Applied Machine Learning

MDS 552

Course Overview

MDS 552 – Applied Machine Learning

Nature of Course: Theory + Practical (Compulsory)

Course Description: This course covers the concept of machine learning and it application in real world task. It includes Supervised, Unsupervised and reinforcement learning algorithms and evaluation metrics to choose the best algorithm for a particular task.

Laboratory Works: Student are advised to implement supervised, unsupervised machine learning algorithm using any high level programing language (*Python and Scikit-Learn preferred*). The deep learning algorithms such as CNN and RNN should be implemented from scratch (*Using library are not preferred*).

Evaluation Criteria

- Internal Evaluation: 30 Marks
  - First Term Exam: 8 Marks
  - Second Term Exam: 8 Marks
  - Assignments : 5 Marks - 5 Assignments each carrying 1 marks (*Each evaluated in 100 points*)
  - Presentation: 5 Marks - Paper based presentation (*Evaluated for content, demo and presentation skills*)
  - Attendance: 2 Marks
  - Teachers Evaluation: 2 Marks
- Final Evaluation: 45 Marks

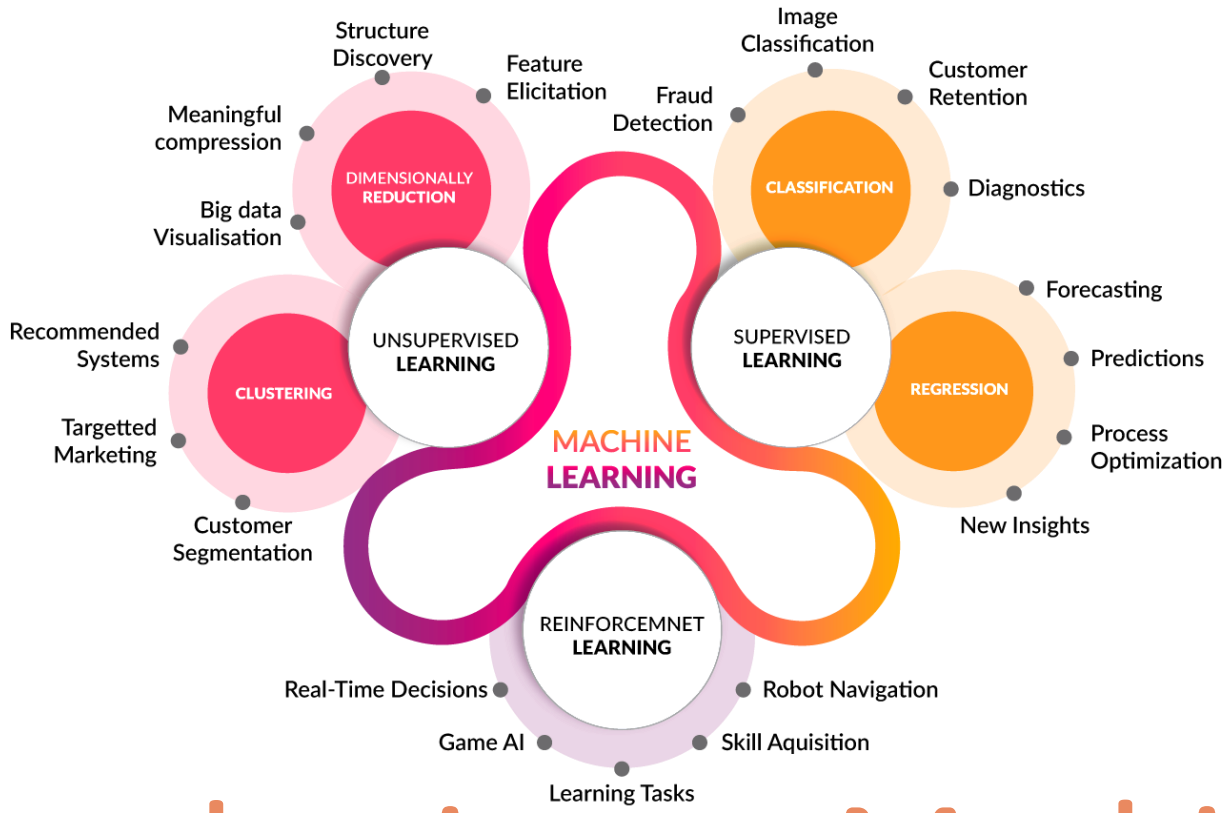
** Note: Reexam will have ceiling to pass mark only.*

Course Contents

- Introduction to Machine Learning
- Supervised Learning
- Unsupervised Learning
- Model Evaluation and Selections
- Reinforcement Learning
- Neural Network and Deep Learning

Course References

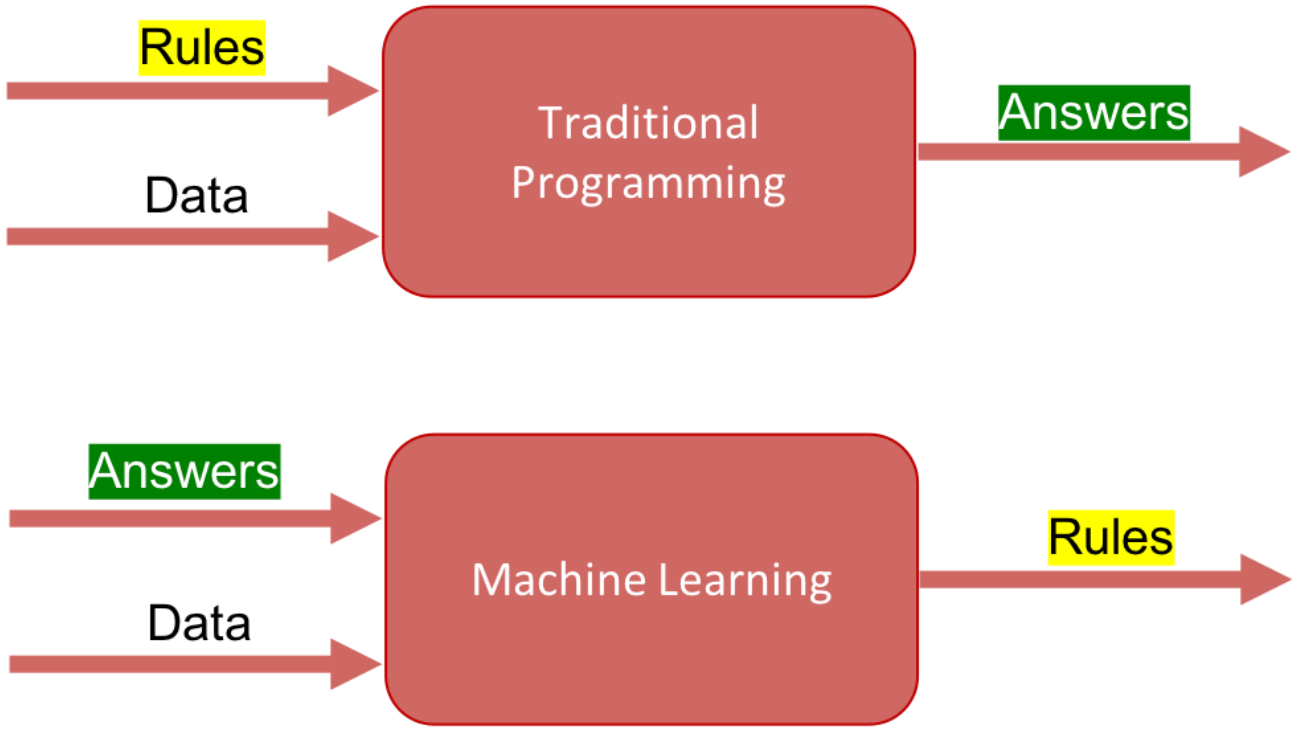
- Machine Learning Specialization by Andrew Ng - YouTube
- Machine Learning | Andrew Ng | Full Course | Stanford University – YouTube
- Lecture 1 | Machine Learning (Stanford) – YouTube
- Machine Learning Playlist - YouTube



Introduction to Machine Learning

Unit 1

What is Machine Learning?



Machine learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed.

What is Machine Learning?

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

- IBM

- In 1959, Arthur Samuel (an AI pioneer at IBM), defined the term "Machine Learning" as

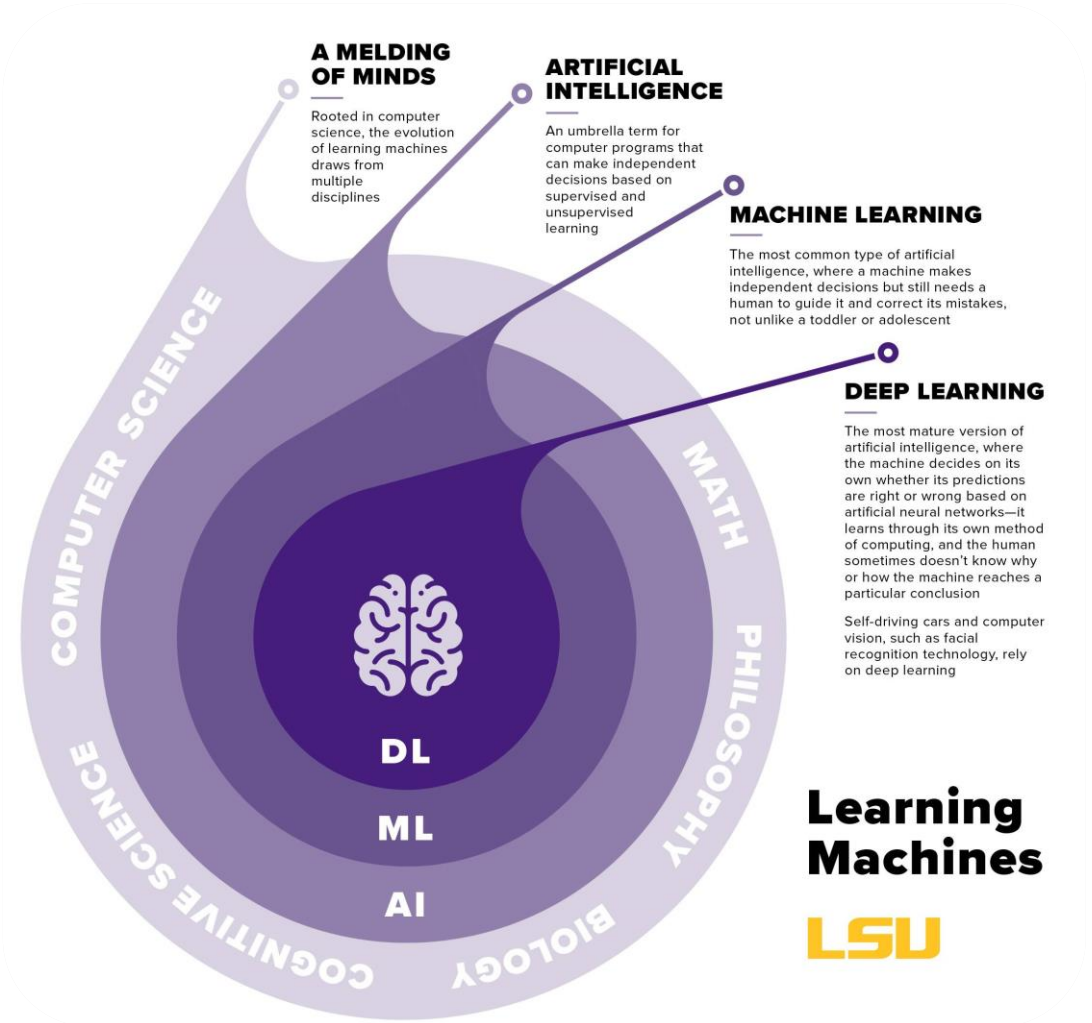
" Field of study that gives computers the ability to learn without being explicitly programmed. "

What is Machine Learning? (contd.)

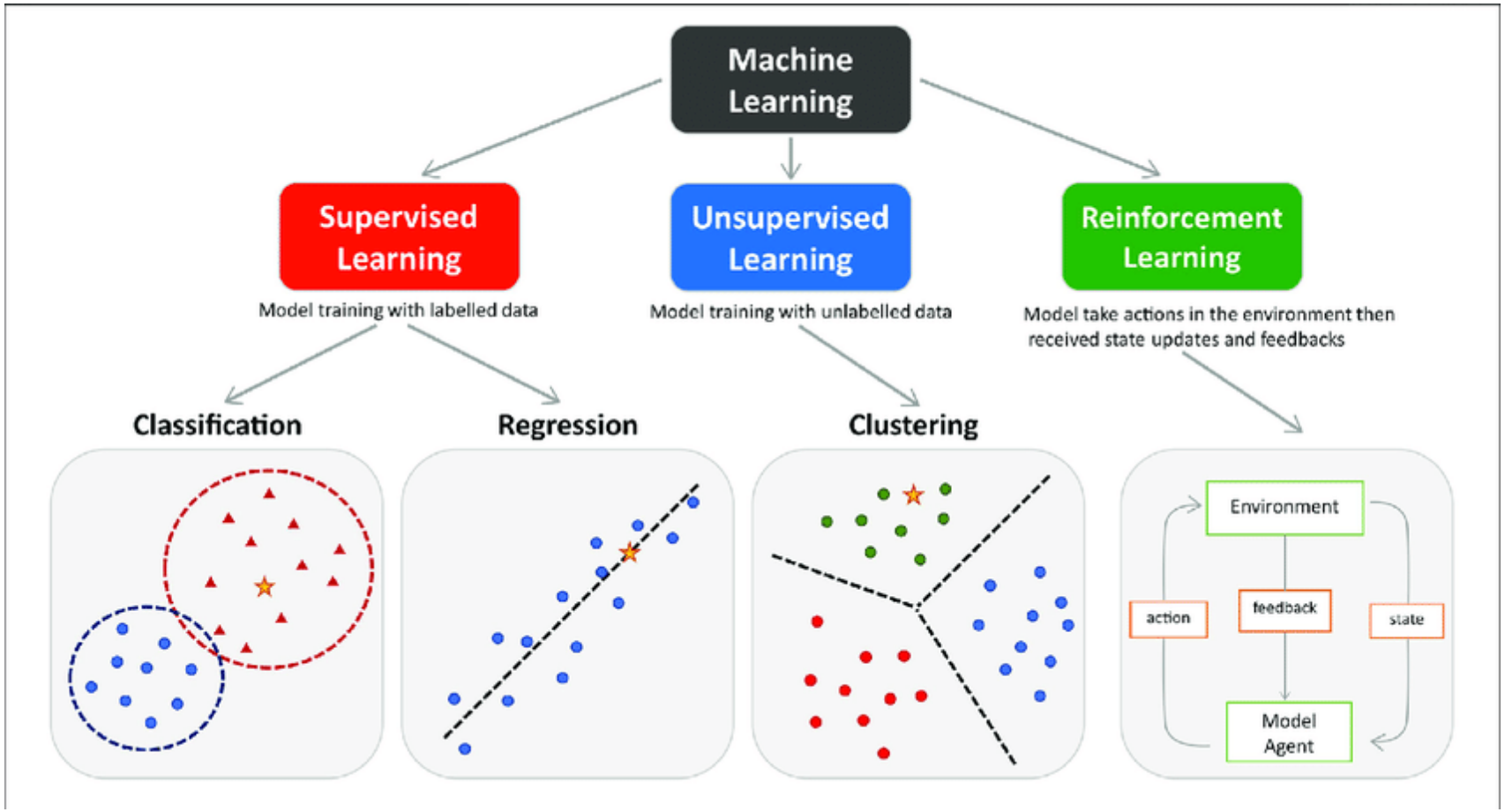
- After that, Tom M. Mitchell in 1997 gave the definition of Machine Learning which is widely quoted all over the globe.
- This is the more formal definition of the algorithms that are studied in the Machine Learning field,

> *"A computer program is said to learn from experience E with respect to some class of tasks T and a performance measure P if its performance in tasks T, as measured by P, improves with experience E."*

Machine Learning is the subfield of AI and is multidisciplinary



Different Types of Machine Learning

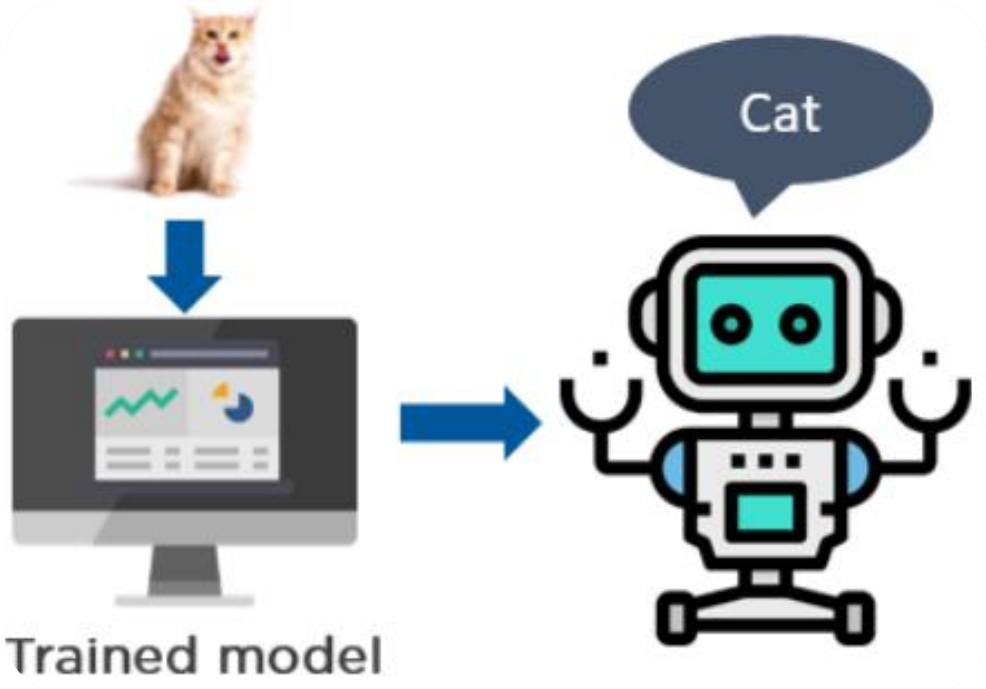


Also, there is something called **Semi Supervised Learning** and **Self Supervised Learning.**

Semi Supervised and Self Supervised Learning is not scope of this course.

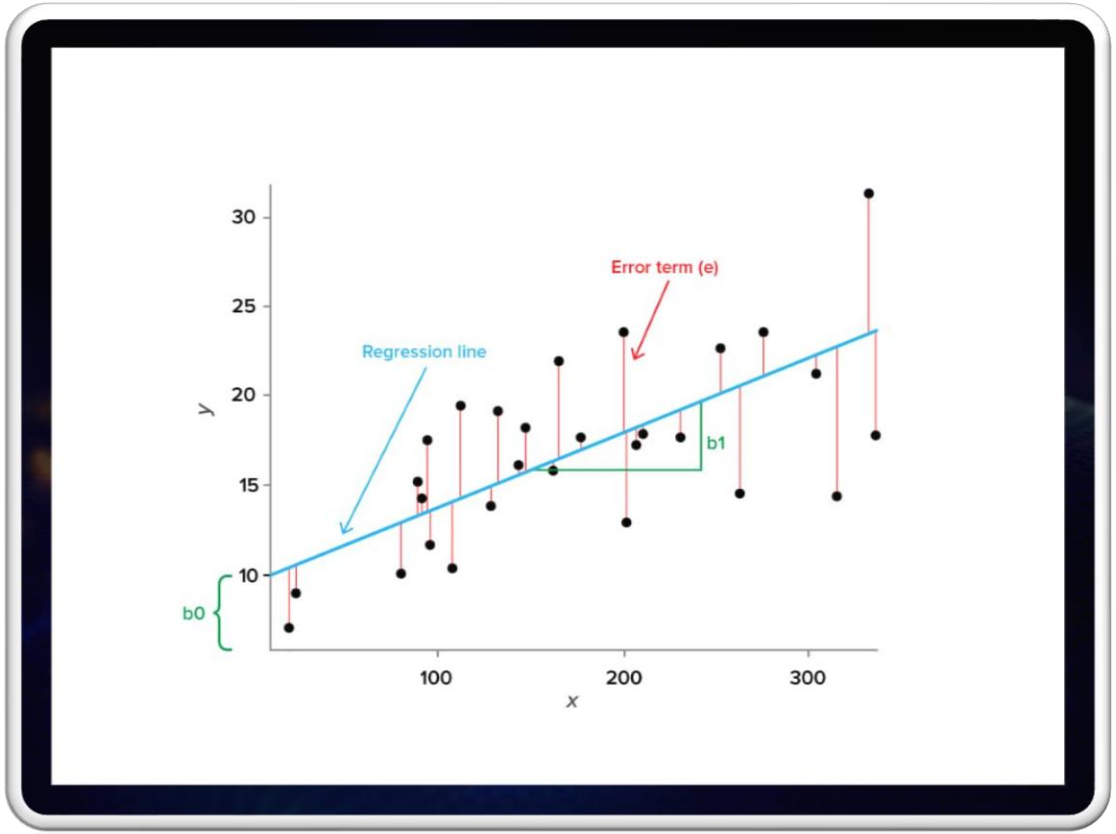
1. Supervised Learning

- A type of learning that uses labelled data (or input with outputs) to train machine learning algorithms.
- The input are provided with corresponding outputs while training ML models.
- Since these algorithm needs external supervision on mapping input and expected output, they are called **supervised**.



1. Supervised Learning (contd.)

- Even though the data needs to be labeled accurately for this method to work, usually done by experts, supervised learning is extremely powerful when used in the right circumstances.



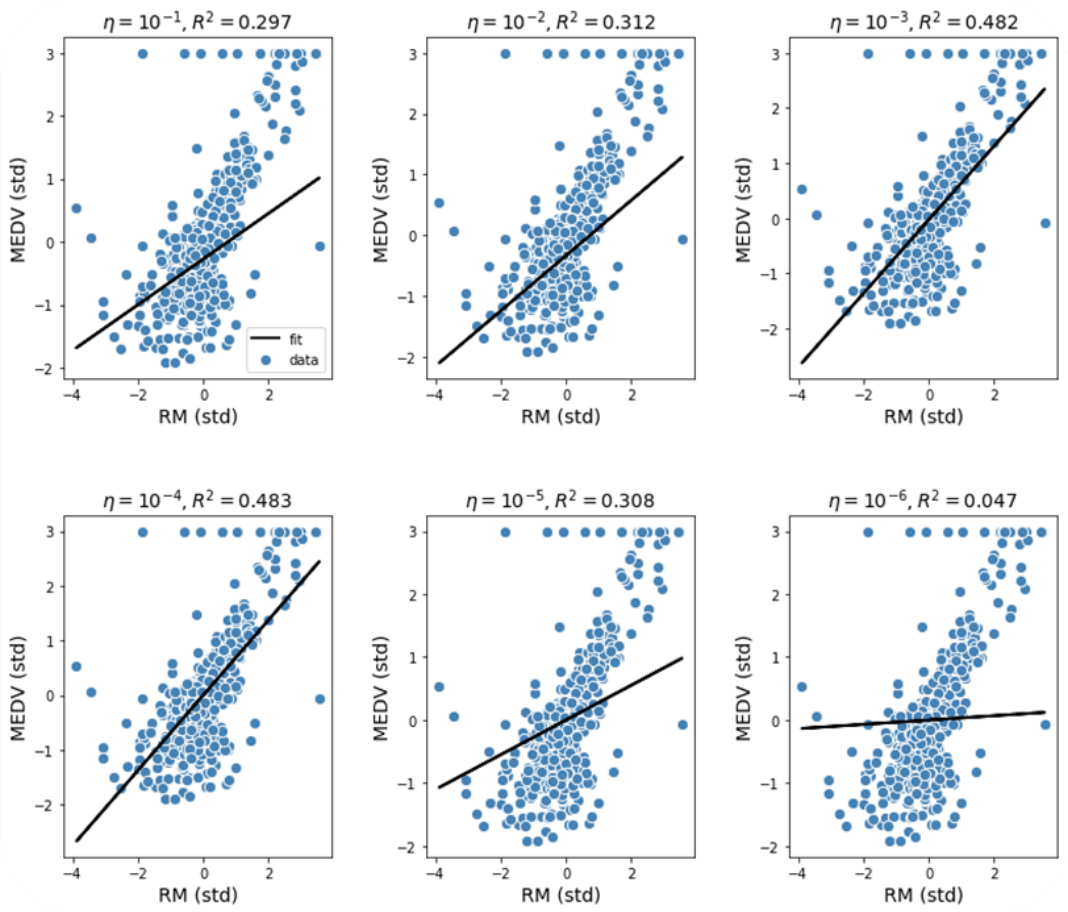
1. Supervised Learning (contd.)

Working Mechanism

- In supervised learning, the ML algorithm is given a training dataset to work with.
- This training dataset is a smaller part of the bigger dataset and serves to give the algorithm a basic idea of the problem, solution, and data points to be dealt with.
- The training dataset is also very similar to the final dataset in its characteristics and provides the algorithm with the labeled parameters required for the problem.

https://www.potentiaco.com/what-is-machine-learning-definition-types-applications-and-examples/

1. Supervised Learning (contd.)



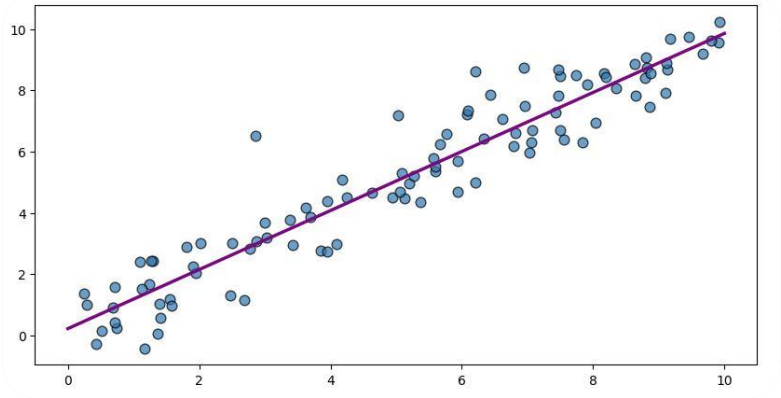
1. Supervised Learning (contd.)

- **Working Mechanism**
  - The algorithm then finds relationships between the parameters given, essentially establishing a cause and effect relationship between the variables in the dataset.
  - At the end of the training, the algorithm has an idea of how the data works and the relationship between the input and the output.
  - This solution is then deployed for use with the final dataset, which it learns from in the same way as the training dataset.
  - This means that supervised machine learning algorithms will continue to improve even after being deployed, discovering new patterns and relationships as it trains itself on new data.

Types of Supervised Learning

Regression

- Learning for prediction of value
- Has continuous output
- Needs corresponding output with input

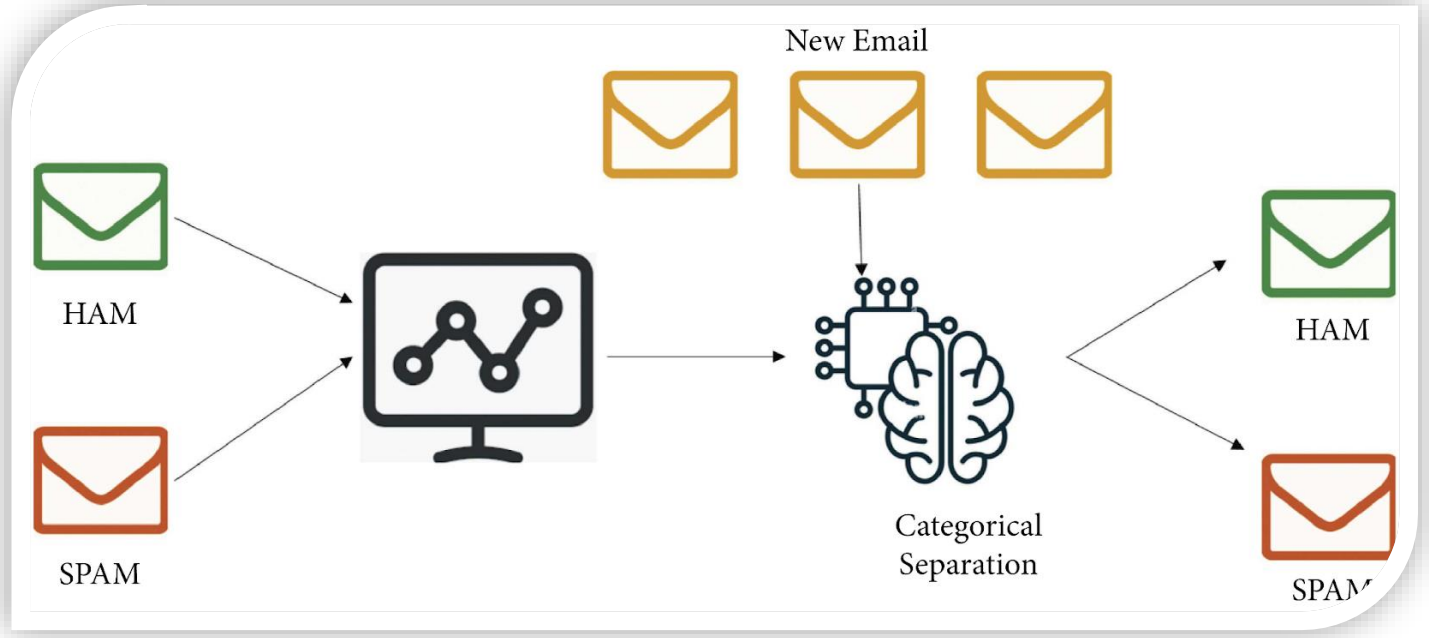


Classification

- Learning for classification of objects
- Has discrete output
- Needs Labelled data with input

1. Supervised Learning (contd.)

- **Applications**
  - Price Prediction
  - Image classification and segmentation
  - Disease identification and medical diagnosis
  - Fraud detection
  - Spam detection
  - Speech recognition
  - Sentiment Analysis
  - Image Captioning
  - Image Generation
  - Text Generation etc.



1. Supervised Learning (contd.)

- **Algorithms**
  - Linear regression
  - Logistic regression
  - Naive Bayes
  - Linear discriminant analysis
  - Decision trees
  - K-nearest neighbor algorithm
  - Neural networks (Multilayer perceptron)
  - Random forest algorithm
  - Support Vector Machine etc.

https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article

2. Unsupervised Algorithm

- A class of algorithm which has input data but no corresponding output or label.
- The goal for unsupervised learning is to model/ understand/ manipulate the underlying structure or distribution of the data in order to learn more about the data.
- There is no supervision from expert through label or output, thus, called unsupervised algorithm.
- Unsupervised machine learning purports to uncover previously unknown patterns in data, but most of the time these patterns are poor approximations of what supervised machine learning can achieve.

Types of Unsupervised Algorithms

- **Clustering** allows you to automatically split the dataset into groups according to similarity. Often, however, cluster analysis overestimates the similarity between groups and doesn't treat data points as individuals. For this reason, cluster analysis is a poor choice for applications like customer segmentation and targeting.
- **Anomaly** detection can automatically discover unusual data points in your dataset. This is useful in pinpointing fraudulent transactions, discovering faulty pieces of hardware, or identifying an outlier caused by a human error during data entry.
- **Association mining** identifies sets of items that frequently occur together in your dataset. Retailers often use it for basket analysis, because it allows analysts to discover goods often purchased at the same time and develop more effective marketing and merchandising strategies.
- **Latent variable models** are commonly used for data preprocessing, such as reducing the number of features in a dataset (dimensionality reduction) or decomposing the dataset into multiple components.

https://www.datarobot.com/wiki/unsupervised-machine-learning/

2. Unsupervised Algorithms (contd.)

- **Applications**
  - Market Basket Analysis
  - Semantic Clustering
  - Delivery Store Optimization
  - Identifying Accident Prone Areas
  - Customer Segmentation
  - Dimensionality Reduction
  - Image Segmentation
  - Audience segmentation
  - Market research
  - Recommendation System

2. Unsupervised Algorithms (contd.)

- **Algorithms**
  - K-means clustering
  - Hierarchical clustering
  - Gaussian Mixture Models
  - Apriori algorithms
  - FP Growth
  - Principal Component Analysis
  - Singular Value Decomposition
  - Autoencoders
  - Local Outlier Factor
  - Expected-Maximization

3. Reinforcement Learning

- We learning to make good decisions under uncertainty.
- Reinforcement learning problems involve learning what to do how to map situations to actions so as to maximize a numerical reward signal.
- In an essential way they are closed-loop problems because the learning systems actions influence its later inputs.
- Moreover, the learner is not told which actions to take, as in many forms of machine learning, but instead must discover which actions yield the most reward by trying them out.
- In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards.

Reinforcement Learning (contd.)

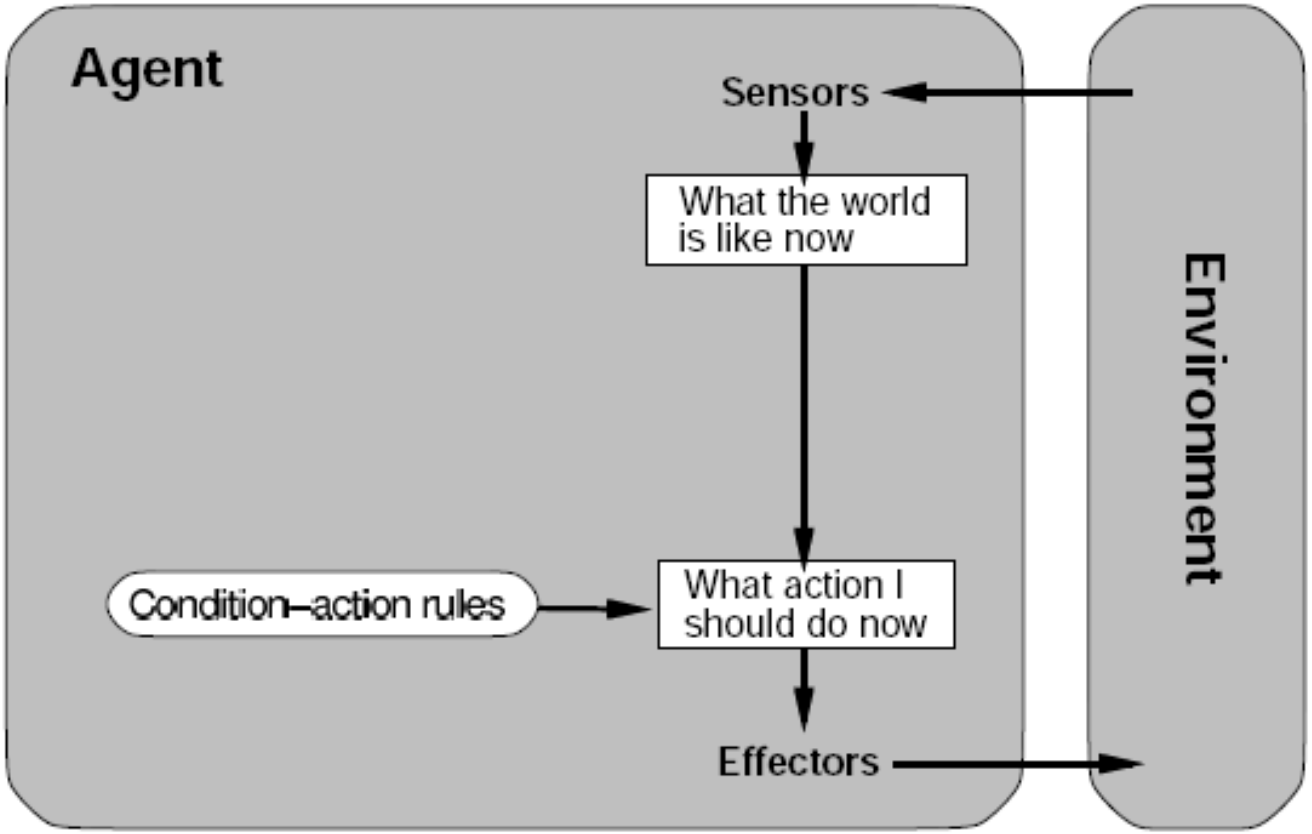
These three characteristics being

- closed-loop in an essential way,
- not having direct instructions as to what actions to take, and where the consequences of actions,
- including reward signals,

play out over extended time periods are the three most important distinguishing features of reinforcement learning problems.

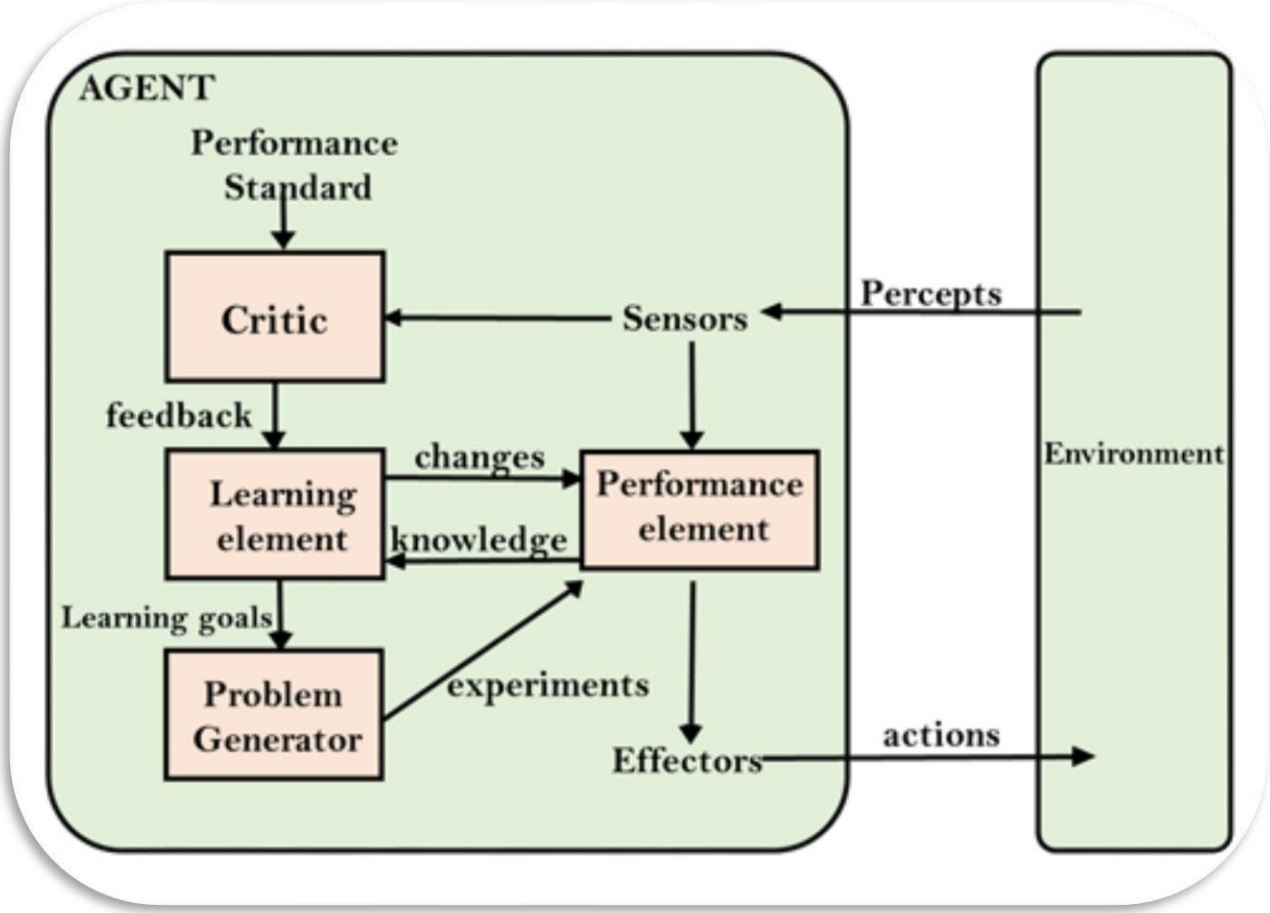
The basic idea is simply to capture the most important aspects of the real problem facing a learning agent interacting with its environment to achieve a goal. Clearly, such an agent must be able to sense the state of the environment to some extent and must be able to take actions that affect the state. The agent also must have a goal or goals relating to the state of the environment. The formulation is intended to include just these three aspects sensation, action, and goal in their simplest possible forms without trivializing any of them.

Reinforcement Learning (contd.)



A simple agent that interacts with environment.

Reinforcement Learning (contd.)



A learning agent that interacts with environment.

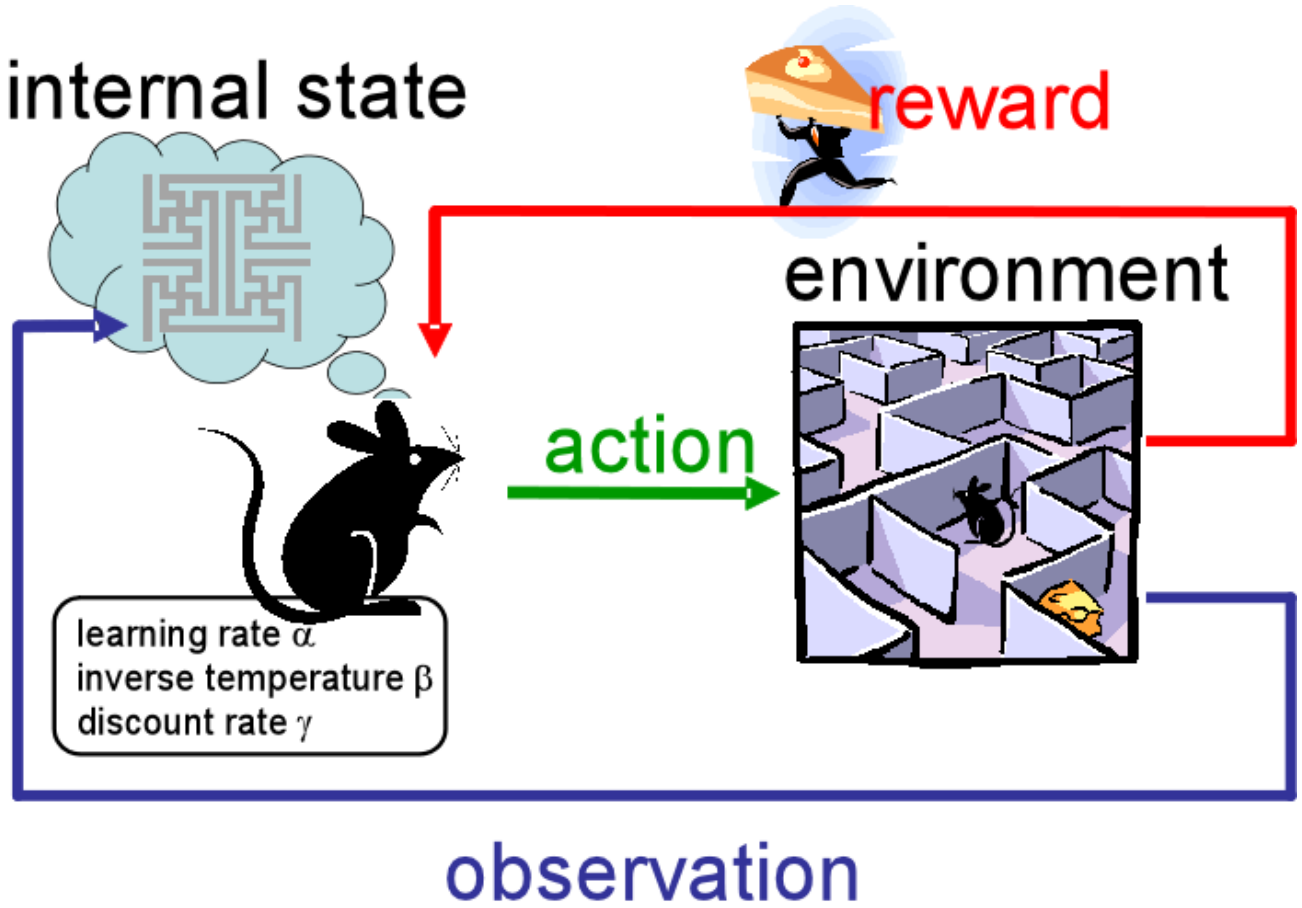
1.Learning element

2.Critic

3.Performance element

4.Problem Generator

Reinforcement Learning (contd.)



End of the Chapter

Next: Chapter 2 – Supervised Learning