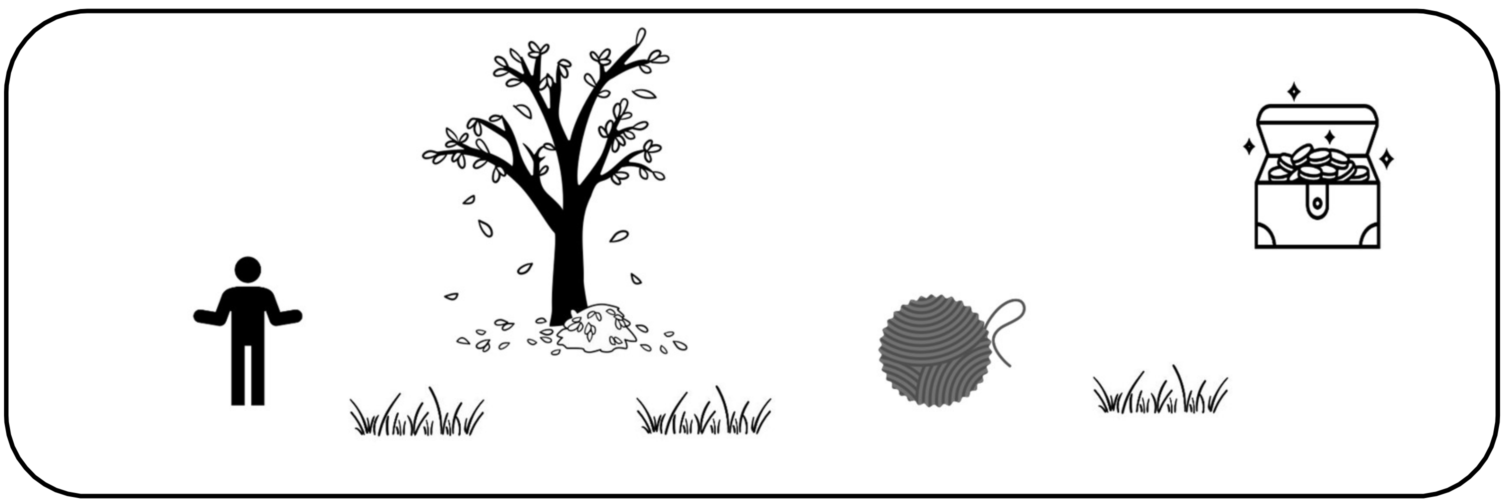
RL Elements : Agent

Agent: Learner in RL problem



RL Elements: Environment

- Environment: training situation
- The real-world environment with which the agent interacts as part of its operation



Types of Environments in RL

Deterministic and Stochastic Environments

- Deterministic Environment : The next state can be predicted from the current state and the actions.
- Stochastic Environment: The next state of the environment can not be predicted from the current state and action.

Types of Environments in RL

Single Agent and Multi-Agent Environments

- Single Agent Environment : One agent in an environment.
- Multi-Agent Environment: Multiple agents interacting with the environment

Types of Environments in RL

Discrete and Continuous Environment

- Discrete Environment: Action space is discrete
- Continuous Environment: action space is continuous.

Types of Environments in RL

Fully Observable and Partially Observable

- Fully Observable : Agent knows full status of the environment
- Partially Observable: Agent does not know the entire state of the environment

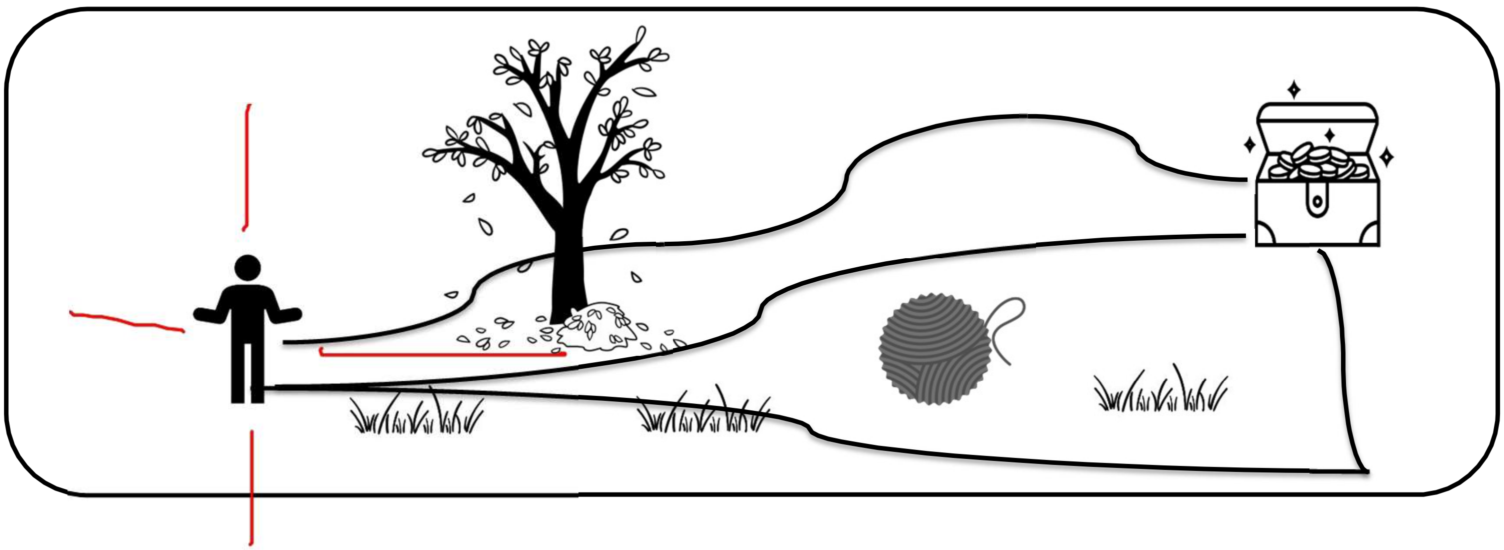
Types of Environments in RL

Sequential and Episodic Environment

- Sequential Environment: agent’s current action is related to previous actions
- Episodic Environment: agent's actions are limited to the current episode

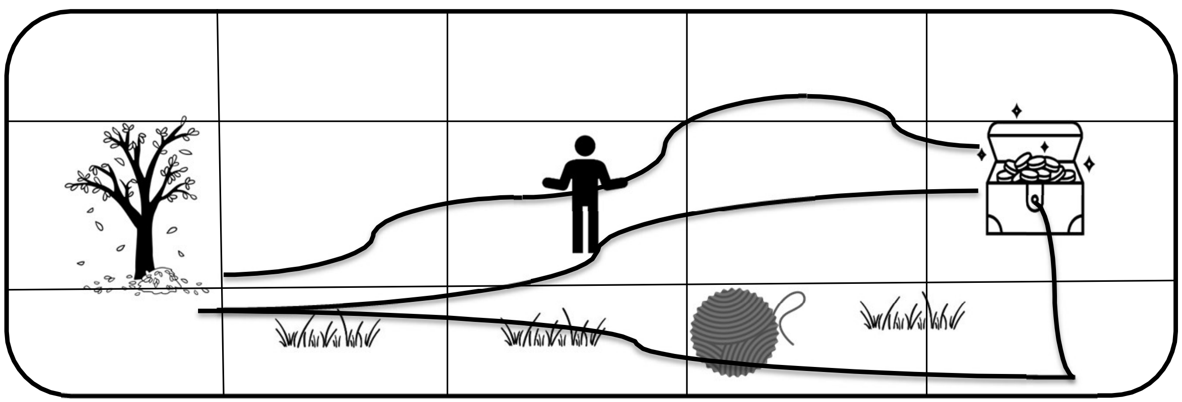
RL Elements: Action

Action : Possible steps taken by an agent within the environment based on its observation



RL Elements : State

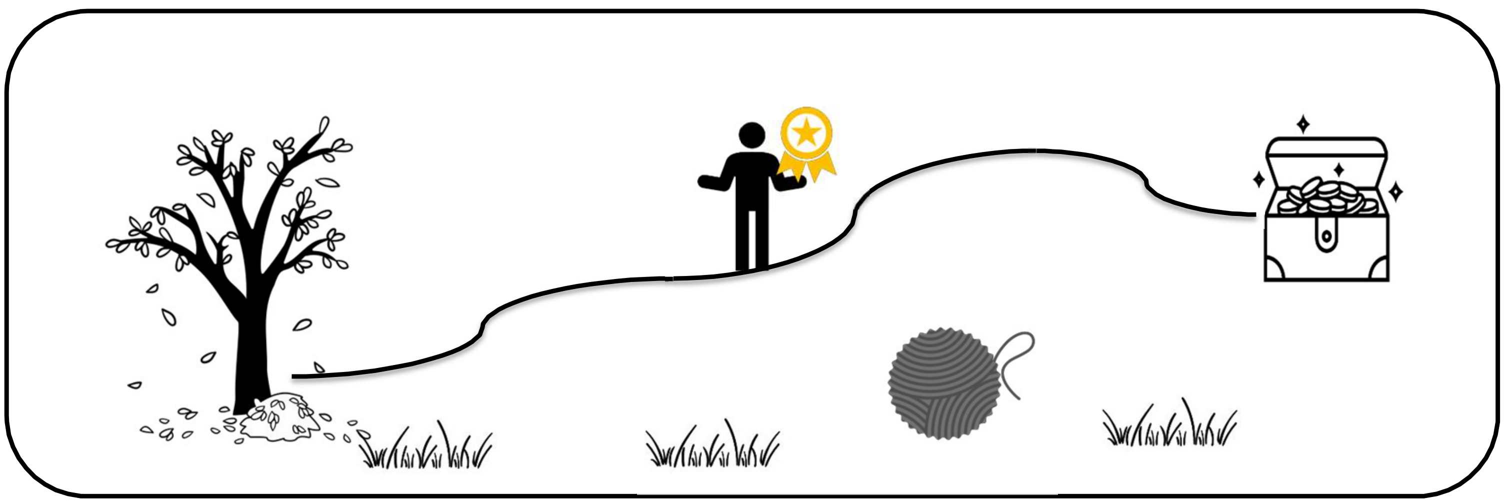
State : The current position or condition returned by the model



1	2	3	4	5
6	7	8	9	10
11	12	13	14	15

RL Elements : Reward

- Reward : positive or negative feedback the agent recieves
- Receives from the environment as a result of its actions
- To help the model move in the right direction
- Games : won/lost
- Navigation : getting close to treasure/obstacle

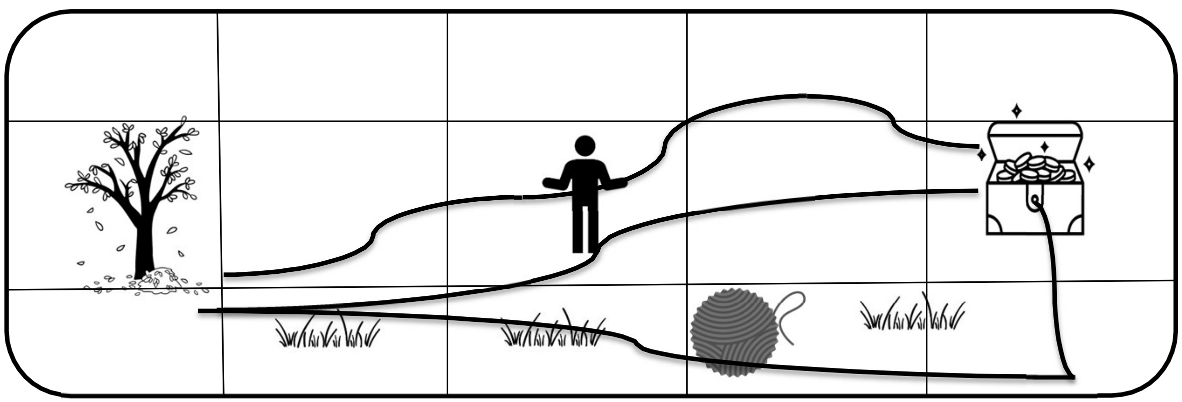


Reward Characteristics

- Positive and negative real values can be used as rewards.
- It is possible that the reward for a specific action will be delayed.
- Actions can have both short - term and long-term rewards
- Discounted rewards

RL Elements : Value

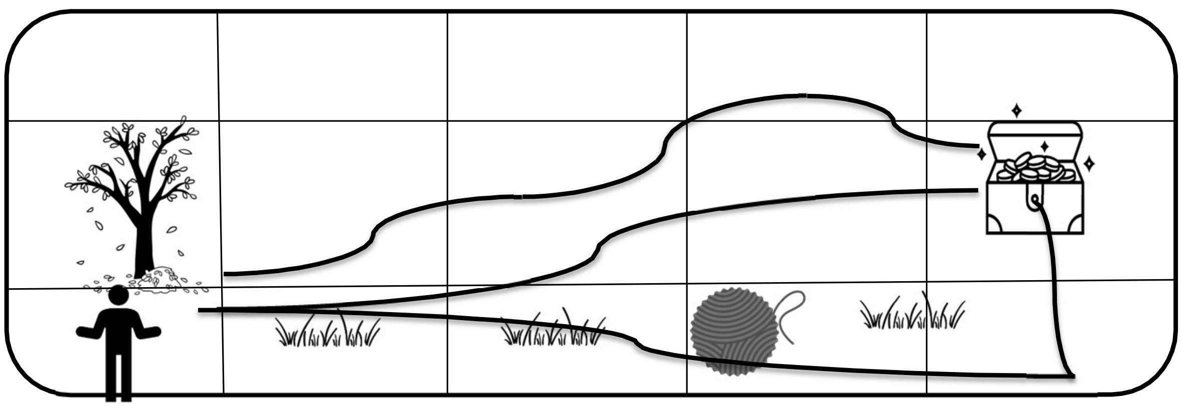
A function describes how good each state or action is.



11	25	30	80	90
10	17	80	90	100
11	60	10	-10	90

RL Elements : Policy

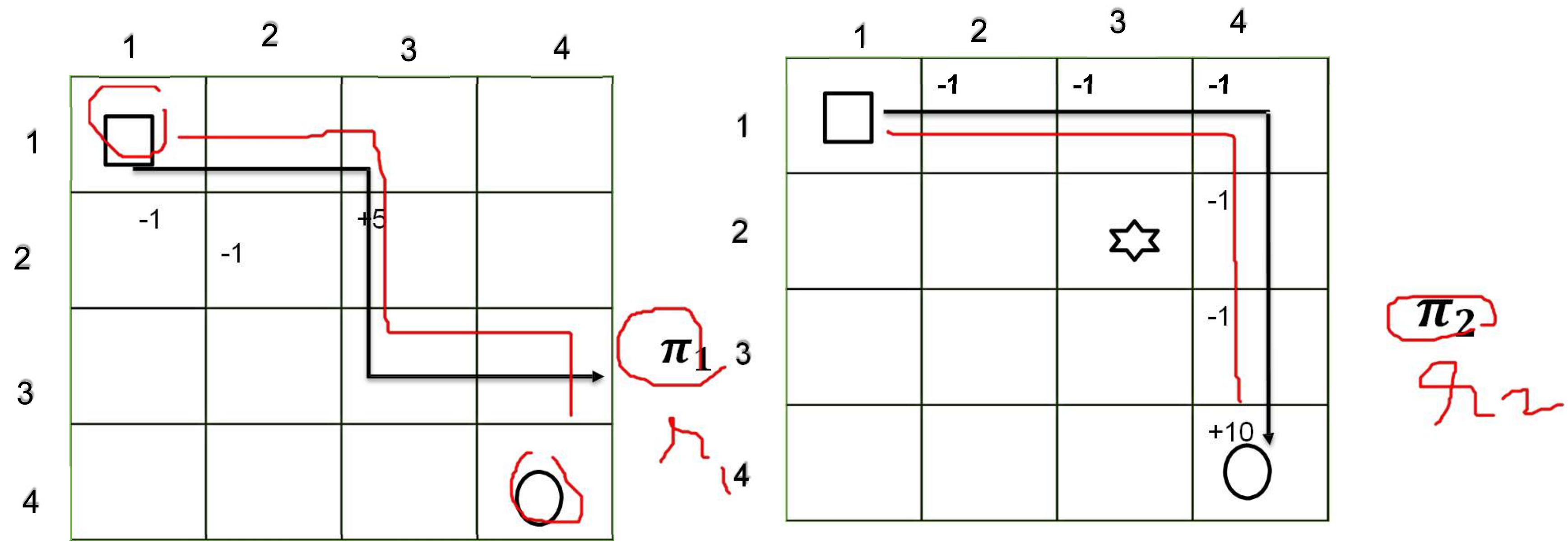
- Policy determines how an agent will behave at anytime
- Mapping between action and present state

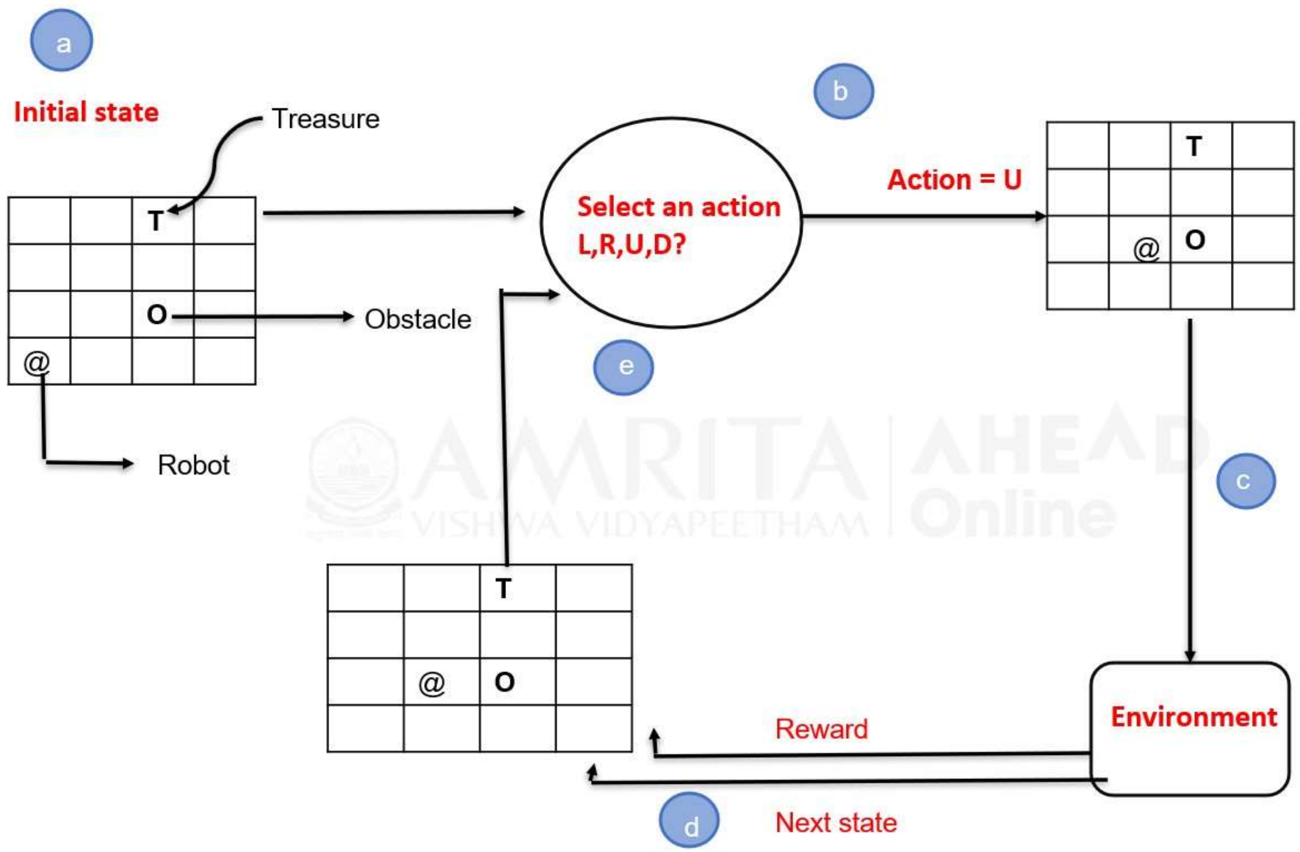


→	→	→	→	T
↑			O	

$\pi: s \rightarrow a$

Policy Example




a
Initial state
Treasure
T
O
Obstacle
@
Robot
b
Select an action
L,R,U,D?
Action = U
T
@
O
c
e
T
@
O
Environment
Reward
Next state
d

Take away

- Components of RL
- How RL works?

References

- [Chapter 1] Sutton and Barto, Reinforcement Learning: An Introduction, The MIT Press Cambridge, Massachusetts London, England, 2015
- Csaba Szepesvari, Algorithms for Reinforcement Learning, Morgan &Claypool, United States, 2010