Introduction to Machine Learning

Mark A. Austin

University of Maryland

austin@umd.edu
ENCE 688P, Fall Semester 2021

October 16, 2021

Overview

Part 04

Types of Machine Learning

Types of

Machine Learning Systems

Machine Learning Systems

Types of Machine Learning Systems:

- Supervised machine learning.
- Unsupervised machine learning.
- Semi-supervised machine learning.
- Reinforcement machine learning.

Supervised Machine Learning

Supervised Machine Learning

Learning algorithms are trained with labeled data and adjust the model parameters to minimize the discrepancy between the computed output and desired output.

Data(x,y):

- x is data, y is the label.

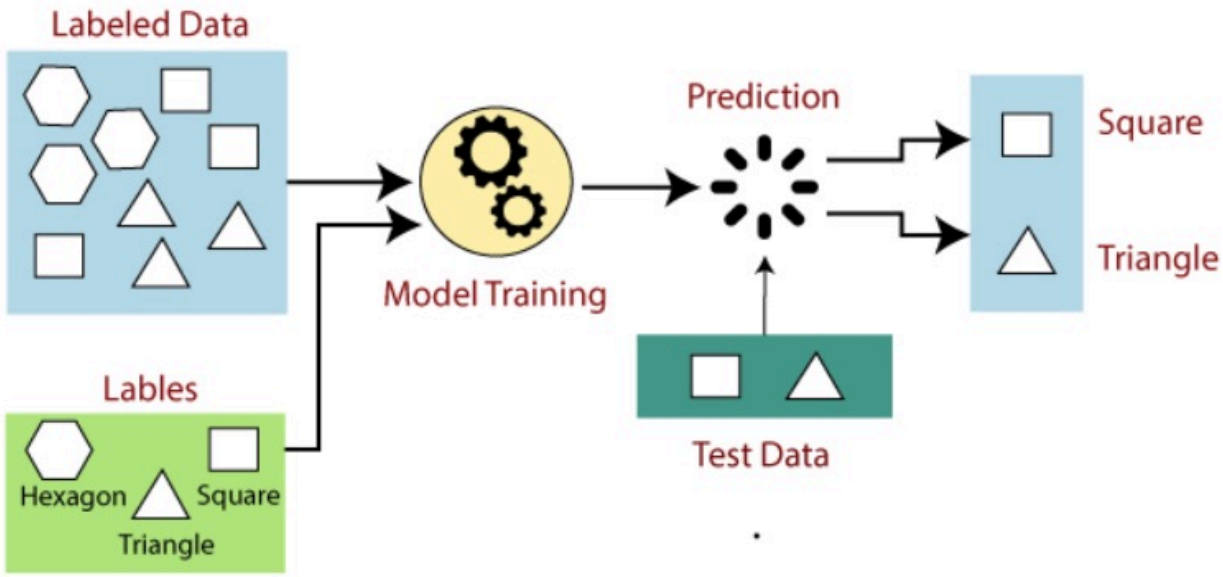
Goal:

- Learn function to map $x \rightarrow y$.

Common Algorithms: Regression, classification, naive bayes, object detection, neural networks, random forests, convolution neural networks.

Supervised Machine Learning

Supervised Machine Learning Process and Testing



Challenges:

- Data preparation and pre-processing; avoid unlikely and incomplete data.
- Identifying the right features to train the machine on.

Supervised Machine Learning

Advantages

- Can predict output based on previous experiences.
- Can have an exact idea about the classes of objects.
- Very useful in real-world applications such as fraud detection.

Disadvantages

- Not suitable for solution of complex tasks.
- Domain of expertise is very narrow – cannot predict correct output if test data is different from training dataset.
- Training requires prior knowledge of the classes of objects.
- Manual labeling of a large data set can be very time consuming.

Unsupervised Machine Learning

Unsupervised Machine Learning

Learning algorithms examine the structure of unlabeled data, and divide it into groups having the closest features.

Data(x):

- x is data, no labels.

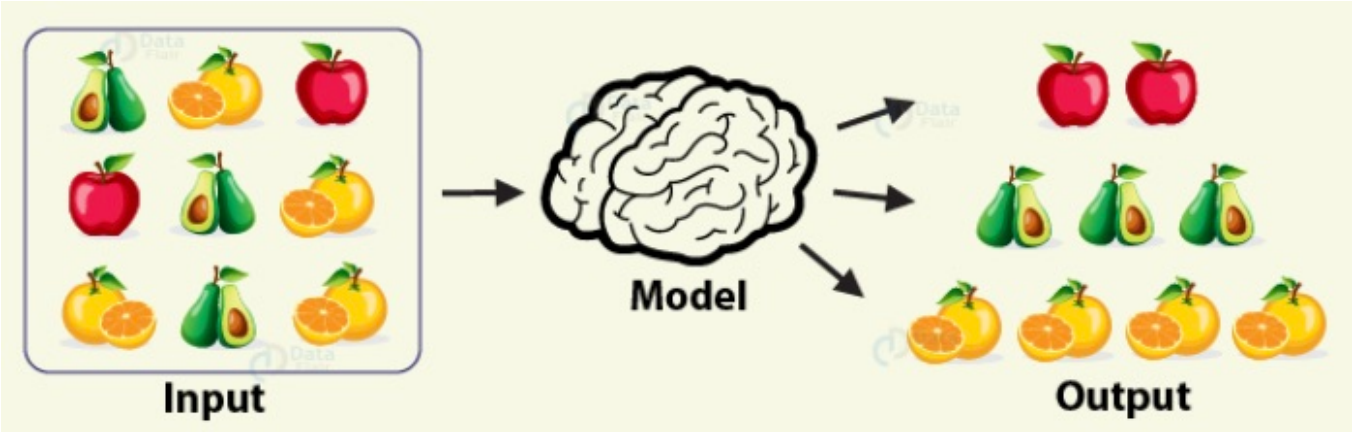
Goal:

- Learn hidden or underlying structure (or patterns in) of the data.

Common Algorithms: K-means clustering, feature or dimensionality reduction.

Unsupervised Machine Learning

Unsupervised Machine Learning Process



Abilities and Challenges

- No supervision needed.
- Unsupervised learning is closer to human cognitive function – it deduces patterns from a wide variety of application and learns over time.

Unsupervised Machine Learning

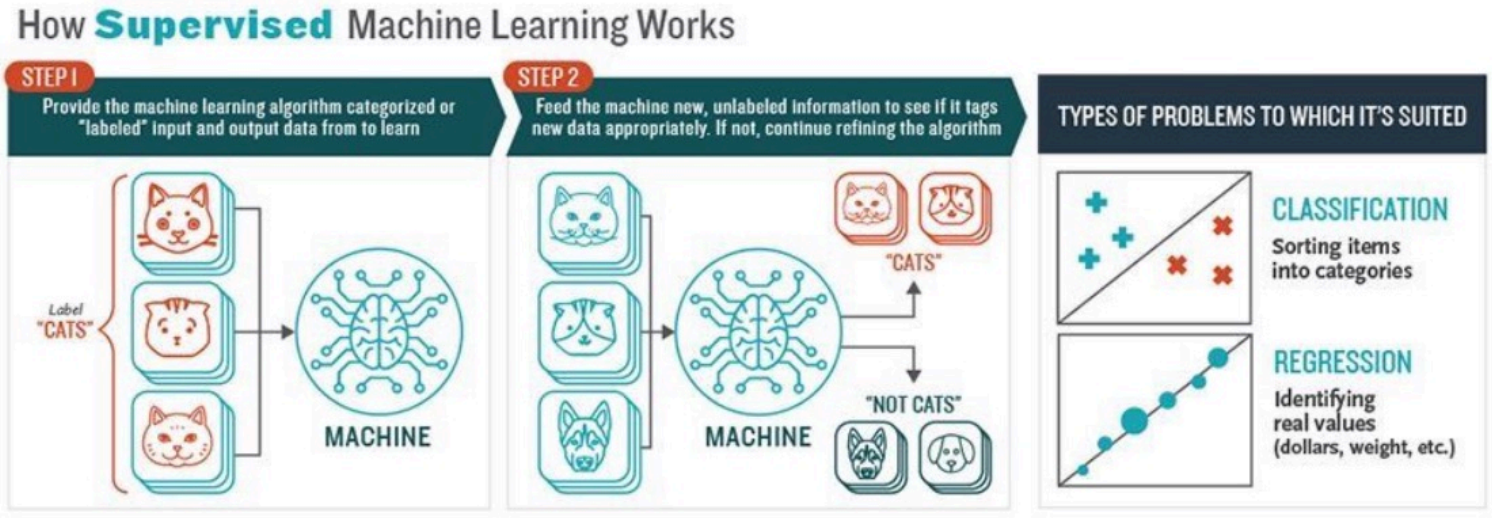
Advantages

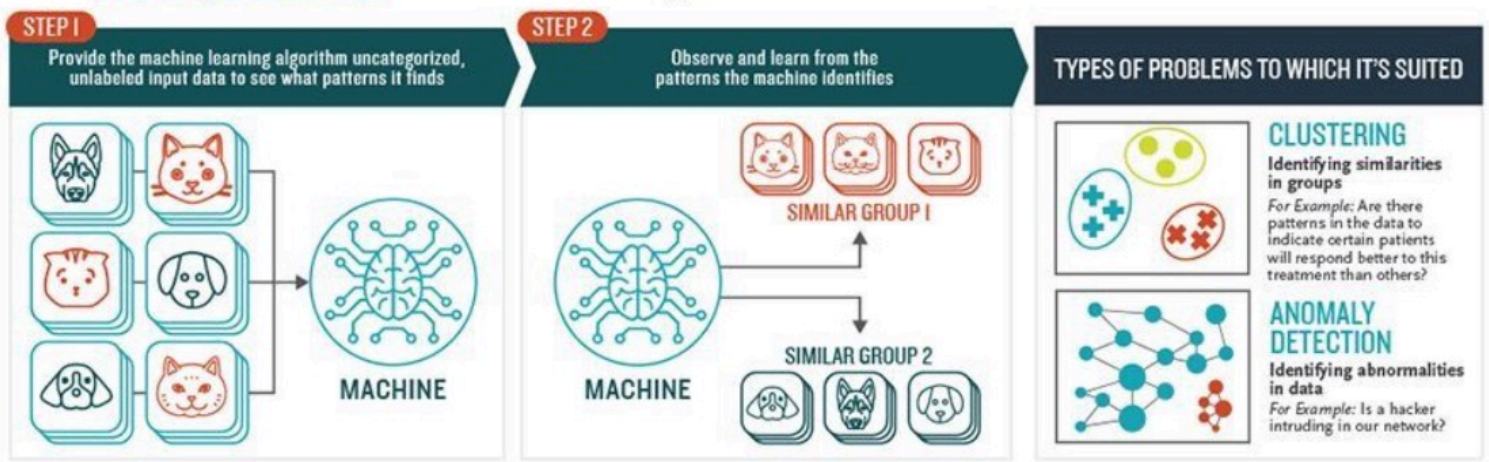
- Ability for a machine to tackle problems that humans might find insurmountable.
- Ideal for exploring raw and unknown data – training data does not need to be labelled.

Disadvantages

- Lower accuracy of results because the input data is now known and not labeled by people in advance.
- User needs to spend time interpreting and labeling classes/groups which follow classification.

Summary: Supervised Learning vs Unsupervised Learning





Semi-Supervised Learning

Semi-Supervised Learning

Semi-supervised learning is an approach to machine learning where algorithms use large amounts of unlabeled data to augment small amounts of labeled data to improve predictive accuracy.

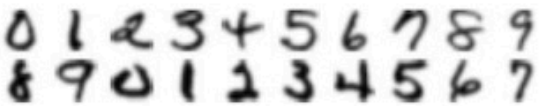
Semi-Supervised Learning in Humans

Concept learning in Children:

- Let x = animal, y = concept (e.g., cat).
- Parent points to animal and says: cat!
- Children subsequently observe animals by themselves and incrementally refine understanding.

Semi-Supervised Learning

Unlabeled and Labeled Data



Human expert/
Special equipment/
Experiment

"Crystal" "Needle" "Empty"

"0" "1" "2" ...

"Sports"
"News"
"Science"
...

Unlabeled data, X_i

Cheap and abundant !

Labeled data, Y_i

Expensive and scarce !

Semi-Supervised Learning

Algorithms

- Self-training, generative models, co-training.
- Graph-based algorithms.
- Semi-supervised support vector machines.

Applications

- Speech recognition and analysis.
- Spam detection and filtering.
- Video surveilance.
- 2D and 2D structure prediction.

Semi-Supervised Learning

Advantages

- Provides the benefits of both unsupervised and supervised learning while avoiding the challenge of finding large amounts of labeled data.

Disadvantages

- Cannot provide significant benefits over supervised learning unless one is absolutely sure that an assumption holds on the relationship between labels and the unlabeled data distribution.

Mathematically, we need:

$$p(x, y) = p(y)p(x|y), \tag{1}$$

where $p(x|y)$ is an identifiable mixture model.

Reinforcement Learning

Reinforcement Learning

Reinforcement learning algorithms use trial-and-error procedures to determine which action can provide the greatest reward.

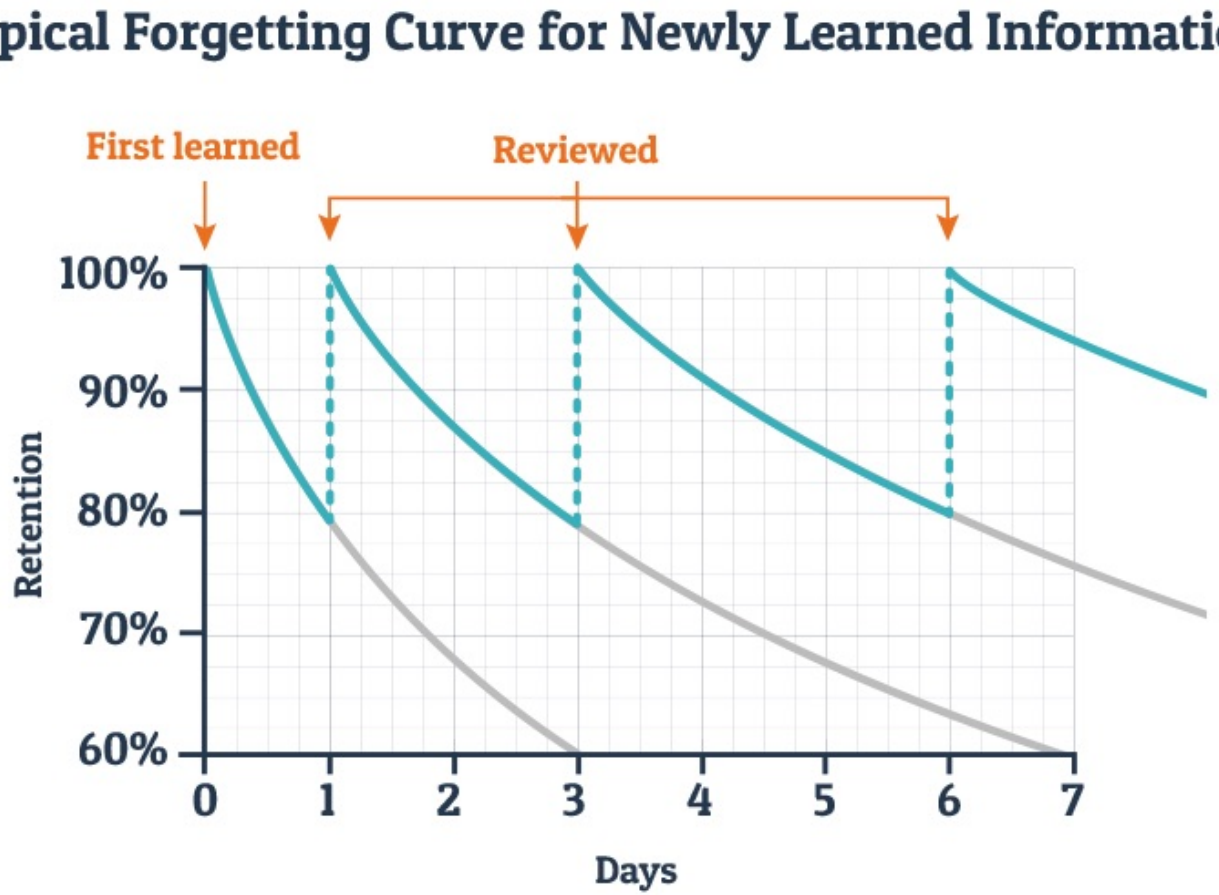
Data: state-action pairs.

Goal: Maximize future rewards over many time steps.

Examples: Taking actions to enhance survival/performance in gaming, robotics, optimization of operations for industrial machinery.

Reinforcement Learning

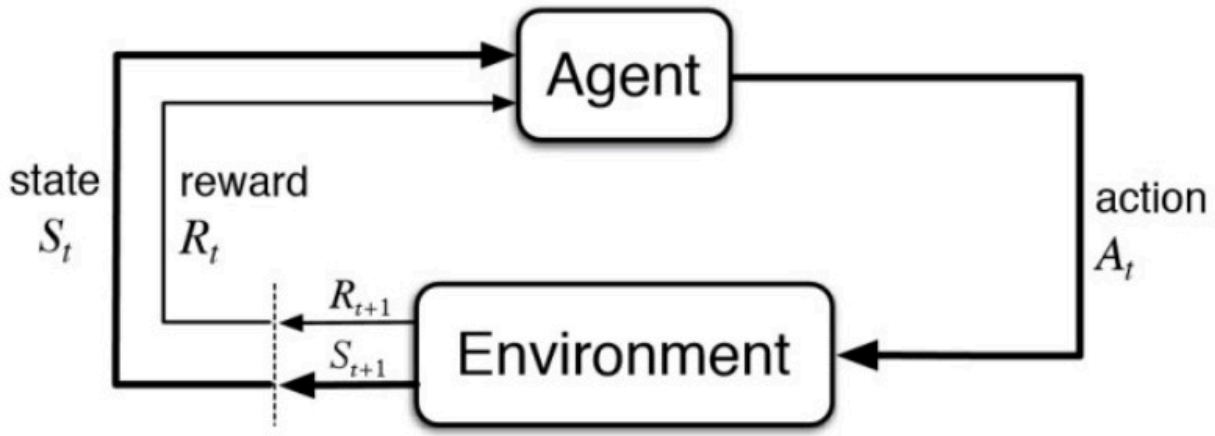
Using Reinforcement to Improve Memory Retention



Reinforcement Learning

Reinforcement Learning Process

In technical terms, reinforcement learning is a process in which a software agent makes observations and takes actions within an environment, and in return, it receives rewards.



The main objective is to maximize long-term rewards.

Reinforcement Learning

Definitions:

- **Environment:** Physical world in which the agent is operating.
- **State:** Current situation of the agent.
- **Reward:** Feedback from the environment.
- **Policy:** Method of map agent's state to actions.
- **Value:** Future reward that an agent would receive by taking an action in a particular state.

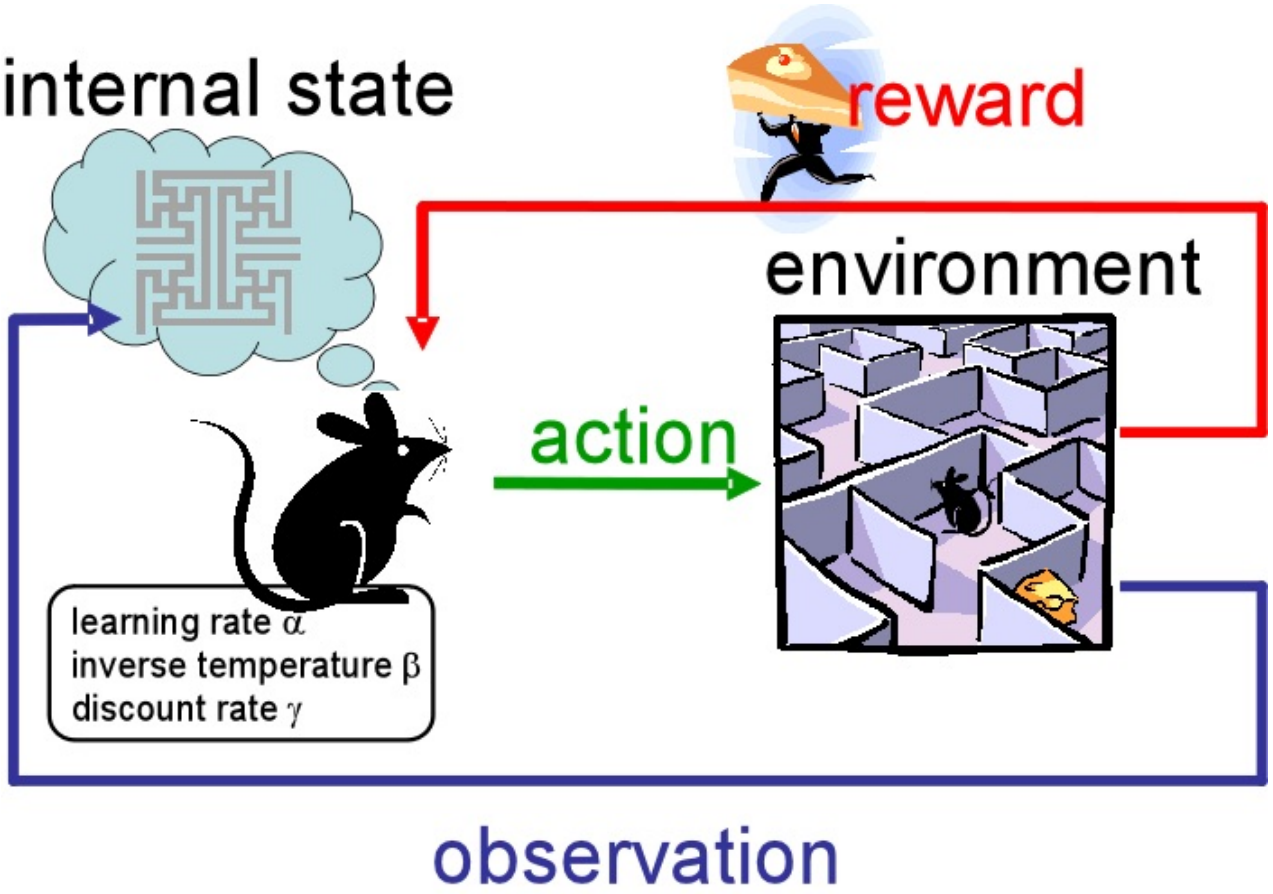
Note:

- These conditions may not always be present in real-world applications.

More Details: See Technical Tutorial on RL by Pieter Abbeel and John Schulman at UC Berkeley.

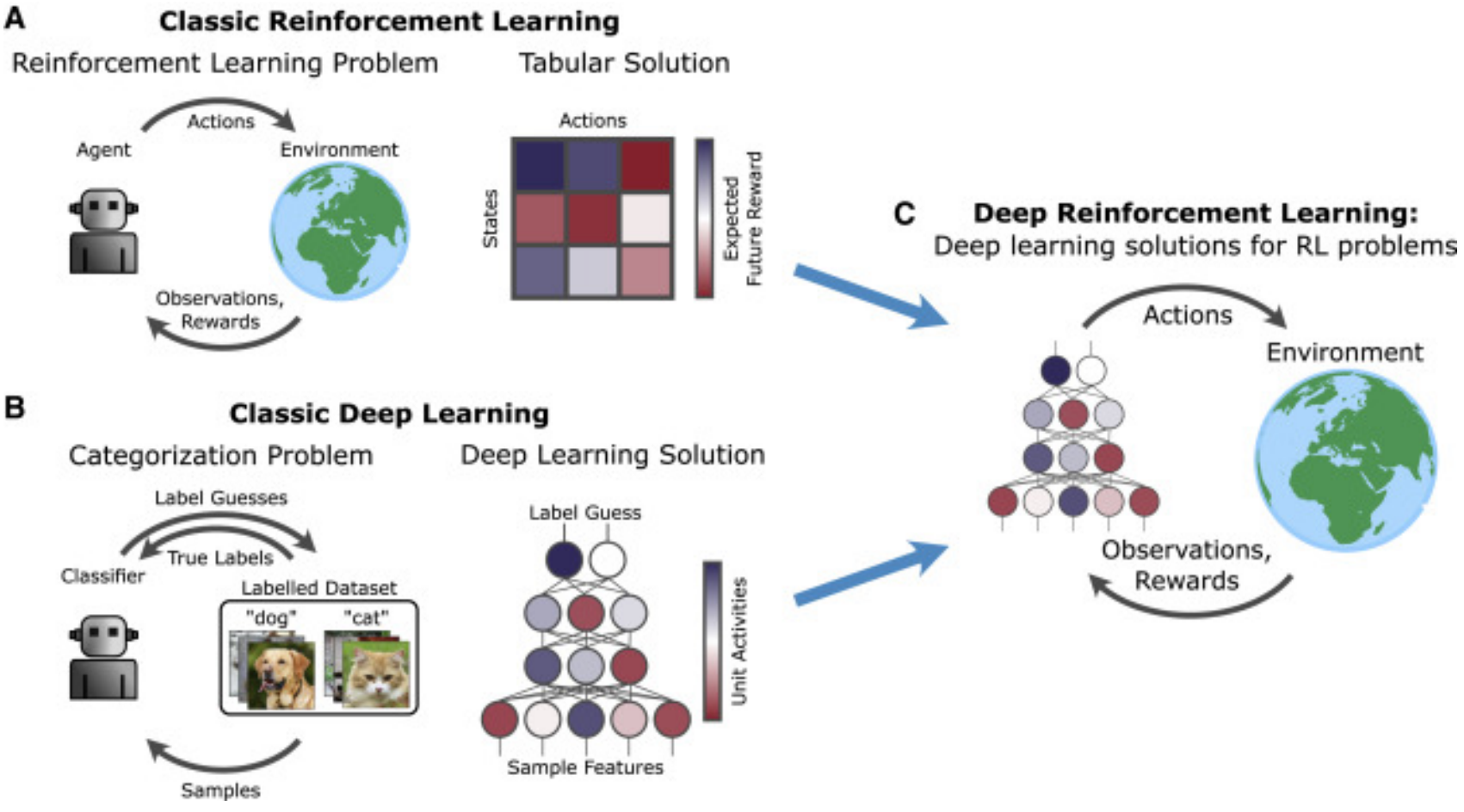
Reinforcement Learning

Simple Example: Mouse Searches Maze to find Cheese



Reinforcement Learning

Classic and Deep Reinforcement Learning



Reinforcement Learning

Advantages

- Reinforcement learning can be used to solve very complex problems that cannot be solved by conventional techniques.
- Errors can be corrected during the learning process.
- Learning process is very similar to humans, but it can often outperform humans ...

Disadvantages

- Not suitable for solving simple problems.
- Reinforcement learning requires lots of data and computation.
- Assumes incorrectly that the World follows a Markovian model, described in terms of sequences of possible events in which the probability of each event depends only on the current state.

References

- Austin M.A., Delgoshaei P., Coelho M. and Heidarinejad M. , Architecting Smart City Digital Twins: Combined Semantic Model and Machine Learning Approach, Journal of Management in Engineering, ASCE, Volume 36, Issue 4, July, 2020.
- Coelho M., and Austin M.A. , Teaching Machines to Understand Urban Networks, The Fifteenth International Conference on Systems (ICONS 2020), Lisbon, Portugal, February 23-27, 2020, pp. 37-42.
- Bhiksha R., Introduction to Neural Networks, Lisbon Machine Learning School, June, 2018.
- Lu T., Fundamental Limitations of Semi-Supervised Learning, MS Thesis in Mathematics in Computer Science, University of Waterloo, Canada, 2009.
- Van Engelen J.E., and Hoos H.H., A Survey on Semi-Supervised Learning, Machine Learning, Vol. 109, 2020, pp. 373-440.