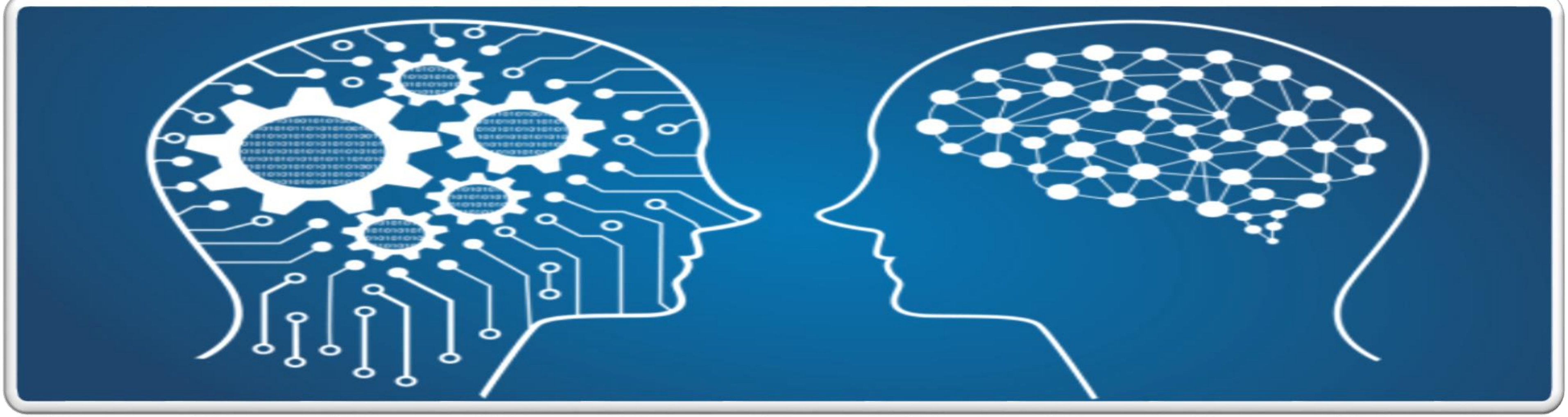


Machine Learning

Dr. Dinesh Kumar Vishwakarma

Professor,

Department of Information Technology,

Delhi Technological University, Delhi-110042

dinesh@dtu.ac.in

http://www.dtu.ac.in/Web/Departments/InformationTechnology/faculty/dkvishwakarma.php

Course Detail

- **Faculty:** Dinesh K Vishwakarma, *Ph.D. in Computer Vision*
  - ❖Email: dinesh@dtu.ac.in
  - ❖Webpage: http://www.dtu.ac.in/Web/Departments/InformationTechnology/faculty/dkvishwakarma.php
- **Course Code:**
  - ❖Credit: L T P: 3 0 2 : 4C

Evaluation Schedule

Class Test – I & Practical Test – I (On any working day **with prior information** to the Students)	Between 24.01.2022 & 04.02.2022
1st Review of Innovative Work* for **MTE** components for courses having **MTE** 1st Review of Innovative Practical Work* for **PRS** components for courses having **PRE**	Between 28.02.2022 & 04.03.2022
Class Test – II & Practical Test – II (On any working day **with prior information** to the Students)	Between 07.03.2022 & 17.03.2022
Submission & Evaluation of Innovative Work* for **MTE** component for courses having **MTE** Submission & Evaluation of Innovative Practical Work* for **PRS** component for courses having **PRE**	Between 30.03.2022 & 10.04.2022
Class Test – III & Practical Test – III (On any working day **with prior information** to the Students)	Between 11.04.2022 & 22.04.2022
Teaching Ends	22.04.2022
Online Submission of Marks of All Components of Evaluation	25.04.2022 (Action: Faculty)
Online Submission of Grades to Examination Branch by Course Coordinators	27.04.2022
Declaration of Results	29.04.2022

Evaluation Criteria

- CWS (15%) PRS (25%)
  - ❖ Assignments
  - ❖ Tutorials
  - ❖ Quiz's/Random Questions
- MTE (20%)
  - ❖ 1 Innovative Work in the form of Small Project, Startup Idea, Collaborative Projects, Automation, Simulation, Case study, Solutions to Real time social, economic and technical problems etc. (group of maximum 2 students): Graphical abstract
- ETE (40%)
  - ❖ (15x2=30%) **3 Class Tests** after every 4 weeks, Best 2 will be considered for evaluation.
  - ❖ (10x1=10%) **Minor Tests** in the form of Quizzes, Short Answer Questions, MCQs, Open Ended/Essay, Questions, etc. Better of the two will be considered for evaluation.

Course Content

Unit No.	Contents	Contact Hours
1	Introduction to Machine Learning: Overview of different tasks: classification, regression, clustering, control, Concept learning, information theory and decision trees, data representation, diversity of data, data table, form of learning, Basic Linear Algebra in machine learning techniques.	8
2	Supervised Learning: Decision trees, nearest neighbours, linear classifiers and kernels, neural networks, linear regression, logistic regression, Support Vector Machines.	12
3	Unsupervised Learning: Clustering, Expectation Maximization, K-Mean clustering, Dimensionality Reduction, Feature Selection, PCA, factor analysis, manifold learning.	10
4	Reinforcement Learning: Element of Reinforcement learning, Basic of Dynamic Programming; fining optimal policies, value iteration; policy iteration; TD learning; Q learning; actor-critic.	8
5	Recent applications & Research Topics: Applications in the fields of web and data mining, text recognition, speech recognition, finance.	4
Total Contact Hours		42

Books

Text Books	
1	Introduction to Machine Learning, Alpaydin, E., MIT Press, 2004
2	Machine Learning, Tom Mitchell, McGraw Hill, 1997
3	Elements of Machine Learning, Pat Langley Morgan Kaufmann Publishers
4.	**Applied Machine Learning, M. Gopal, McGraw Hill, 2018**
Reference	
1	The elements of statistical learning, Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. Vol. 1. Springer, Berlin: Springer series in statistics, 2001.
2	Machine Learning: A probabilistic approach, by David Barber.
3	Pattern recognition and machine learning by Christopher Bishop, Springer Verlag, 2006
4	An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics) 1st ed. 2013, Corr. 7th printing 2017 Edition

Resources: Journals

Ranking

1	IEEE Transactions on Pattern Analysis and Machine Intelligence
2	IEEE Transactions on Neural Networks and Learning Systems
3	Pattern Recognition
4	International Journal of Computer Vision
5	IEEE Transactions on Fuzzy Systems
6	Journal of Machine Learning Research
7	Expert Systems with Applications
8	Fuzzy Sets and Systems
9	Information Sciences
10	Artificial Intelligence
11	Machine Learning
12	Pattern Recognition Letters

Resources: Conferences

H5-index	Publisher	Conference Details
240	IEEE	CVPR : IEEE/CVF Conference on Computer Vision and Pattern Recognition Jun 16, 2020 - Jun 18, 2020 - Seattle , United States http://cvpr2020.thecvf.com/
169	Neural Information Processing Systems Foundation	NeurIPS : Neural Information Processing Systems (NIPS) Dec 6, 2020 - Dec 12, 2020 - Vancouver , Canada https://nips.cc/Conferences/2020/CallForPapers
137	Springer	ECCV : European Conference on Computer Vision Aug 23, 2020 - Aug 28, 2020 - Glasgow , United Kingdom https://eccv2020.eu/
135	PMLR	ICML : International Conference on Machine Learning (ICML) Jul 12, 2020 - Jul 18, 2020 - Vienna , Austria https://icml.cc/Conferences/2020
129	IEEE	ICCV : IEEE/CVF International Conference on Computer Vision Oct 11, 2021 - Oct 17, 2021 - Montreal , Canada http://iccv2021.thecvf.com/home Deadline : to be co
106	ACL	ACL : Meeting of the Association for Computational Linguistics (ACL) Aug 1, 2021 - Aug 6, 2021 - Bangkok , Thailand https://2021.aclweb.org/ Deadline : to be co
95	AAAI	AAAI : AAAI Conference on Artificial Intelligence Feb 2, 2021 - Feb 9, 2021 - Vancouver , Canada https://aaai.org/Conferences/AAAI-21/ Deadline : Tue 01
88	ACL	EMNLP : Conference on Empirical Methods in Natural Language Processing (EMNLP Nov 16, 2020 - Nov 20, 2020 - Online , Online https://2020.emnlp.org/
87	acm Association for Computing Machinery	CHI : Computer Human Interaction (CHI) May 8, 2021 - May 13, 2021 - Yokohama , Japan https://chi2021.acm.org/ Deadline : Thu 10
86	Association for	SIGKDD : ACM SIGKDD International Conference on Knowledge discovery and data mining

http://www.guide2research.com/topconf/machine-learning

A Few Quotes

- "A breakthrough in machine learning would be worth ten Microsofts" (Bill Gates, Chairman, Microsoft)
- Machine learning is the hot new thing" (John Hennessy, President, Stanford)
- "Web rankings today are mostly a matter of machine learning" (Prabhakar Raghavan, Dir. Research, Yahoo)
- "Machine learning is going to result in a real revolution" (Greg Papadopoulos, CTO, Sun)
- **Machine learning (ML)** is the study of computer algorithms that improve automatically through experience.

What is Machine Learning?

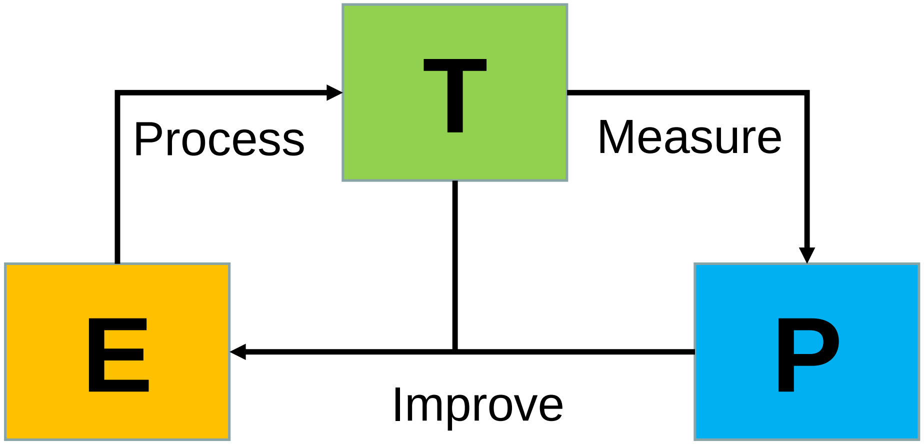
- A branch of artificial intelligence, concerned with the design and development of algorithms that allow computers to evolve ***behaviors based on empirical data***.
- Machine Learning is the science (and art) of programming computers so they can learn from data.
- As intelligence requires knowledge, it is necessary for the computers to acquire knowledge.
- Getting computers to program themselves
- Writing software is the bottleneck
- Let the data do the work instead!

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.
—Arthur Samuel, 1959

The term *machine learning* was coined in 1959 by Arthur Samuel

What is Machine Learning?

- 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, "Machine Learning" 1997.***



What is Machine Learning?

E	T	P
Experience	Task	Performance
Having Labelled Data: No. of students (male, female), etc.	Processing	Measuring Performance
Supervised Learning	Classification, Regression	Accuracy, Precession, Recall

What is Machine Learning?

T: Playing checkers
P: Percentage of games won against an arbitrary opponent
E: Playing practice games against itself

T: Recognizing hand-written words
P: Percentage of words correctly classified
E: Database of human-labeled images of handwritten words

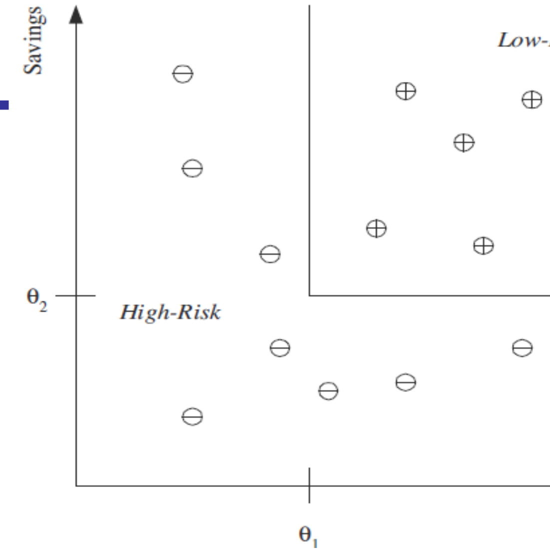
T: Driving on four-lane highways using vision sensors
P: Average distance traveled before a human-judged error
E: A sequence of images and steering commands recorded while observing a human driver.

T: Categorize email messages as spam or legitimate.
P: Percentage of email messages correctly classified.
E: Database of emails, some with human-given labels

Example 1: Class of ML Analysis

- Typical customer: **Admin/ Instructor**.
- Database:
  - ❖Current students registered
  - ❖basic parameters (Height, weight)
  - ❖Basic classification.
- Goal: predict/decide whether student is **FIT**?

Example 2: Credit Risk Analysis



- Typical customer: bank.
- Database:
  - ❖Current clients data, including:
  - ❖basic profile (income, house ownership, delinquent account, etc.)
  - ❖Basic classification.
- Goal: predict/decide whether to grant credit.

Example 2: Credit Risk Analysis

- Rules learned from data:

  IF Other-Delinquent-Accounts > 2 and

  Number-Delinquent-Billing-Cycles >1

  THEN DENY CREDIT

  IF Other-Delinquent-Accounts = 0 and

  Income > $30k

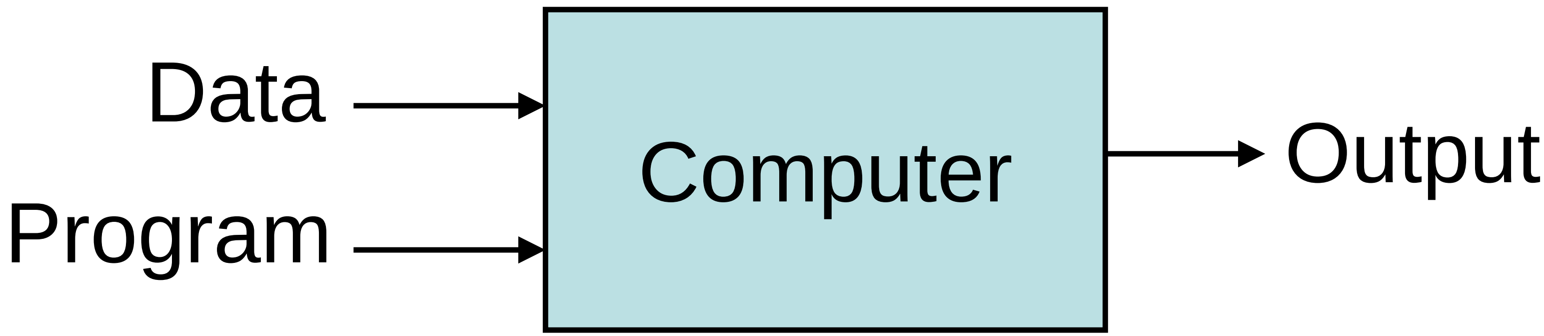
  THEN GRANT CREDIT

Example 3: Clustering news

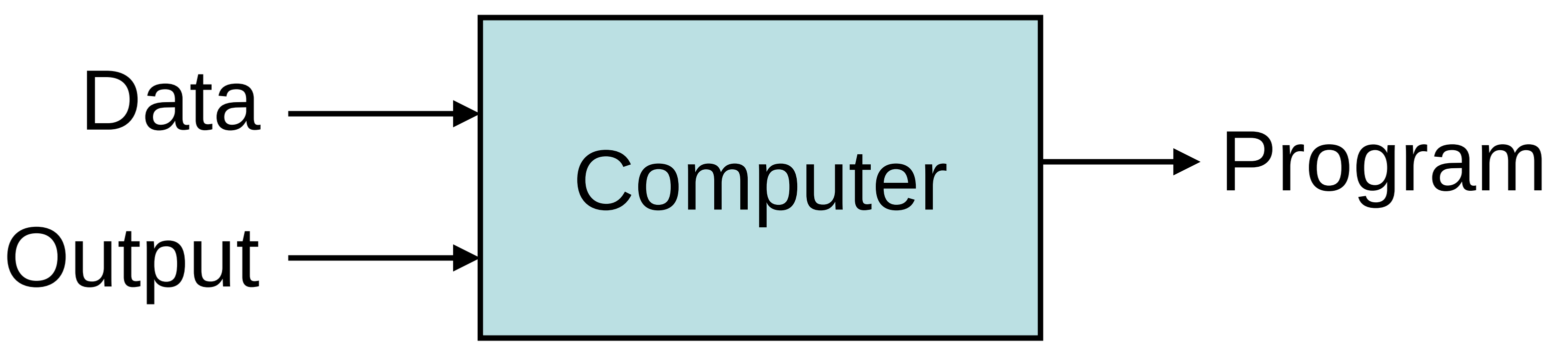
- Data: Reuters news / Web data
- Goal: Basic category classification:
  - ❖Business, sports, politics, etc.
  - ❖classify to subcategories (unspecified)
- Methodology:
  - ❖consider “typical words” for each category.
  - ❖Classify using a “distance “ measure.

What is Machine Learning?

Traditional Programming



Machine Learning

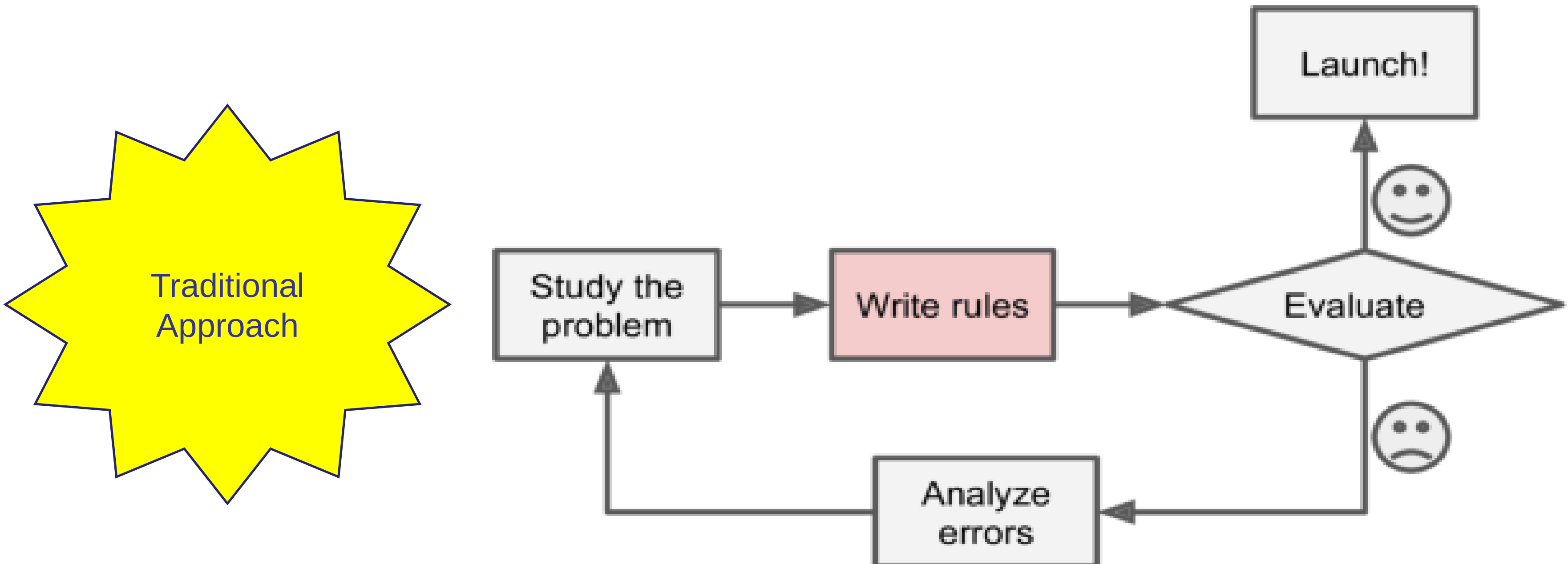


Resources: Datasets

- UCI Repository:
  http://www.ics.uci.edu/~mlearn/MLRepository.html
- UCI KDD Archive:
  http://kdd.ics.uci.edu/summary.data.application.html
- Statlib: http://lib.stat.cmu.edu/
- Delve: http://www.cs.utoronto.ca/~delve/
- Kaggle : https://www.kaggle.com/notebook

Why Machine Learning?

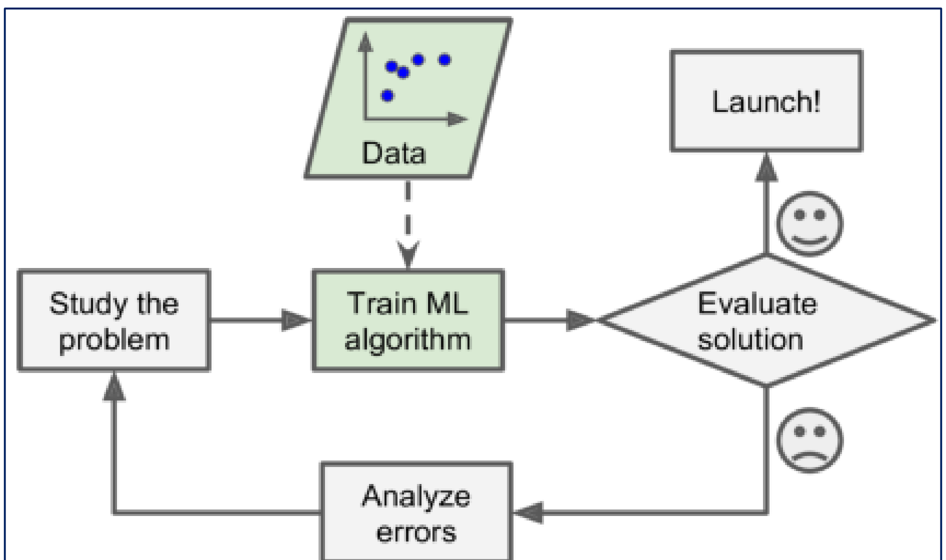
- Consider an example of Spam filtering.
  - ❖ First we look, how spam typically looks like, such as (“4U,” “credit card,” “free,” and “amazing”)
  - ❖ Then we write a detection algorithm for each patterns and flagged if pattern is detected.
  - ❖ We test our program and repeat step 1 and 2 until is good enough

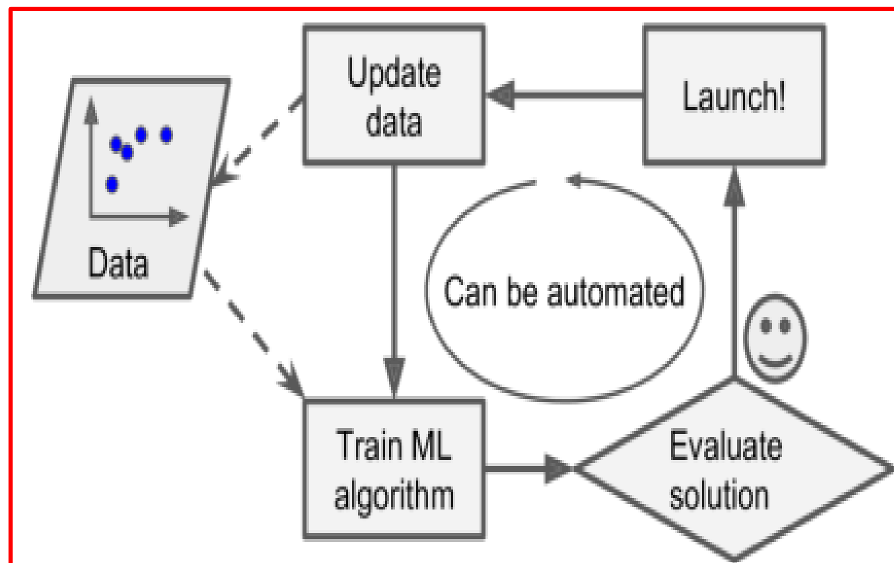


Since the problem is not trivial, your program will likely become a long list of complex rules—pretty hard to maintain

Why Machine Learning?...

- ML techniques automatically learns which words and phrases are good predictors of spam by detecting unusually frequent patterns of words in the spam examples compared to the ham example.
- The program is much shorter, easier to maintain, and most likely more accurate.

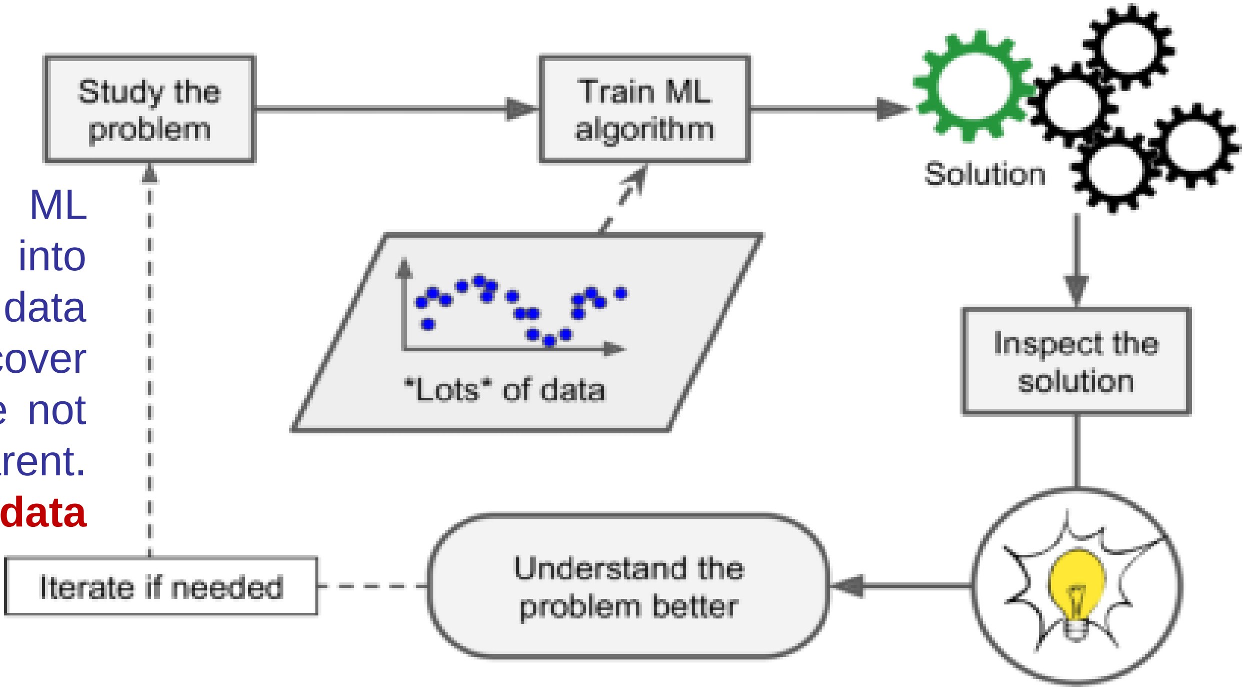




Why Machine Learning?...

- ML algorithms can be inspected to see what has been learned. For instance, once the spam filter has been trained on enough spam, it can easily be inspected to reveal the list of words and combinations of words that it believes are the best predictors of spam.
- Sometimes this will reveal unsuspected correlations or new trends, and thereby lead to a better understanding of the problem.

Applying ML techniques to dig into large amounts of data can help discover patterns that were not immediately apparent. This is called **data mining.**



Why Machine Learning?...

- No human experts
  - ❖ industrial/manufacturing control
  - ❖ mass spectrometer analysis, drug design, astronomic discovery
- Black-box human expertise
  - ❖ face/handwriting/speech recognition
  - ❖ driving a car, flying a plane
- Rapidly changing phenomena
  - ❖ credit scoring, financial modeling
  - ❖ diagnosis, fraud detection
- Need for customization/personalization
  - ❖ personalized news reader
  - ❖ movie/book recommendation

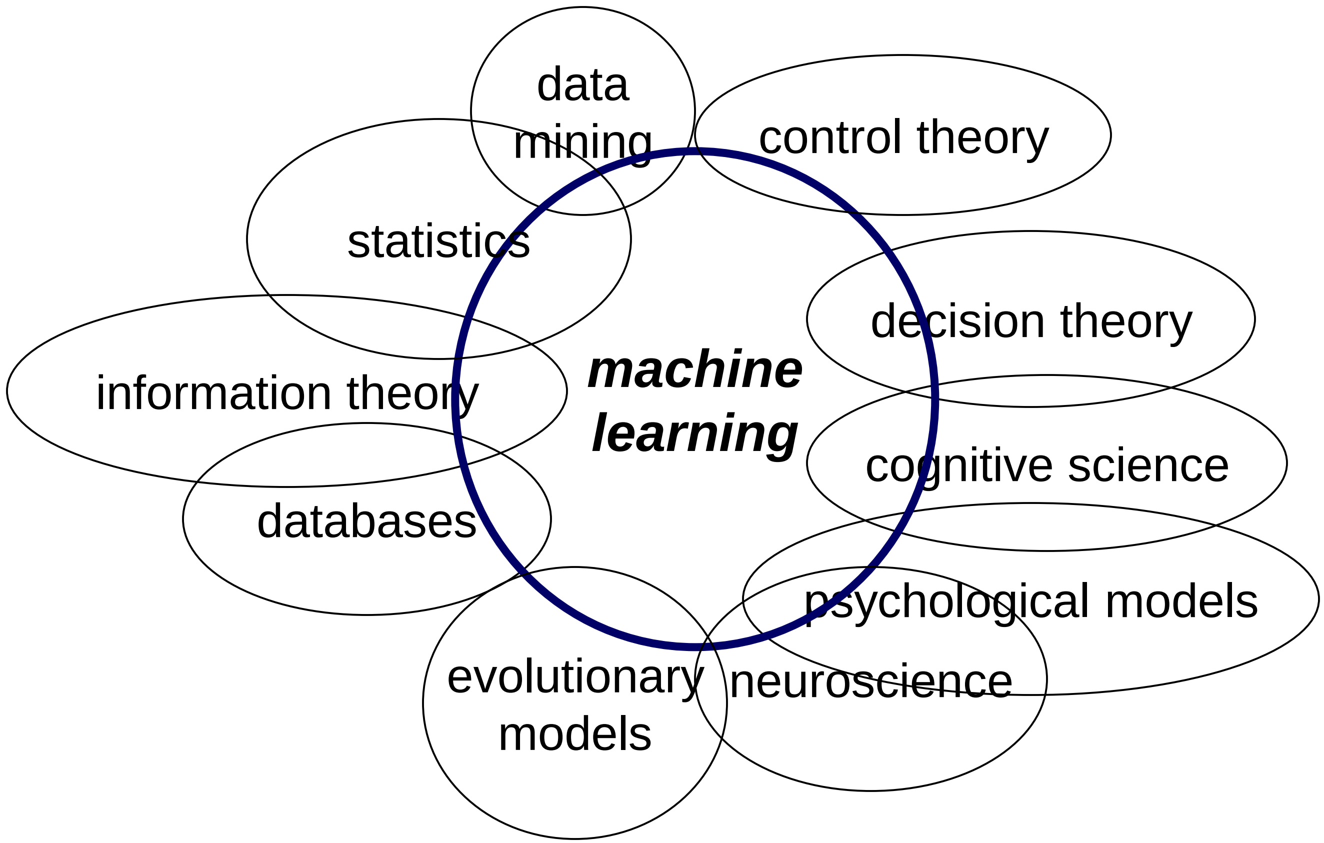
Benefit of ML over Rule Based

- Problems for which existing solutions require a lot of hand-tuning or long lists of rules: one ML algorithm can often simplify code and perform better.
- Complex problems for which there is no good solution at all using a traditional approach: the best ML techniques can find a solution.
- Fluctuating environments: a ML system can adapt to new data.
- Getting insights about complex problems and large amounts of data.

Applications

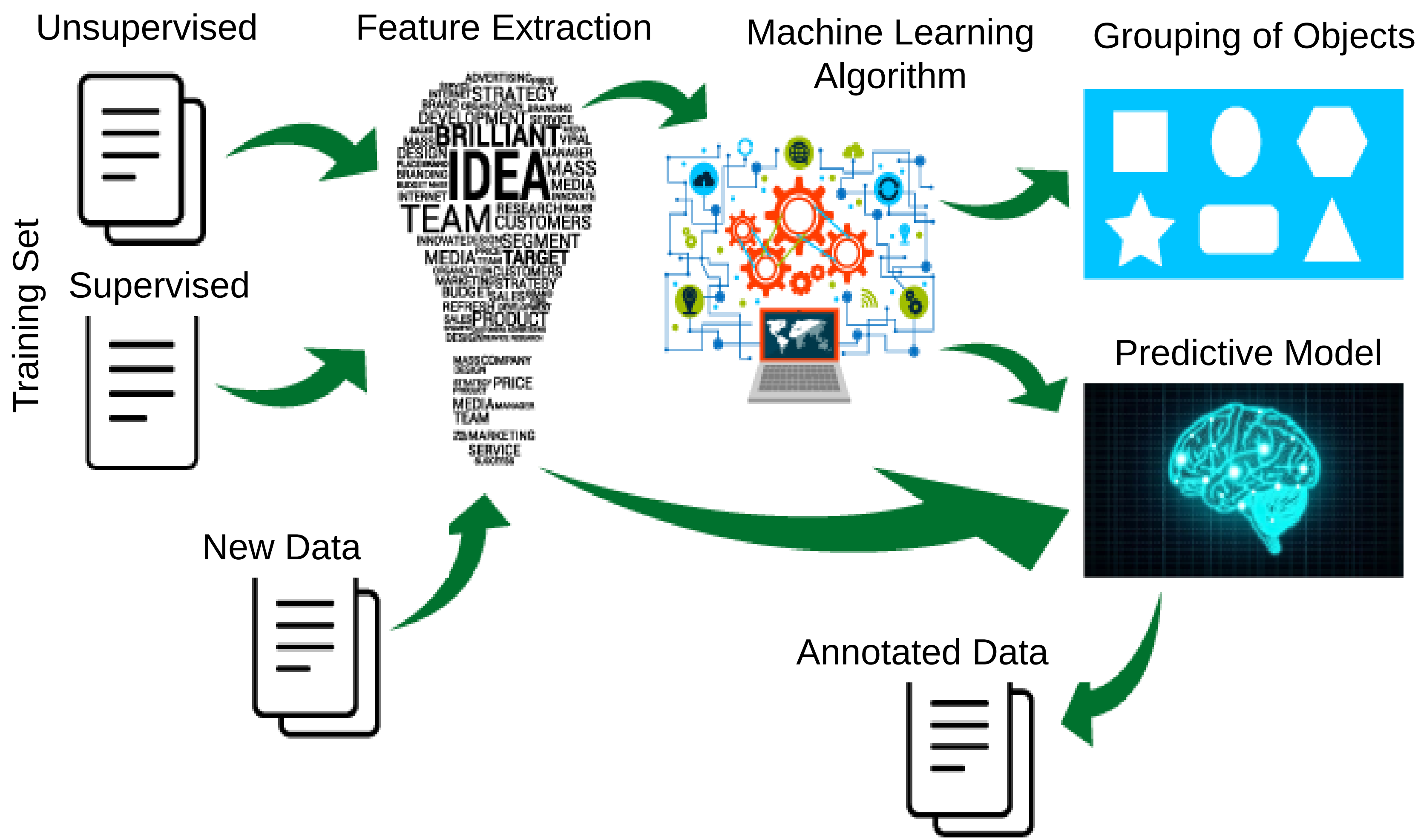
- Traffic Alerts
- Image Recognition
- Video Surveillance
- Sentiment Analysis
- Product Recommendation
- Online support using Chatbots
- Google Translate
- Online Video Streaming Applications
- Virtual Professional Assistants
- Machine Learning Usage in Social Media
- Stock Market Signals Using Machine Learning
- Auto-Driven Cars
- Fraud Detection

Related Field

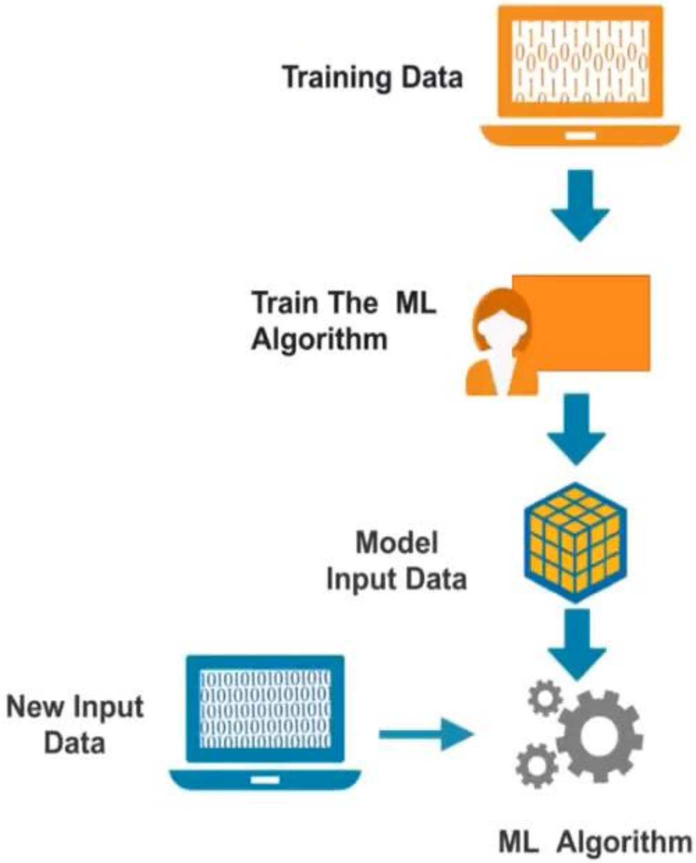


Machine learning is primarily concerned with the accuracy and effectiveness of the *computer system*.

Machine Learning System



Machine Learning System



Machine Learning in a Nutshell

- Tens of thousands of machine learning algorithms.
- Hundreds new every year
- Every machine learning algorithm has three components:

  ❖ **Representation**

  ❖ **Evaluation**

  ❖ **Optimization**

Representation

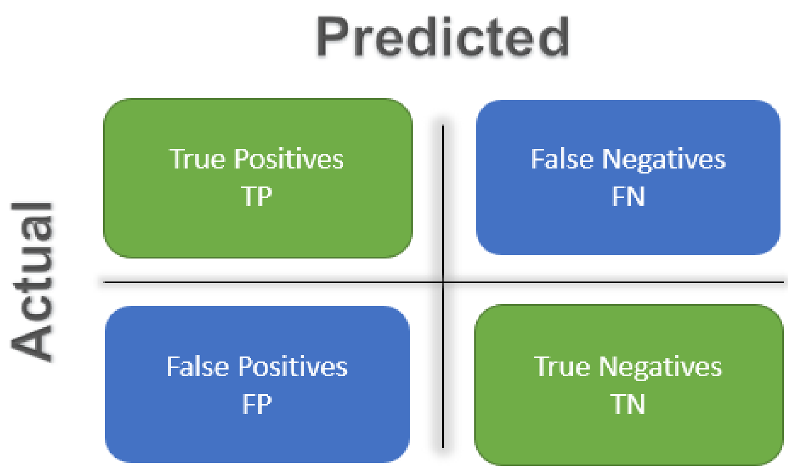
- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

Evaluation

		Prediction c^+	Prediction c^-	Total
Actual	c^+	10	2	12
Actual	c^-	2	8	10
	Total	12	10	**22**

- Confusion Matrix
- Accuracy
- Recall/Sensitivity/True Positive Rate
- Specificity
- Error Rate
- ROC
- Squared error
- Likelihood
- Posterior probability

- Cost / Utility
- Margin
- Specificity
- F-Score
- etc.



Optimization

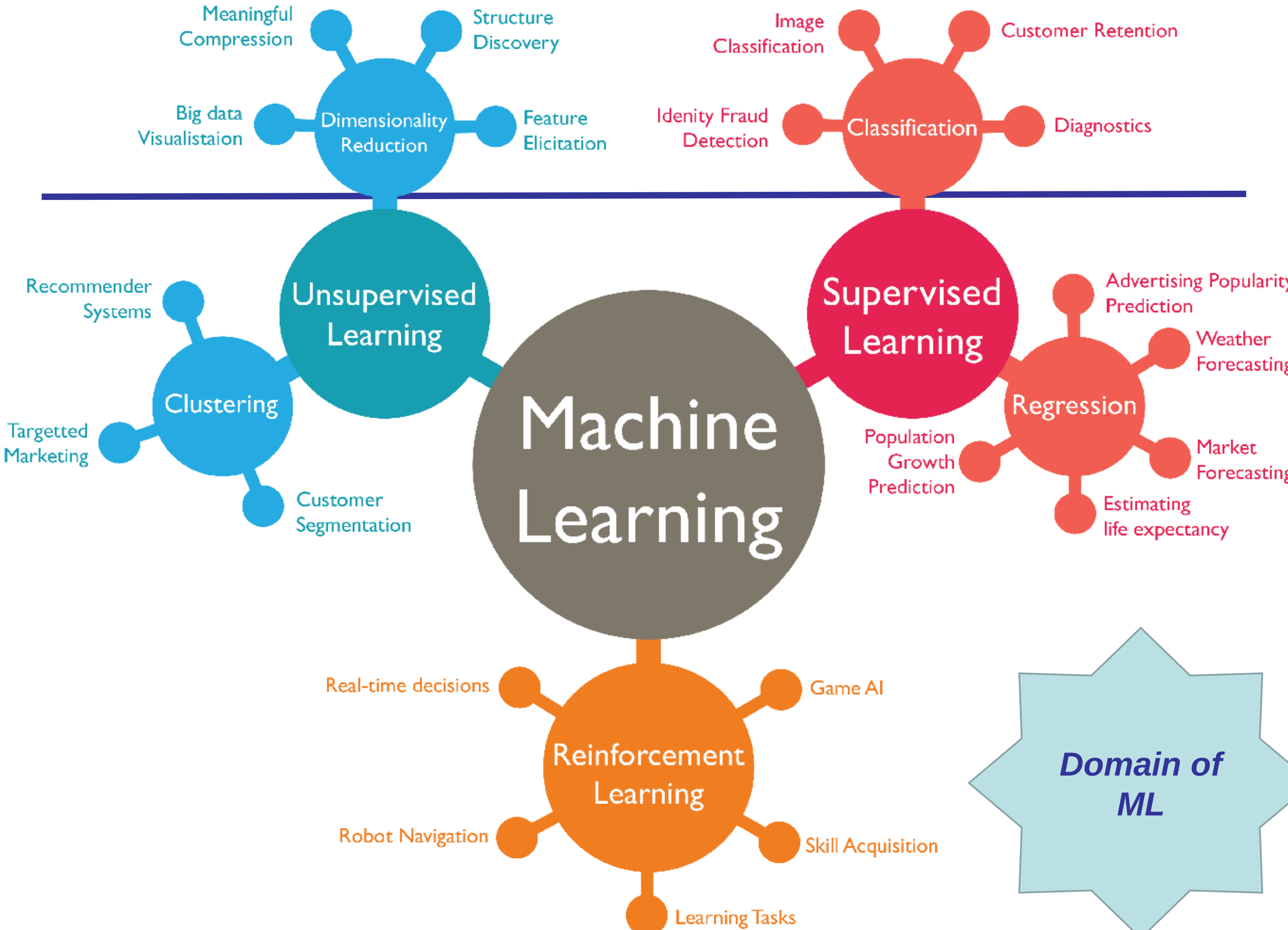
- **Combinatorial optimization**
  - E.g.: Greedy search,
  - finding an optimal object from a finite set of objects
- **Convex optimization**
  - E.g.: Gradient descent
  - Finding the minimum of a function.
- **Constrained optimization**
  - E.g.: Linear programming
  - Optimizing an objective function with respect to some variables in the presence of constraints on those variables

Examples of Machine Learning Problems

- **Pattern Recognition**
  - ❖ Facial identities or facial expressions
  - ❖ Handwritten or spoken words (e.g., Siri)
  - ❖ Medical images
  - ❖ Sensor Data/IoT
- **Optimization**
  - ❖ Many parameters have “hidden” relationships that can be the basis of optimization
- **Pattern Generation**
  - ❖ Generating images or motion sequences
- **Anomaly Detection**
  - ❖ Unusual patterns in the telemetry from physical and/or virtual plants (e.g., data centers)
  - ❖ Unusual sequences of credit card transactions
  - ❖ Unusual patterns of sensor data from a nuclear power plant
    - or unusual sound in your car engine or …
- **Prediction**
  - ❖ Future stock prices or currency exchange rates

Web-based E.g. of ML

- Web data is huge and tasks have to performed with very big datasets often use ML.
  - ❖ especially if the data is noisy or non-stationary.
- Spam filtering, fraud detection:
  - ❖ The enemy adapts so we must adapt too.
- Recommendation systems:
  - ❖ Lots of noisy data. Million dollar prize!
- Information retrieval:
  - ❖ Find documents or images with similar content.
- Data Visualization:
  - ❖ Display a huge database in a revealing way


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Domain of ML

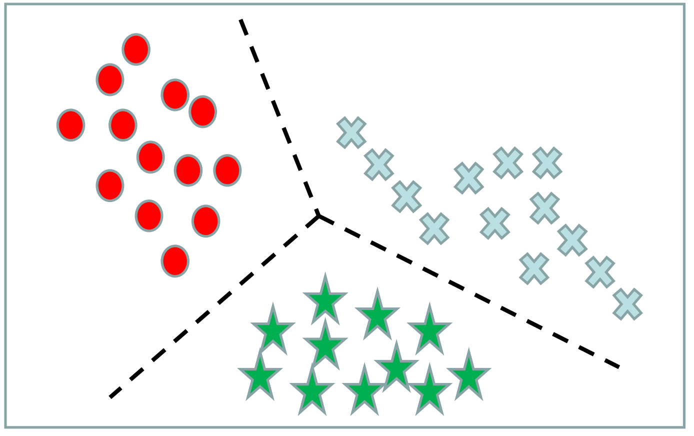
Types of Learning

- **Supervised (inductive) learning**
  - ❖Training data includes desired outputs
- **Unsupervised learning**
  - ❖Training data does not include desired outputs
- **Semi-supervised learning**
  - ❖Training data includes a few desired outputs
- **Reinforcement learning**
  - ❖Rewards from sequence of actions

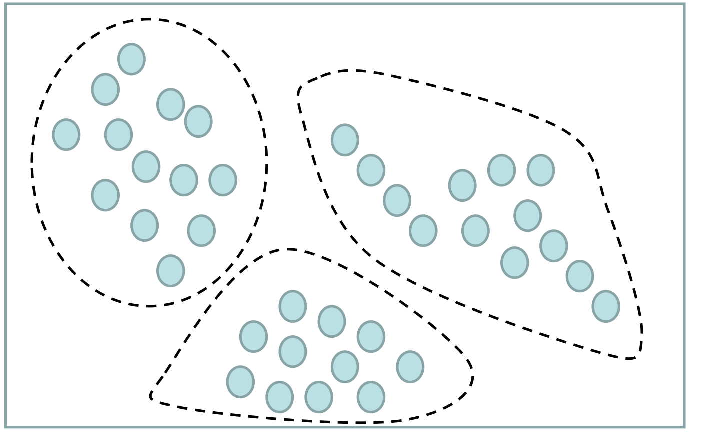
Inductive Learning

- Learner discovers rules by observing examples
- **Given** examples of a function *(X, F(X))*
- **Predict** function *F(X)* for new examples *X*
  - ❖Discrete *F(X)*: Classification
  - ❖Continuous *F(X)*: Regression
  - ❖*F(X)* = Probability(*X*): Probability estimation

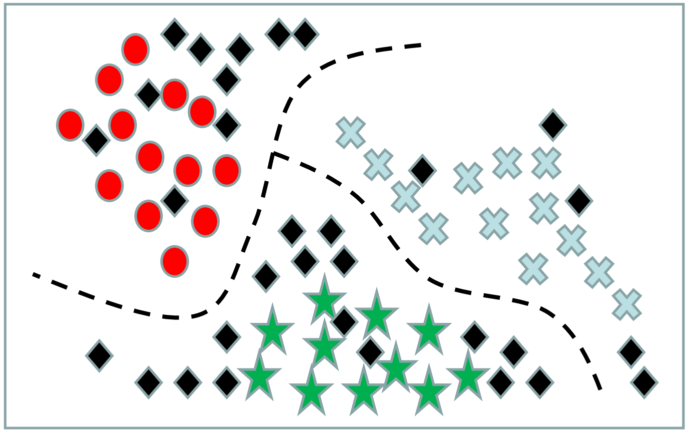
Learning Algorithms



Supervised learning



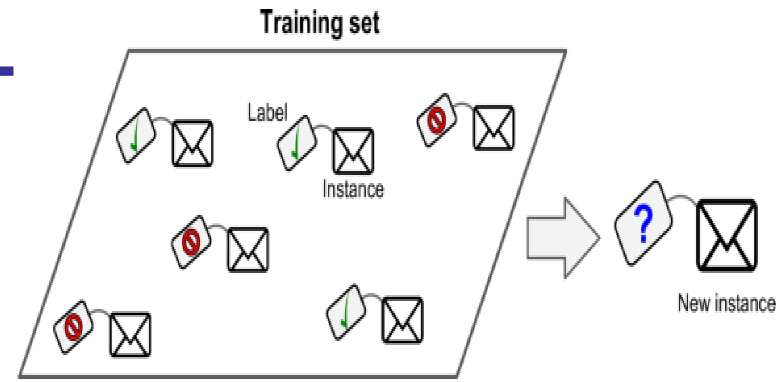
Unsupervised learning

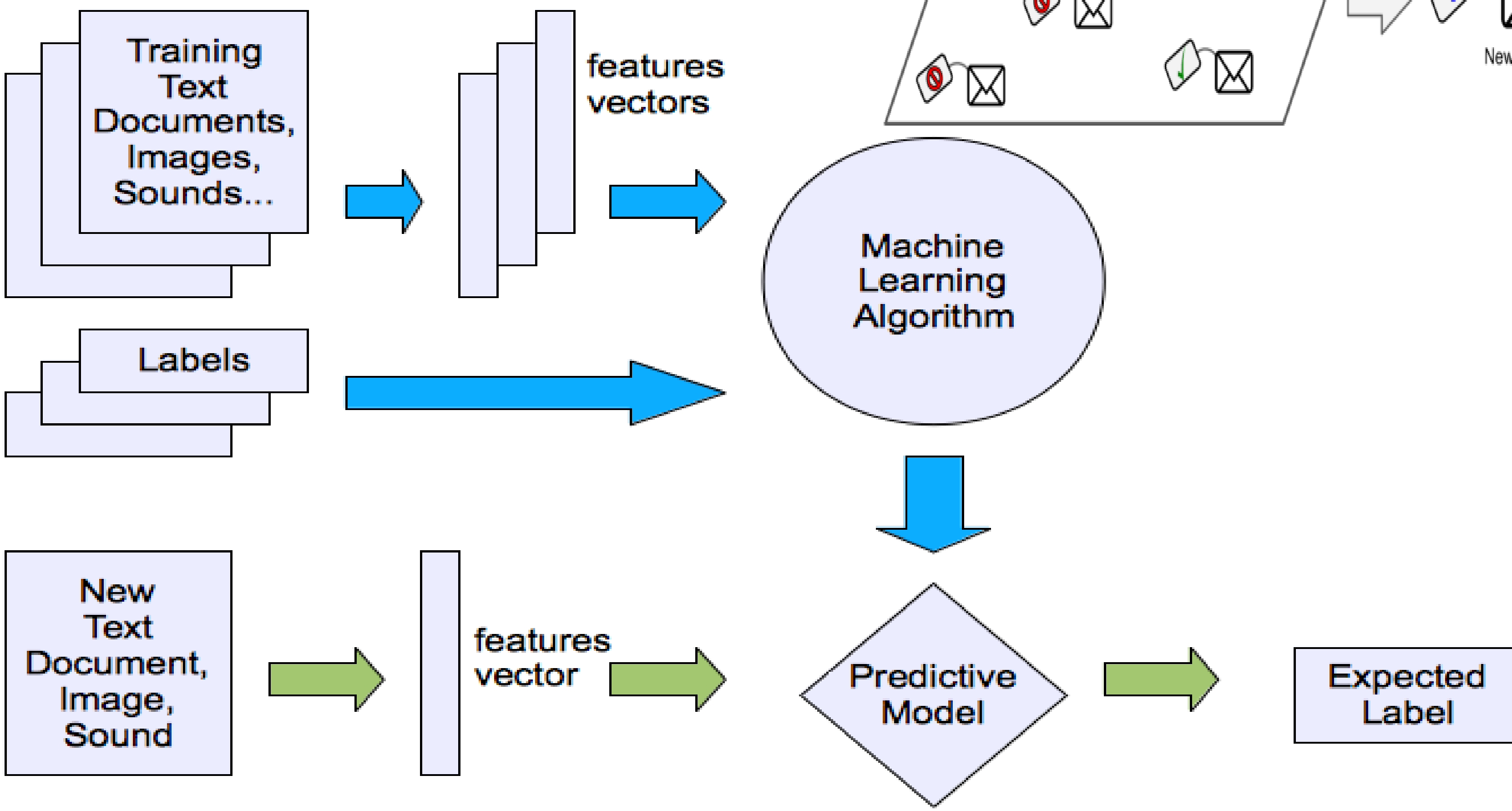


Semi-supervised learning

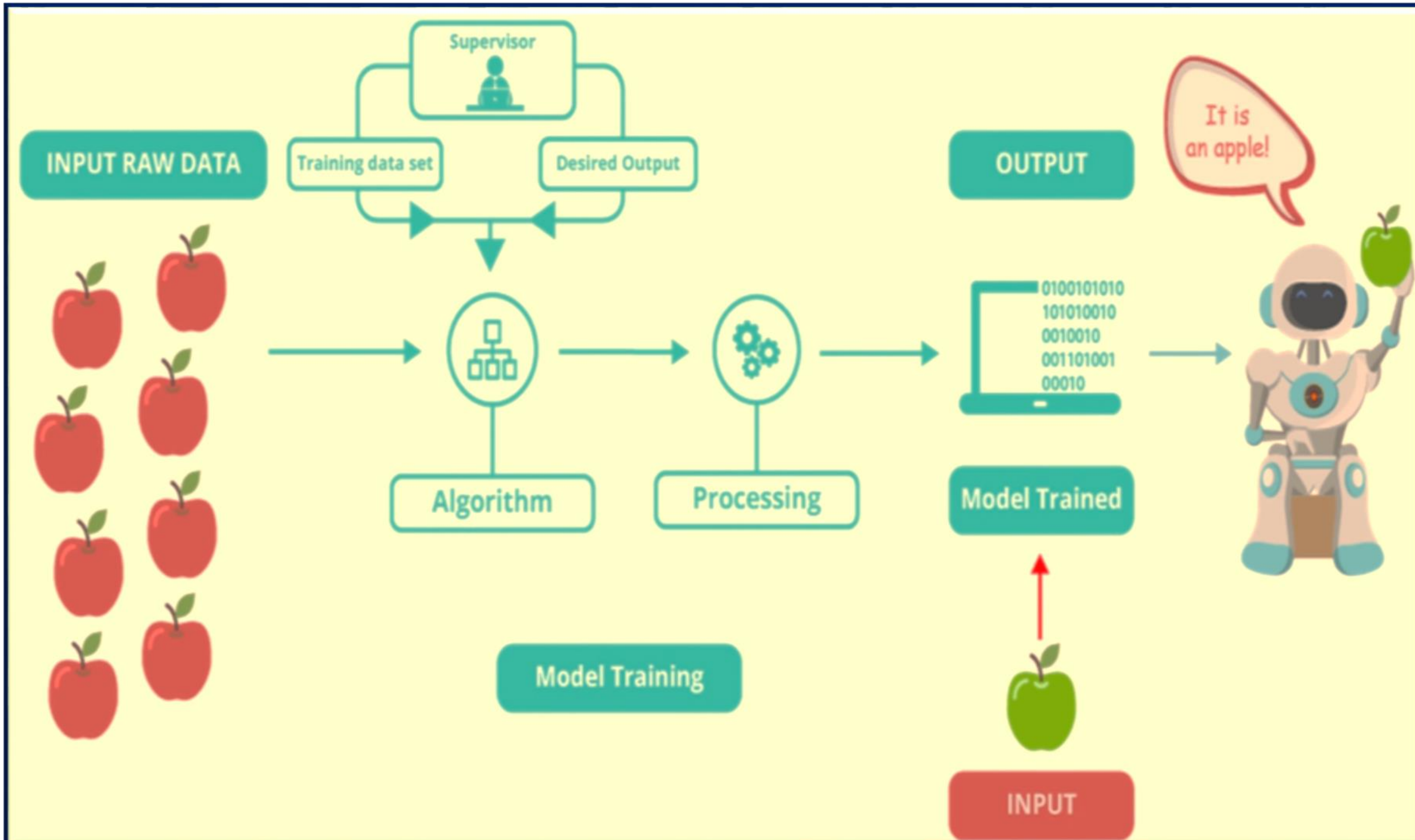
Machine learning structure

- Supervised learning

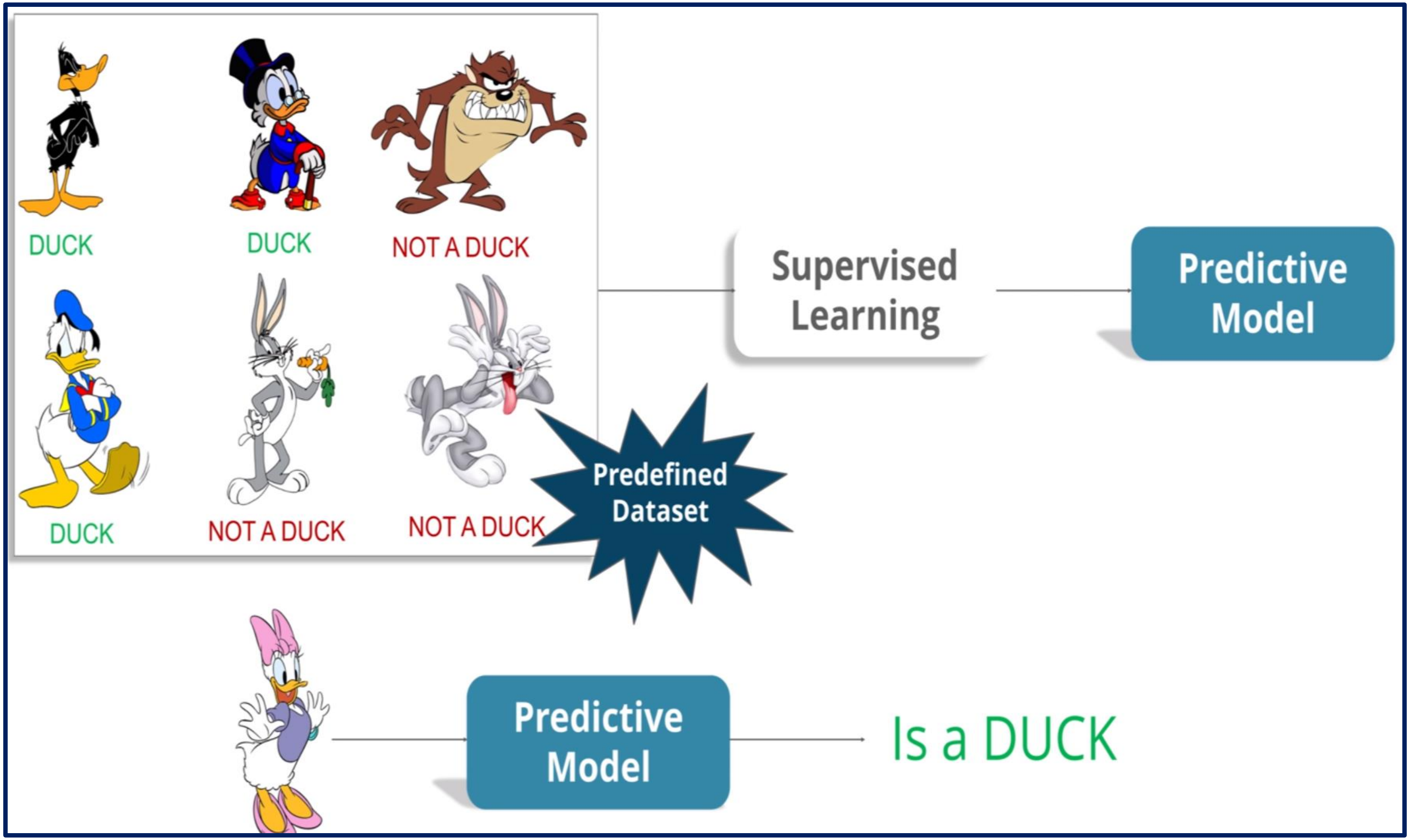




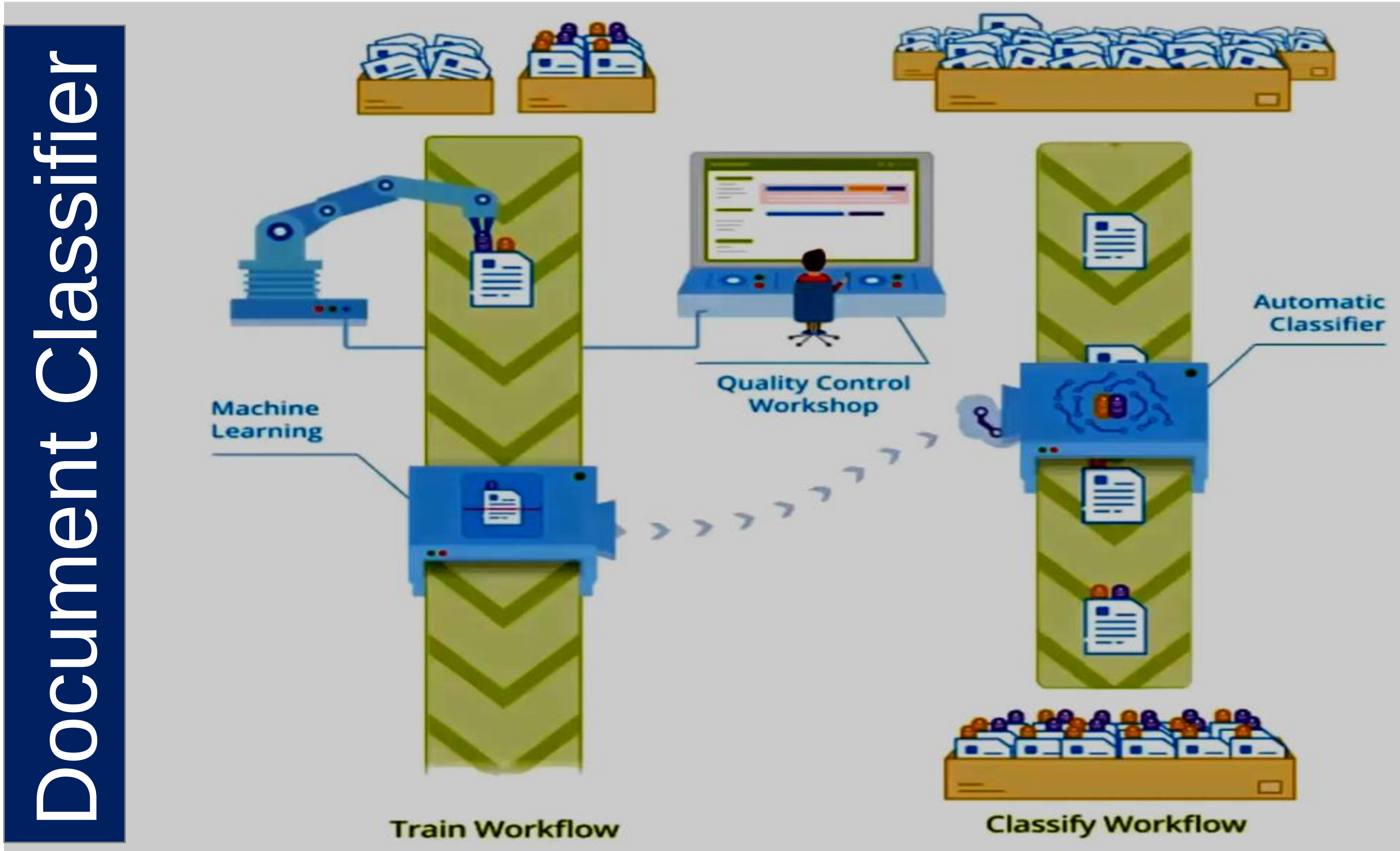
Supervised Learning



E.g. Supervised Learning



E.g. Supervised Learning



Spectrum of Supervision

Unsupervised

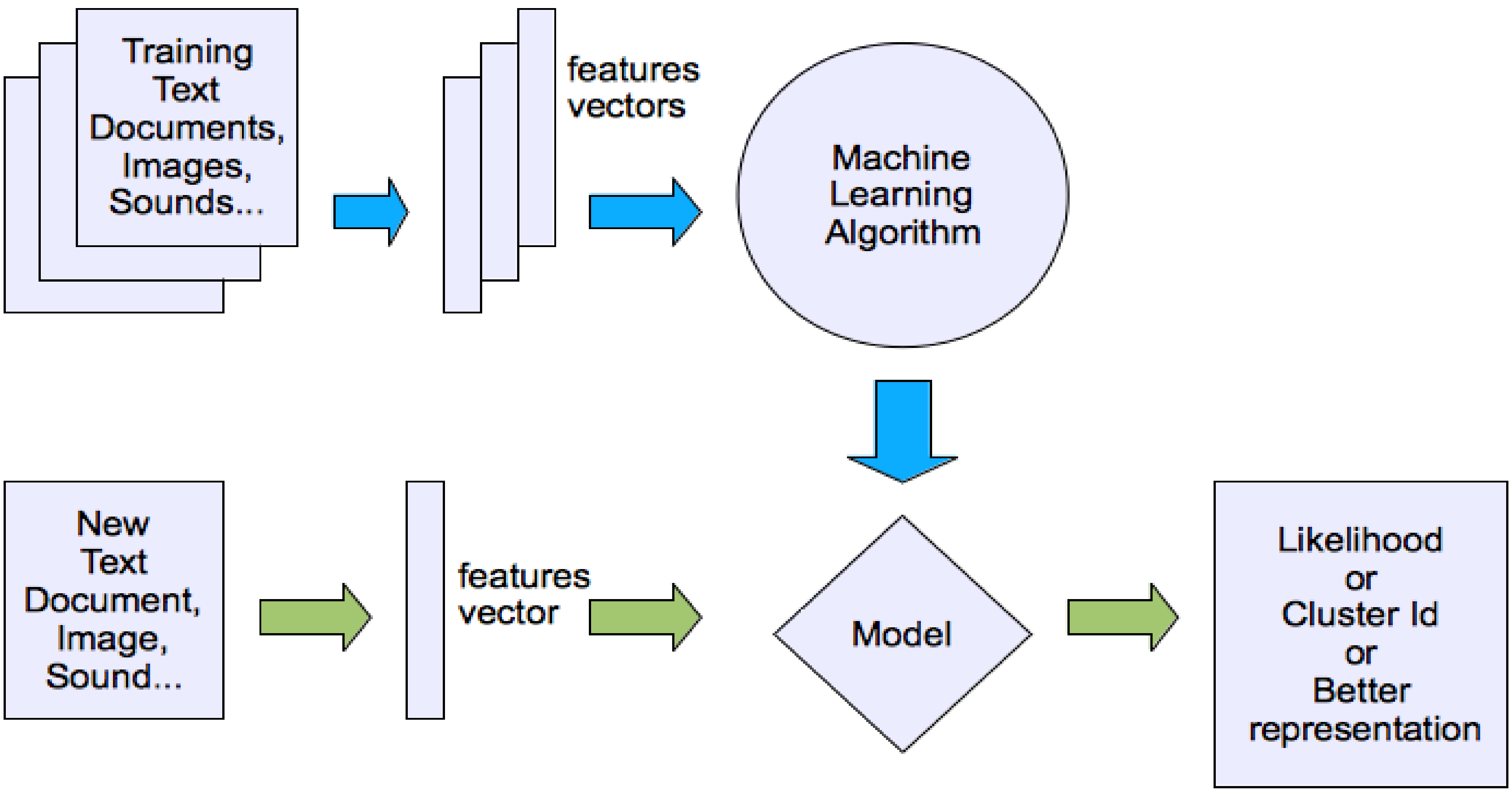
"Weakly" supervised

Fully supervised

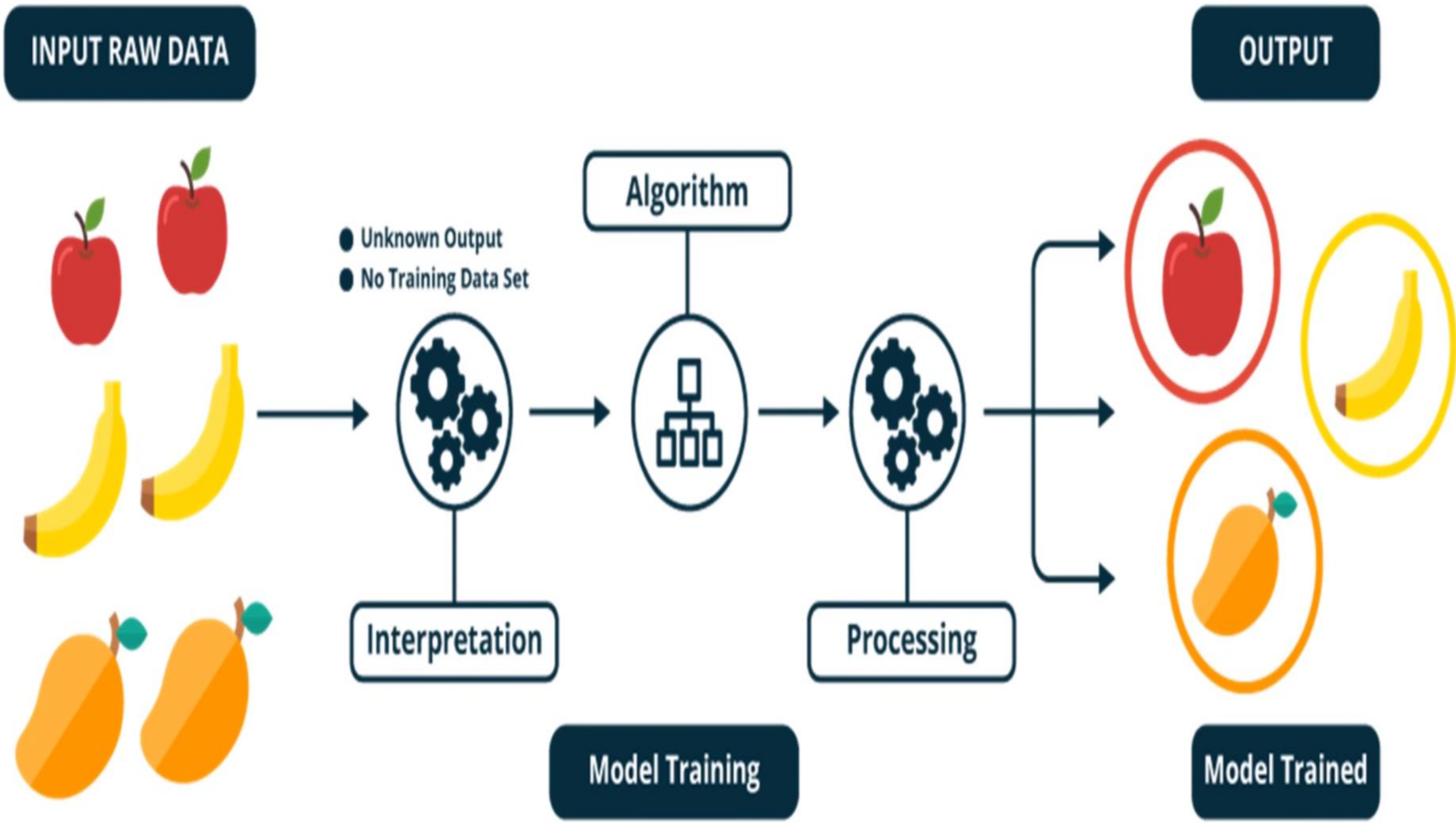
Definition depends on task

Machine learning structure

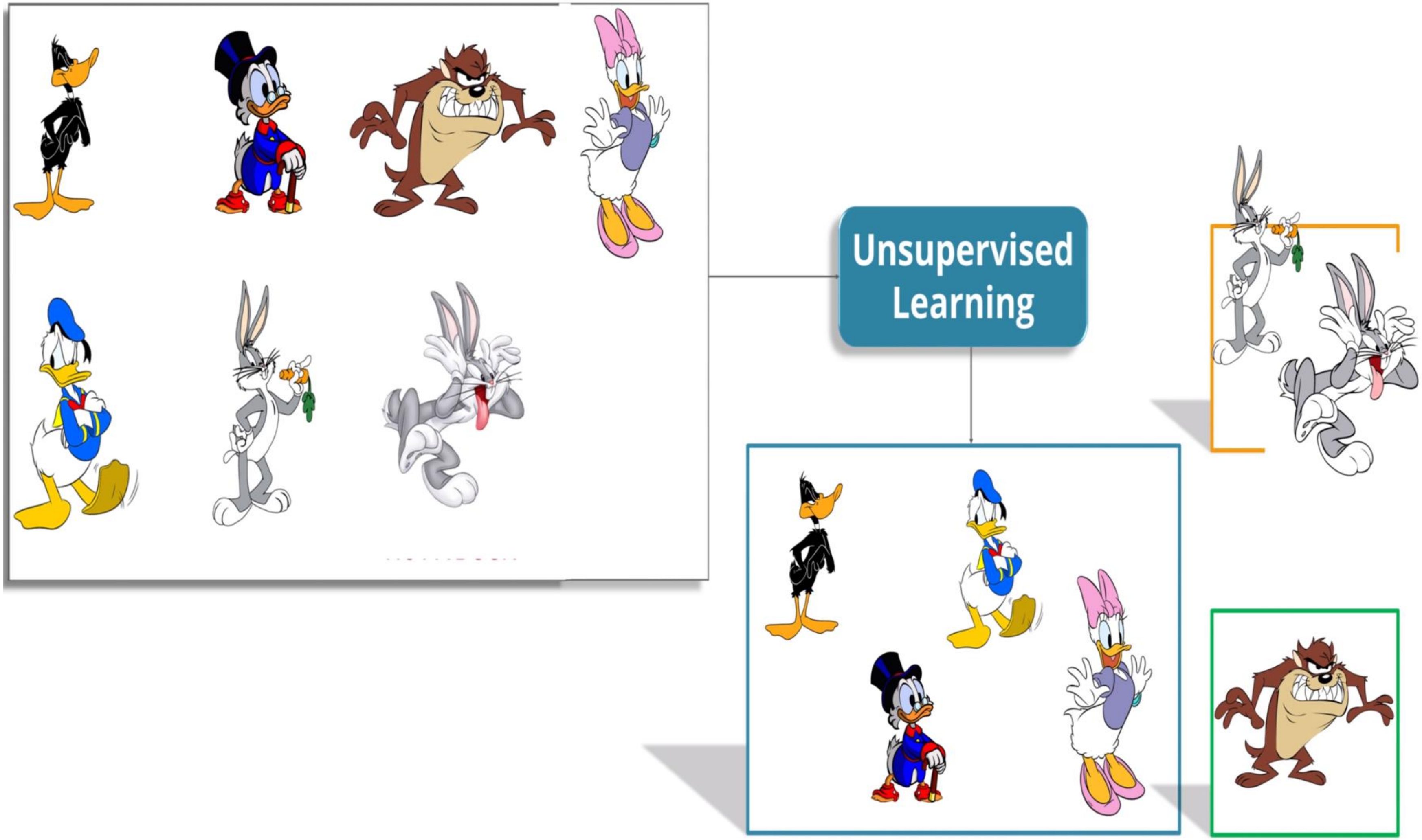
- Unsupervised learning



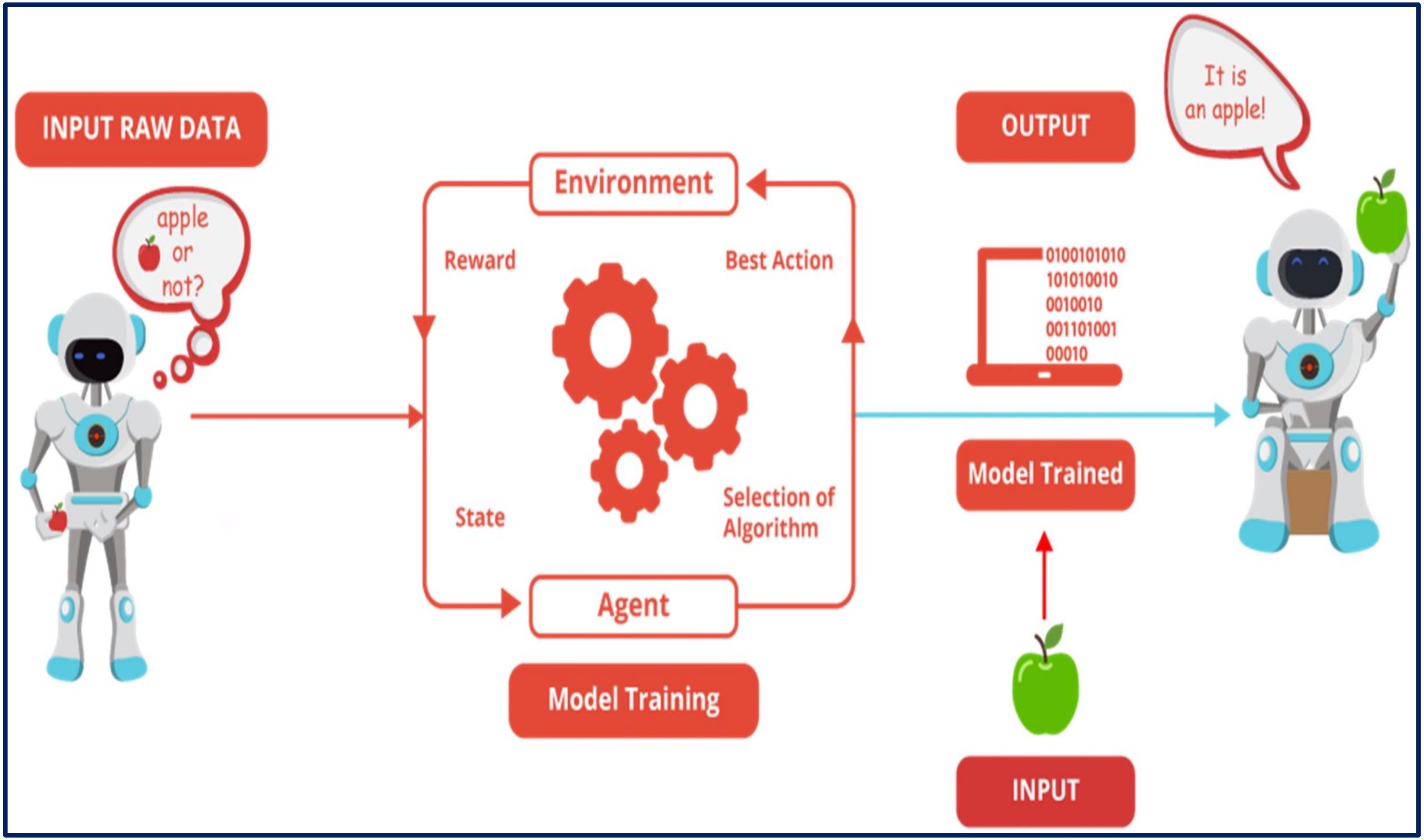
Unsupervised Learning



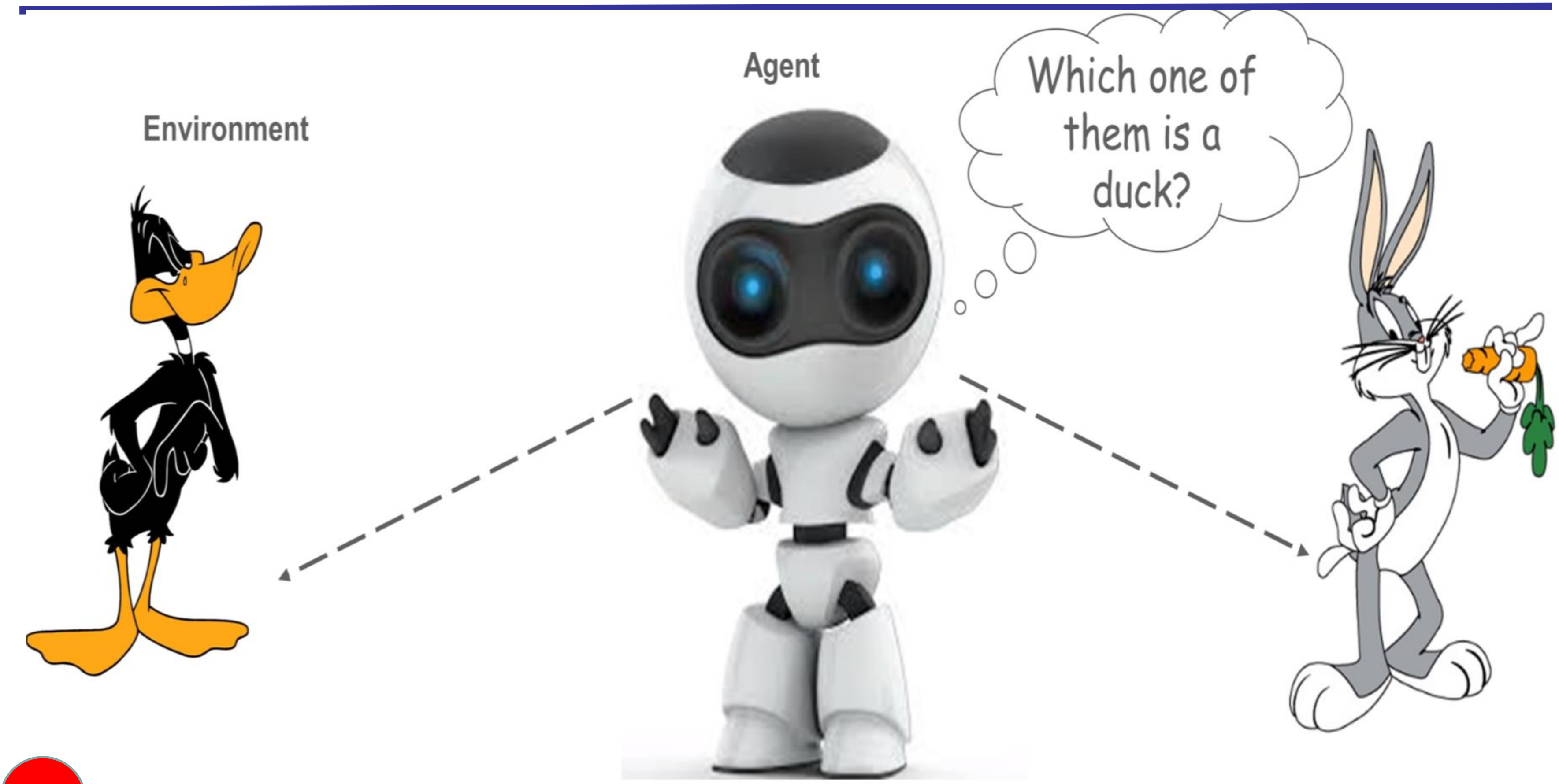
E.g. Unsupervised Learning



Reinforcement Learning

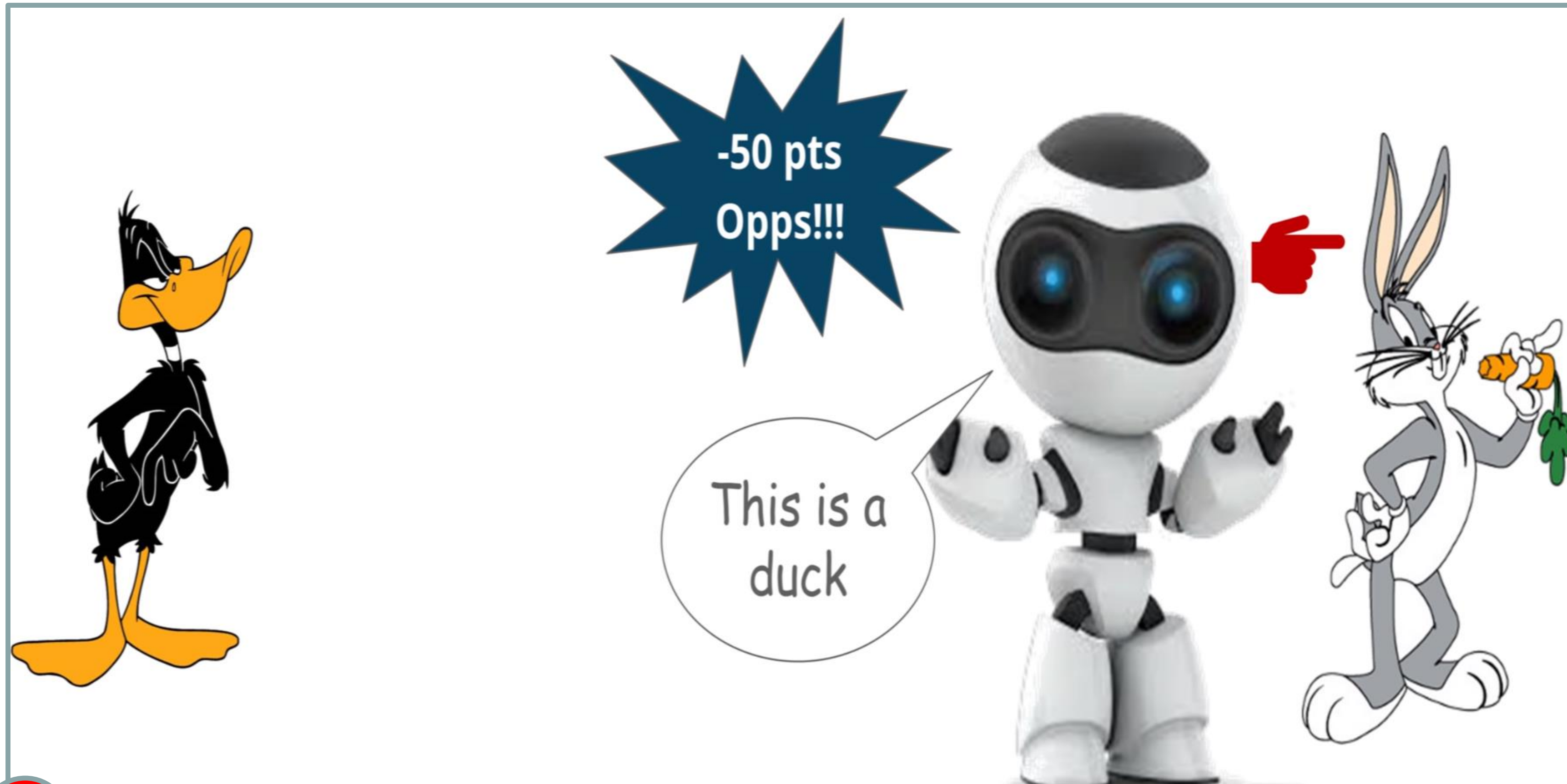


Reinforcement Learning



1. Observe
2. Select Action Using Policy

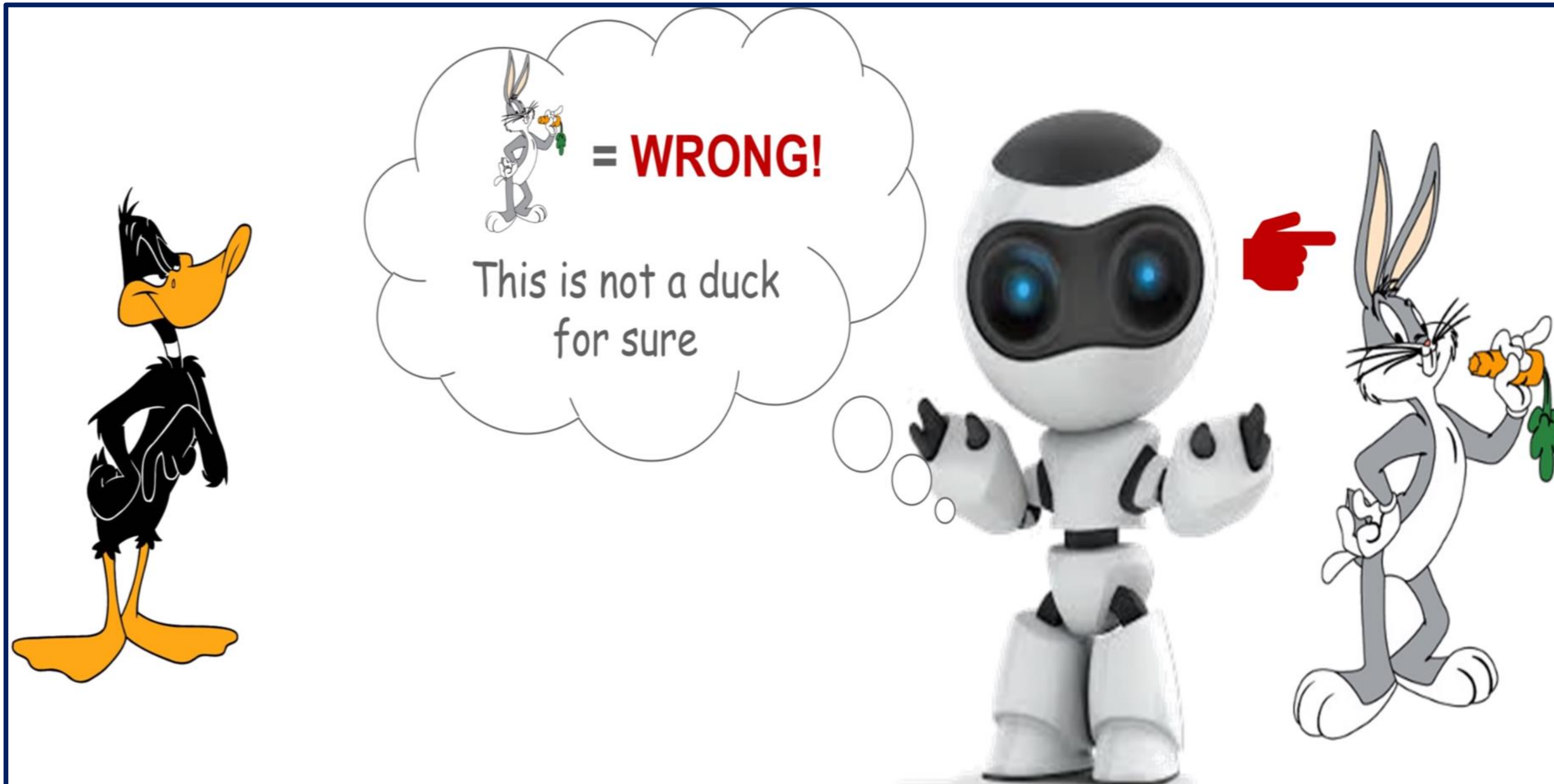
Reinforcement Learning



3 Action!

4 Get Reward or Penalty

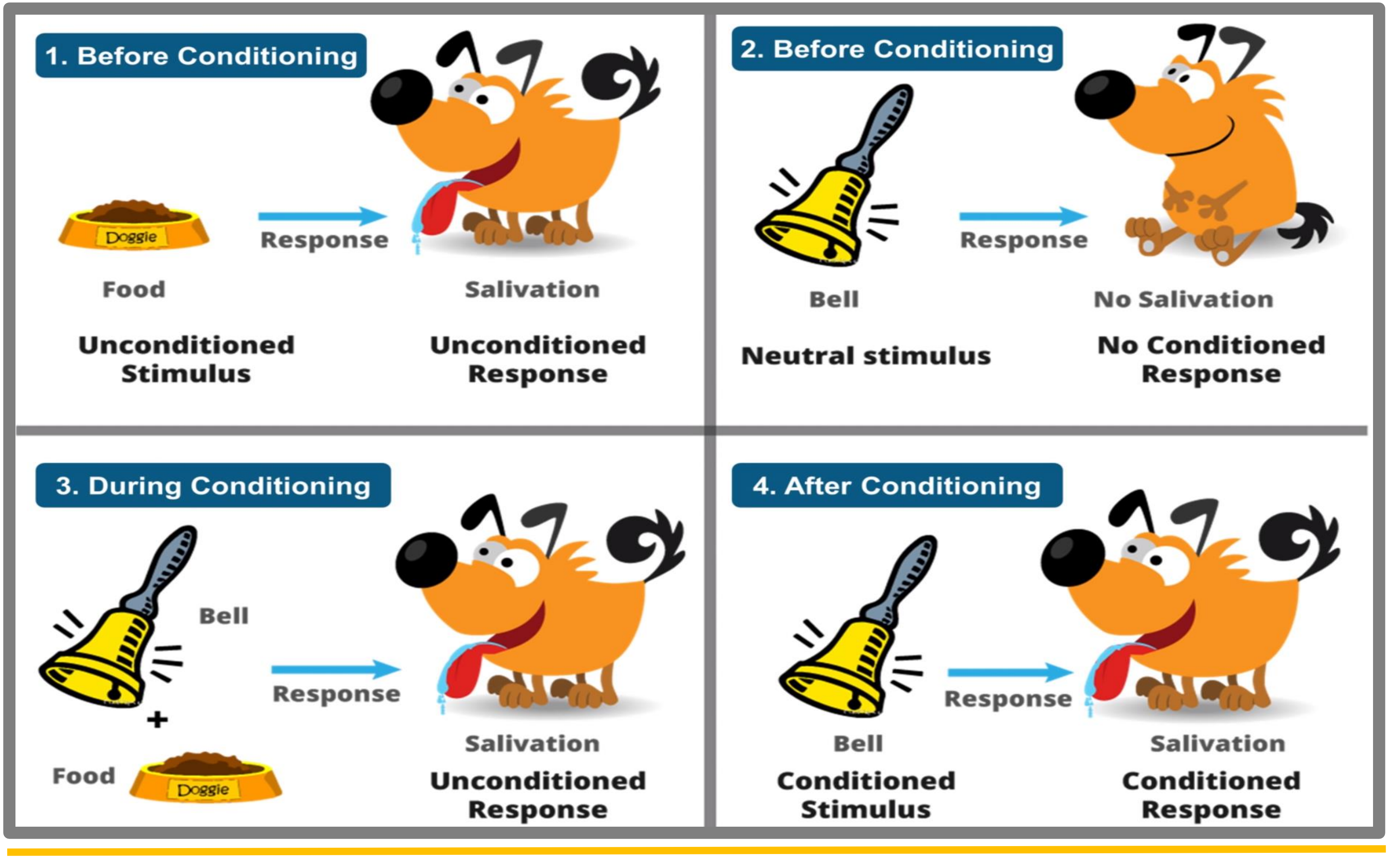
Reinforcement Learning



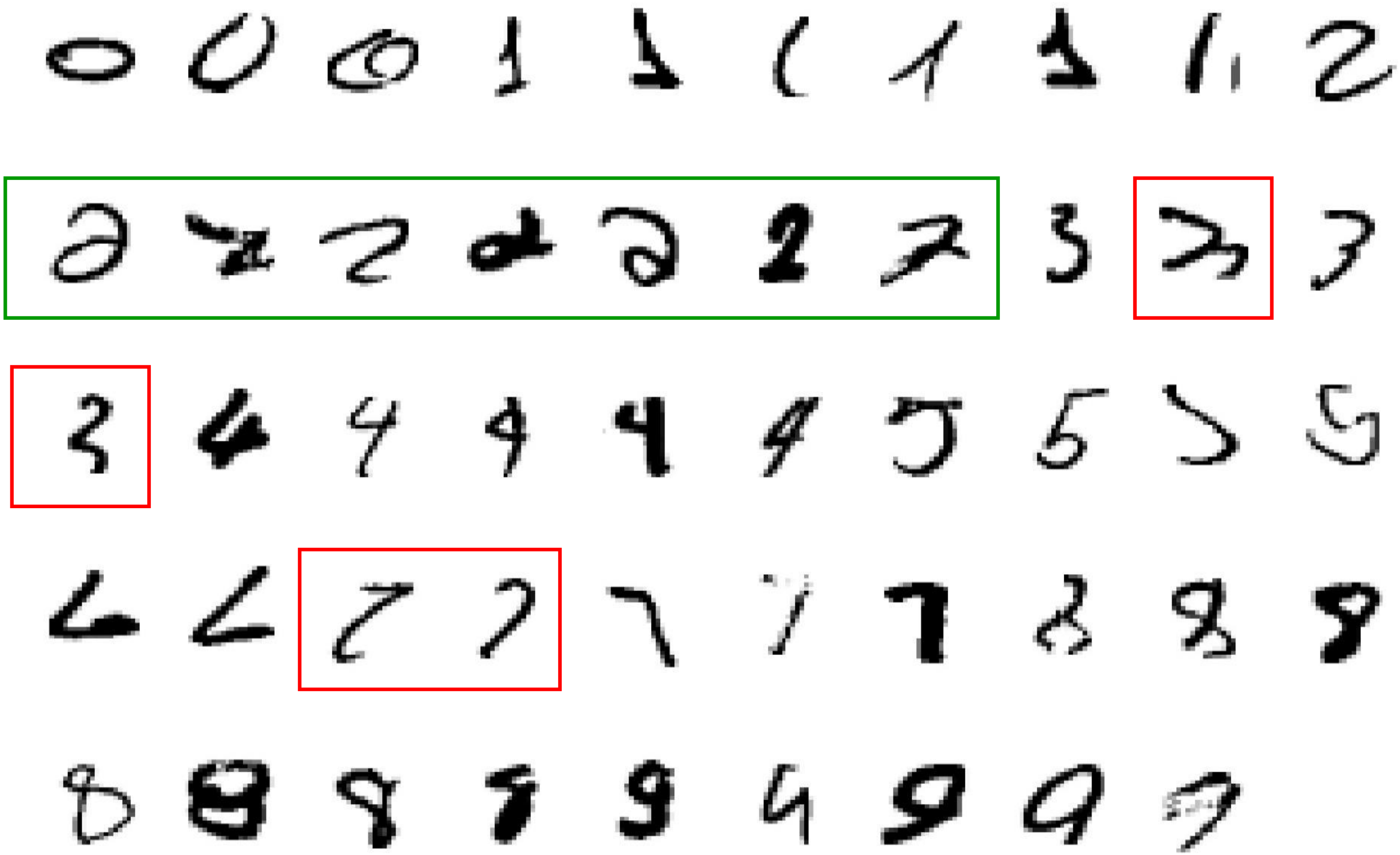
5 Update Policy (learning step)

6 Iterate to get Optimal Policy

E.g. Reinforcement Learning



Why Machine Learning is Hard?



What We'll Cover

- **Fundamentals of Linear Algebra and Probability**
- **Supervised learning**
  - ❖ Linear Regression
  - ❖ Logistic Regression
  - ❖ Decision tree induction
  - ❖ Instance-based learning
  - ❖ Bayesian learning
  - ❖ Neural networks
  - ❖ Support vector machines
  - ❖ Model ensembles
- **Unsupervised learning**
  - ❖ Clustering
  - ❖ Dimensionality reduction
- **Reinforcement Learning**

Data Representation

- **Information systems:**
  - ❖It represents knowledge from RAW data, which is used for decision making.
- **Data warehousing**
  - ❖It provide integrated, consistent and cleaned data to machine learning algorithms.
- **Data Table:**
  - ❖It is used to represent information.

DATA TABLE

- Each row represents a measurements/ observations and each column gives the value of an attribute of the information system for all measurements/ observations.
- Different terms are used to call **'Rows'** information such as **"Instances, examples, samples, measurements, observations, records, patterns, objects, cases, events"**
- Similarly, the '**Column'** information is used to call **"attributes and features"**.

E.G. DATA TABLE

- Consider a patient information in the data table.
- **Features and attributes:** Headache, Muscle-Pain, Temperature. These attributes represented in linguistic form.

Patient	Headache	Muscle Pain	Temperature	Flu
1	NO	YES	HIGH	**YES**
2	YES	YES	HIGH	**YES**
3	YES	YES	VERY HIGH	**YES**
4	NO	YES	NORMAL	**NO**
5	YES	NO	HIGH	**NO**
6	NO	YES	VERY HIGH	**YES**

E.G. DATA TABLE

- An outcome for each observation is known as "a priori" for directed/supervised learning.
- **Decision Attribute:** one distinguished attributes that represent knowledge and information system of this kind called decision system.
- E.g. 'FLU' is decision attribute
- {Flu: Yes}, {Flu; No}.
- **Flu** is a decision attribute with respect to condition attributes: *headache, muscle-pain, temperature.*

E.G. DATA TABLE

- A data file represents inputs as N instances: $S^{(1)}, S^{(2)}, S^{(3)}, \dots\dots\dots S^{(N)}$.

- Each individual instances $S^{(i)}; i = 1, 2, \dots\dots N$ that provides the input to the machine learning tools is characterized by its predefined values for a set of features/attributes $x_1, x_2, \quad x_3, \dots\dots\dots \quad x_n$ *or* $x_j; j = 1,2,3, \dots\dots n$

E.G. DATA TABLE

x_j / $S^{(i)}$	x_1	x_2	x_3	x_3	…….	x_n	**Decision** y
$S^{(1)}$							
$S^{(2)}$							
$S^{(3)}$							
$S^{(4)}$							
. .							
$S^{(\mathbf{N})}$							

Training experience is available in the form of N examples: $S^{(i)} \in S; i = 1, 2, 3 \ldots N$. Where S is a set of possible instances, which come from real world.

DATA REPRESENTATION

- An instance can be represented for n attribute/features: $x_j; j = 1, 2, 3, \ldots \ldots \ldots n$.
- These features can be visualize as n numerical features as a point in n -dimensional state space $\Re^n$.
- $\boldsymbol{x} = [x_1\ x_2\ x_3\ x_4 \ldots \ldots x_n]^{\boldsymbol{T}} \in \Re^n$. The set X is a finite set of feature vector $x^{(i)}$ for all possible instance.
- Also visualized as X region in the state space $\Re^n$ to which instance belongs, i.e. $X \subset \Re^n$

DATA REPRESENTATION

- Here, $x^{(i)}$ is a representation of $s^{(i)}$, X is the *representation space*.
- The pair of (S, X) constitutes the information system. Where S is non-empty set of instances and X is non-empty features.
- Here, index i represents instances and j represents features.
  - ❖ $\left\{\boldsymbol{s}^{(\boldsymbol{i})}; \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \ldots\ldots \boldsymbol{N}\right\} \in \boldsymbol{S}$
  - ❖ $\left\{\boldsymbol{x}^{(\boldsymbol{i})}; \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \ldots\ldots \boldsymbol{N}\right\} \in \boldsymbol{X}$ *(set of features)*
  - ❖ $\left\{\boldsymbol{x}_{\boldsymbol{j}}^{(\boldsymbol{i})}; \boldsymbol{j} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \ldots\ldots \boldsymbol{N}\right\} = \boldsymbol{x}^{(\boldsymbol{i})}$
  - ❖ **Features $x_j; j = 1, 2, \ldots. n$, may be viewed as state variables and feature vector x as a state vector in n-dimensional space.**

DATA REPRESENTATION

- For every feature x_j a set of values can be written as $V_{x_j} \in R$ and called as domain of $x_j; j = 1, 2, \dots . n$.
- $V_{x_j}^{(i)} \in V_{x_j}; i = 1, 2, \dots N$.
- The ***tuple*** (S,X,Y) may be constituted and this is called decision system.

Thank You