INTRODUCTION TO MACHINE LEARNING

19CSC312A

Artificial Intelligence

B. Tech. CSE, 2021

Course Objectives

This course will enable students to

1. Define machine learning and understand the basic theory underlying machine learning.
2. Differentiate supervised, unsupervised and reinforcement learning
3. Understand the basic concepts of learning and decision trees.
4. Understand neural networks and Bayesian techniques for problems appear in machine learning
5. Understand the instant based learning and reinforced learning
6. Perform statistical analysis of machine learning techniques.

Introduction

Ever since computers were invented, we have wondered whether they might be made to learn. If we could understand how to program them to learn-to improve automatically with experience-the impact would be dramatic.

- Imagine computers learning from medical records which treatments are most effective for new diseases
- Houses learning from experience to optimize energy costs based on the particular usage patterns of their occupants.
- Personal software assistants learning the evolving interests of their users in order to highlight especially relevant stories from the online morning newspaper

Examples of Successful Applications of Machine Learning

- Learning to recognize spoken words
- Learning to drive an autonomous vehicle
- Learning to classify new astronomical structures
- Learning to play world-class backgammon

Why is Machine Learning Important?

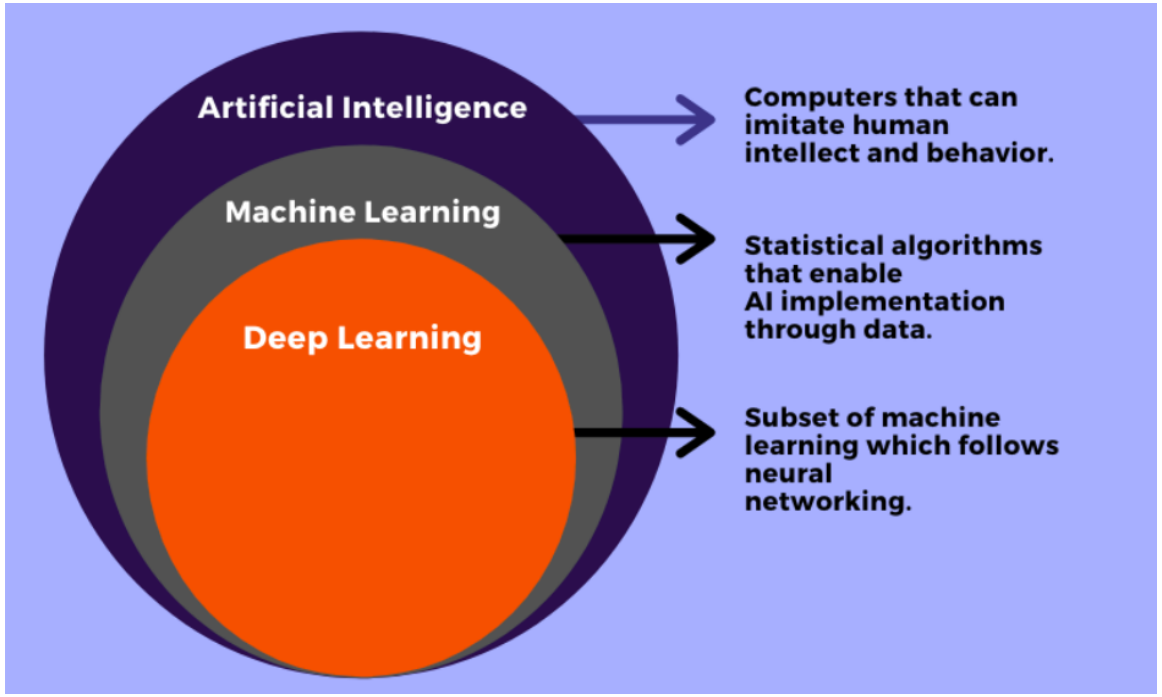
- Some tasks cannot be defined well, except by examples (e.g., recognizing people).
- Relationships and correlations can be hidden within large amounts of data. Machine Learning/Data Mining may be able to find these relationships.
- Human designers often produce machines that do not work as well as desired in the environments in which they are used.
- The amount of knowledge available about certain tasks might be too large for explicit encoding by humans (e.g., medical diagnostic).
- Environments change over time.
- New knowledge about tasks is constantly being discovered by humans. It may be difficult to continuously re-design systems “by hand”.

Areas of Influence for Machine Learning

- **Statistics:** How best to use samples drawn from unknown probability distributions to help decide from which distribution some new sample is drawn?
- **Brain Models**: Non-linear elements with weighted inputs (Artificial Neural Networks) have been suggested as simple models of biological neurons.
- **Adaptive Control Theory**: How to deal with controlling a process having unknown parameters that must be estimated during operation?
- **Psychology**: How to model human performance on various learning tasks?
- **Artificial Intelligence**: How to write algorithms to acquire the knowledge humans are able to acquire, at least, as well as humans?
- **Evolutionary Models**: How to model certain aspects of biological evolution to improve the performance of computer programs?

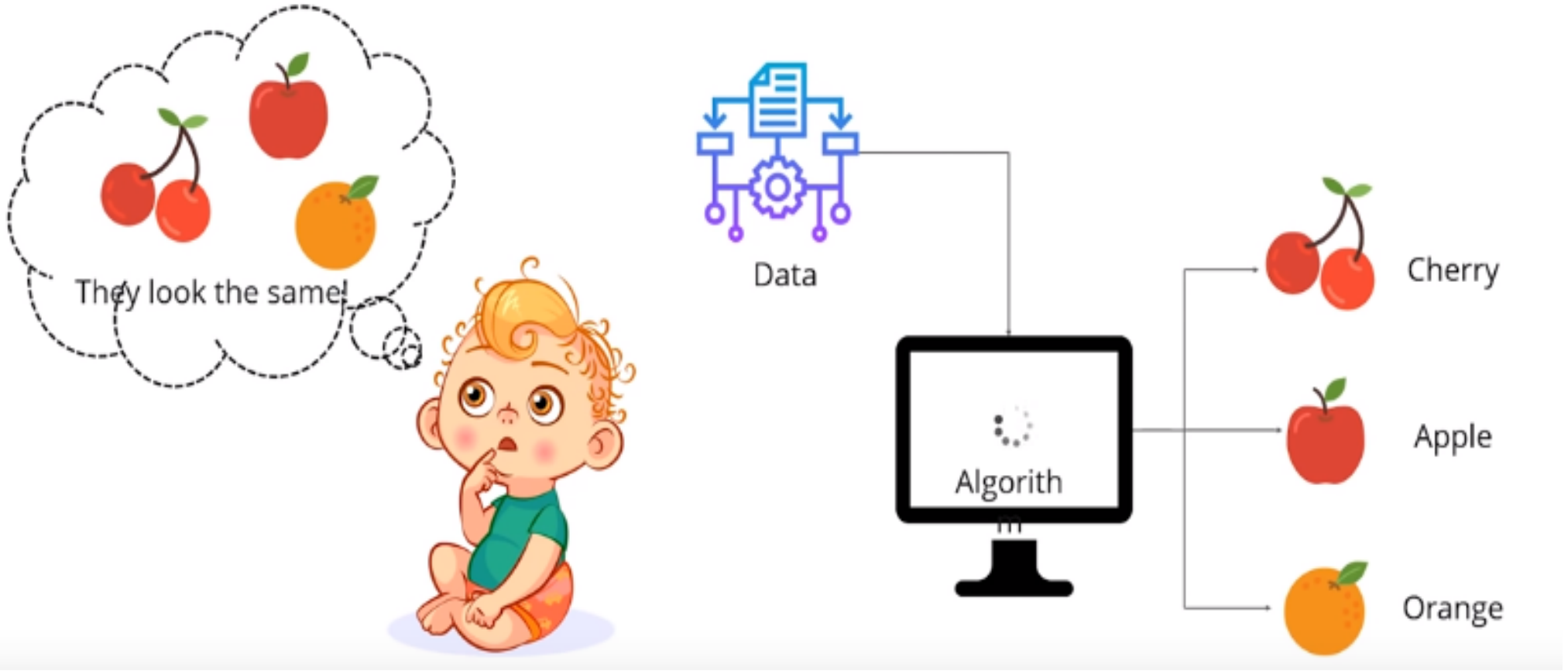
Machine Learning

- Machine Learning
  - Grew out of work in AI
  - New capability for computers
- Examples:
  1. Database mining
     - Large datasets from growth of automation/web.
     - E.g., Web click data, medical records, biology, engineering
  2. Applications can't program by hand.
     - E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision.
  3. Self-customizing programs
     - E.g., Amazon, Netflix product recommendations
  4. Understanding human learning (brain, real AI).



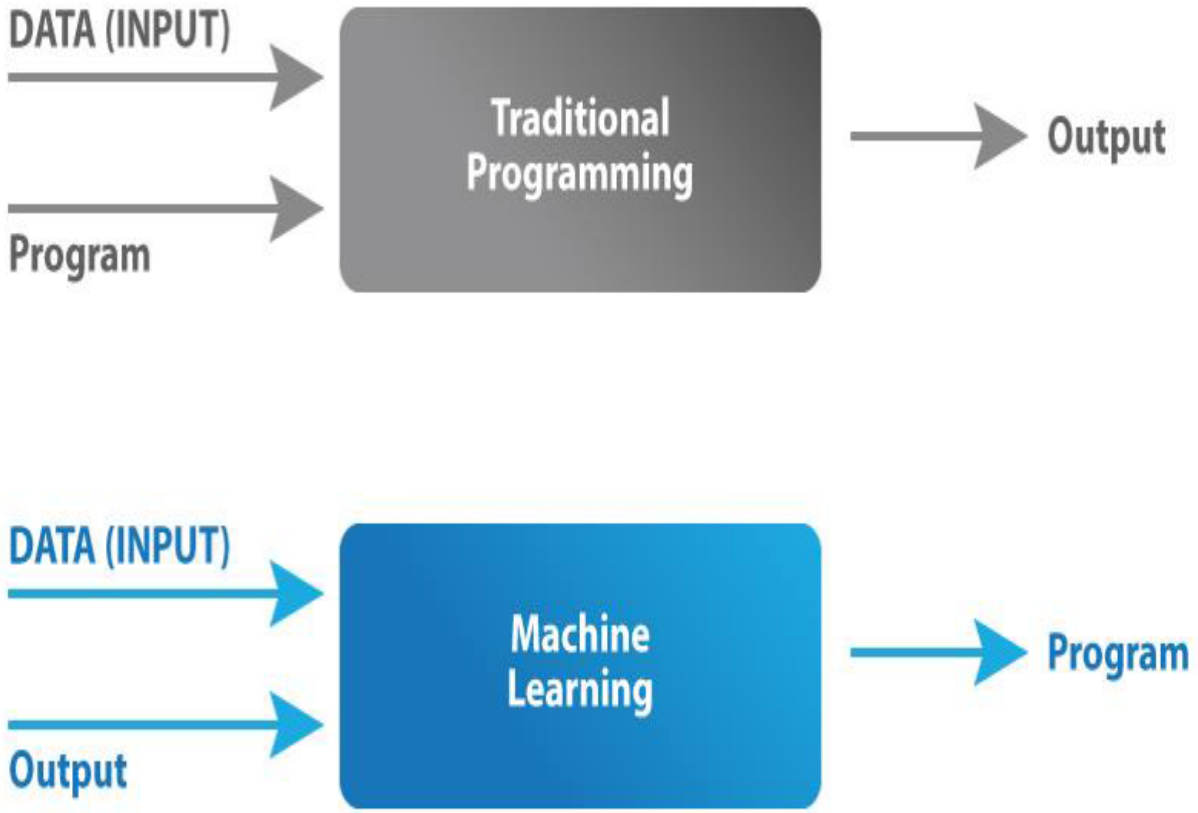
Machine Learning (ML)

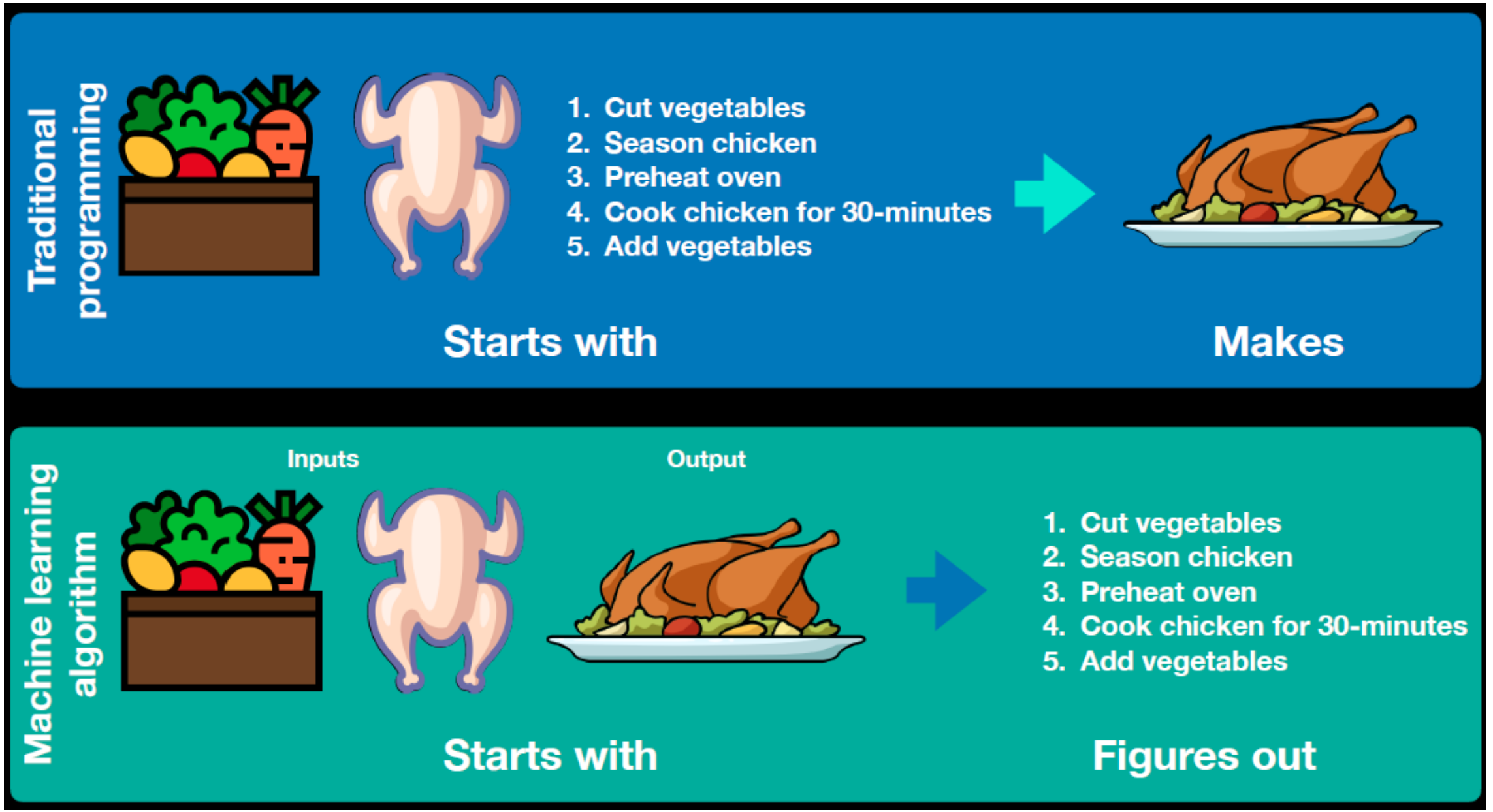
Machine learning is a subset of artificial intelligence (AI) which provides machines the ability to learn automatically & improve from experience without being explicitly programmed.



How is it different from traditional programming

- In traditional programming, we would feed the input data and a well written and tested program into a machine to generate output.
- In machine learning, input data along with the output associated with the data is fed into the machine during the learning phase, and it works out a program for itself. To understand this better, refer to the illustration below:





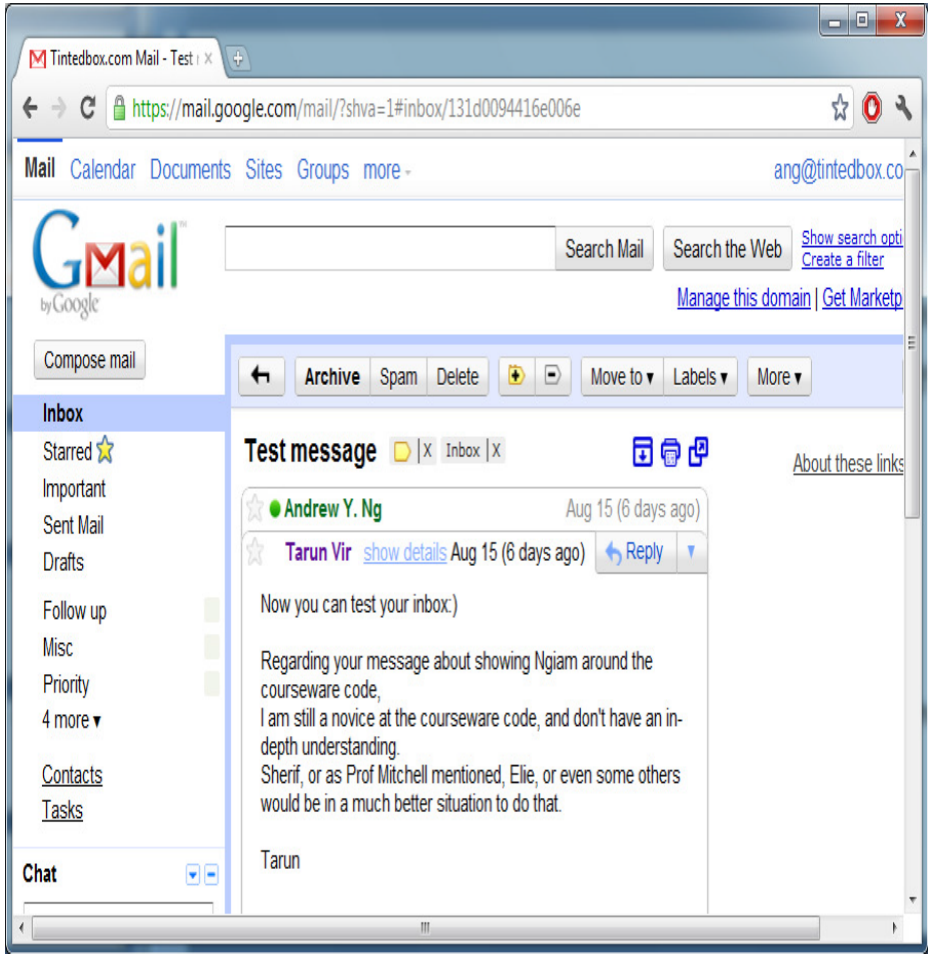
Introduction to machine learning

- Three definitions of Machine Learning are offered.

1. Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.
2. Tom Mitchell (1998) Well-posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.
3. Machine Learning is an application of Artificial Intelligence (AI) which enables a program(software) to learn from the experiences and improve their self at a task without being explicitly programmed.

- "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."
- Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

1. Classifying emails as spam or not spam. ➔ T
2. Watching you label emails as spam or not spam. ➔E
3. The number (or fraction) of emails correctly classified as spam/not spam.➔P
4. None of the above—this is not a machine learning problem.



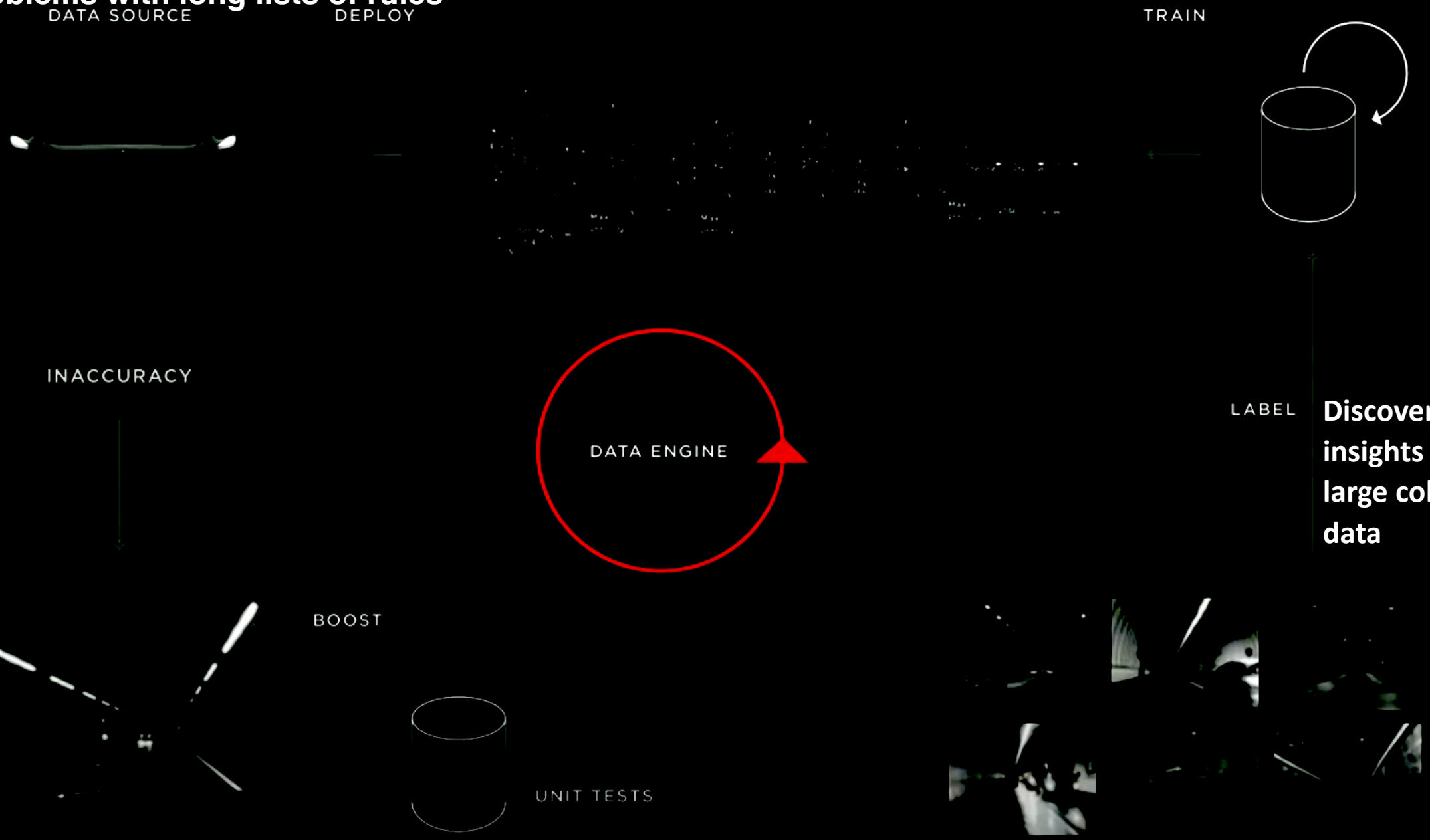
Example

- For example, how would you write a program that can identify fruits based on their various properties, such as color, shape size or any other property?
- One approach is to hardcode everything, make some rules and use them to identify the fruits. This may seem the only way and work but one can never make perfect rules that apply on all cases. This problem can be easily solved using machine learning without any rules which makes it more robust and practical
- we can say that Machine Learning is the study of making machines more human-like in their behaviour and decision making by giving them the ability to learn with minimum human intervention, i.e., no explicit programming. Now the question arises, how can a program attain any experience and from where does it learn? The answer is data. Data is also called the fuel for Machine Learning and we can safely say that there is no machine learning without data.

What machine learning is good for

- **Problems with long lists of rules**—when the traditional approach fails, machine learning may help.
- **Continually changing environments**—machine learning can adapt ('learn') to new scenarios.
- **Discovering insights within large collections of data**—can you imagine trying to go through every transaction your (large) company has ever had by hand?

Problems with long lists of rules



Discovering insights within large collections of data

Continually changing environments

TESLA

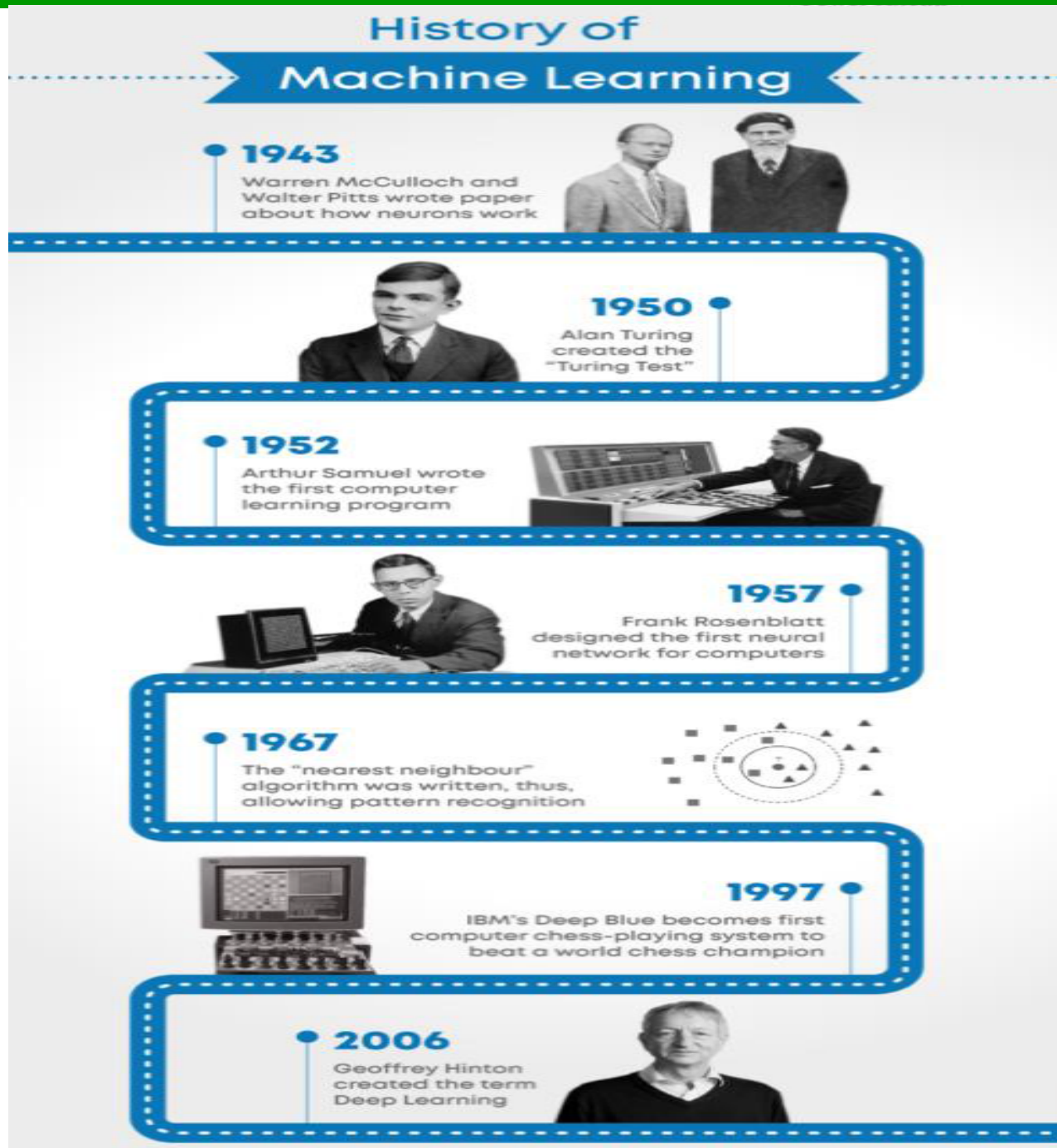
Source: Tesla Autonomy Day video

Why do we need Machine Learning

- Machine learning can automate many tasks, especially the ones that only humans can perform with their innate intelligence. Replicating this intelligence to machines can be achieved only with the help of machine learning.
- Businesses can automate routine tasks
- Automating and quickly create models for data analysis
- Helps in creating models that can process and analyze large amounts of complex data to deliver accurate results.
- Image recognition, text generation, and many other use-cases are finding applications in the real world.

History of Machine Learning

- Machine learning has been here for over 70 years now.
- It all started in 1943, when neurophysiologist Warren McCulloch and mathematician Walter Pitts wrote a paper about neurons, and how they work. They decided to create a model of this using an electrical circuit, and therefore, the neural network was born.
- In 1950, Alan Turing created the “Turing Test” to determine if a computer has real intelligence. To pass the test, a computer must be able to fool a human into believing it is also human.
- In 1952, Arthur Samuel wrote the first computer learning program. The program was the game of checkers,



History of Machine Learning

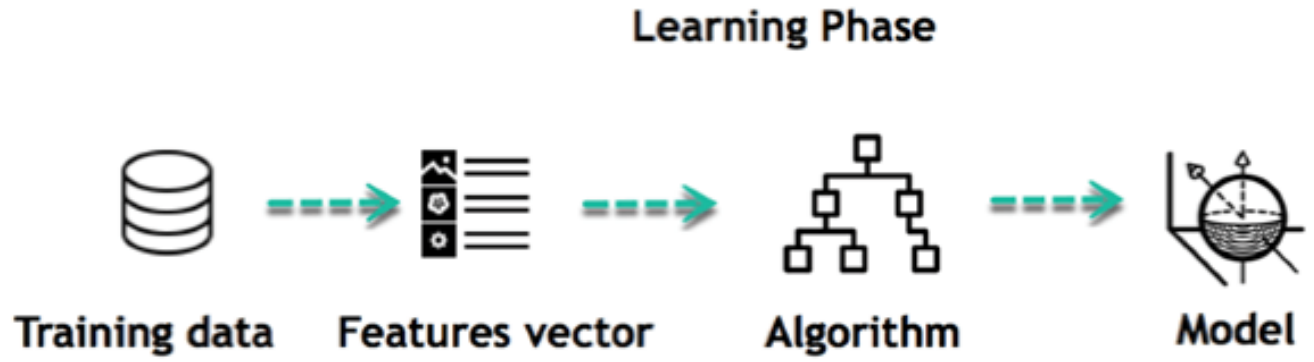
- In 1957, Frank Rosenblatt designed the first neural network for computers (the perceptron), which simulates the thought processes of the human brain.
- Later, in 1967, the “nearest neighbour” algorithm was written, allowing computers to begin using very basic pattern recognition. This could be used to map a route for travelling salesmen, starting at a random city but ensuring they visit all cities during a short tour.
- In the 1990s we saw a big change. Now work on machine learning shifted from a **knowledge-driven approach** to a **data-driven approach**. Scientists began to create programs for computers to analyze large amounts of data and draw conclusions or “learn” from the results.
- In 1997, IBM’s Deep Blue became the first computer chess-playing system to beat a reigning world chess champion. Deep Blue used the computing power in the 1990s to perform large-scale searches of potential moves and select the best move.
- In 2006, Geoffrey Hinton created the term “deep learning” to explain new algorithms that help computers distinguish objects and text in images and videos.

Features of Machine Learning

- **Automation**: Gmail account, there is a spam folder that contains all the spam emails.
- **Improved customer experience**: For any business, one of the most crucial ways to drive engagement, promote brand loyalty and establish long-lasting customer relationships is
- by providing a customized experience and providing better services. Machine Learning helps us to achieve both of them.
- **Automated data visualization**: companies like Google, Twitter, Facebook use this data and visualize the notable relationships, thus giving businesses the ability to make better decisions that can actually benefit both companies as well as customers
- **Business intelligence**: Machine learning characteristics, when merged with big data analytics can help companies to find solutions to the problems that can help the businesses to grow and generate more profit

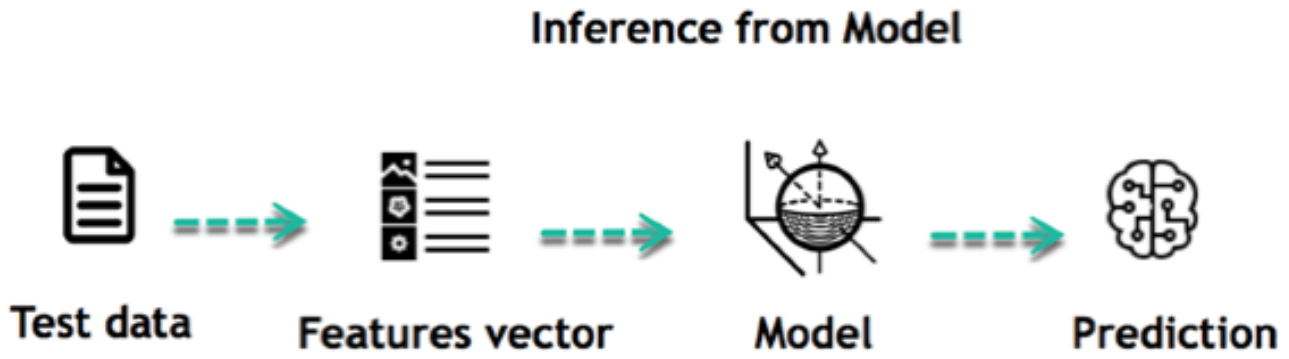
Learning

- The core objective of machine learning is the **learning** and **inference**.
- First of all, the machine learns through the discovery of patterns. This discovery is made thanks to the **data**. One crucial
- part of the data scientist is to choose carefully which data to provide to the machine. The list of attributes used to solve a problem is called a **feature vector.** You can think of a feature vector as a subset of data that is used to tackle a problem.
- The machine uses some fancy algorithms to simplify the reality and transform this discovery into a **model**. Therefore, the learning stage is used to describe the data and summarize it into a model.
- For instance, the machine is trying to understand the relationship between the wage of an individual and the likelihood to go to a fancy restaurant. It turns out the machine finds a positive relationship [illegible] the model



Inferring

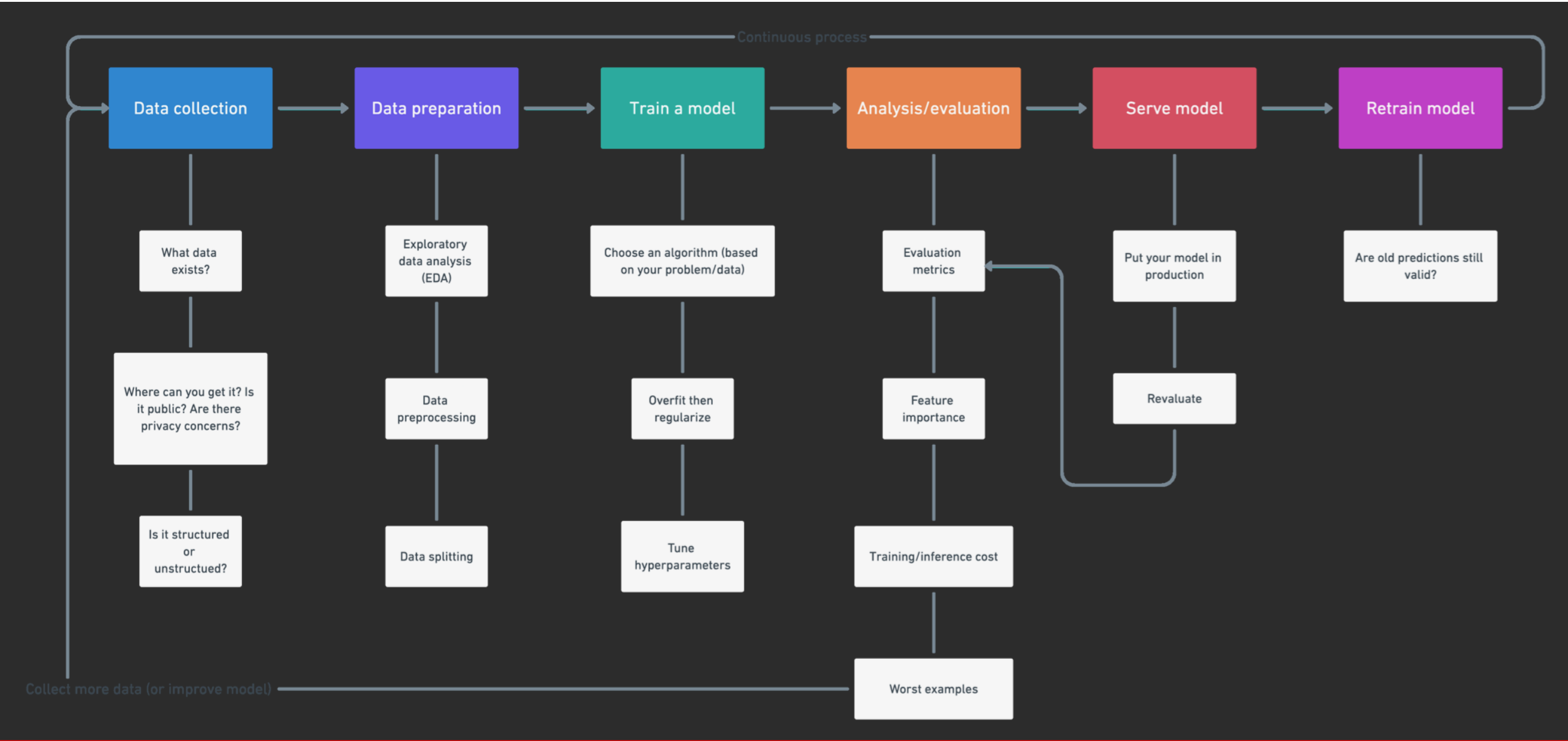
- When the model is built, it is possible to test how powerful it is on never-seen-before data. The new data are transformed into a features vector, go through the model and give a prediction. This is all the beautiful part of machine learning. There is no need to update the rules or train again the model. You can use the model previously trained to make inference on new data.



The life of Machine Learning programs

1. Define a question
2. Collect data
3. Visualize data
4. Train algorithm
5. Test the Algorithm
6. Collect feedback
7. Refine the algorithm
8. Loop 4-7 until the results are satisfying
9. Use the model to make a prediction

Machine Learning Process

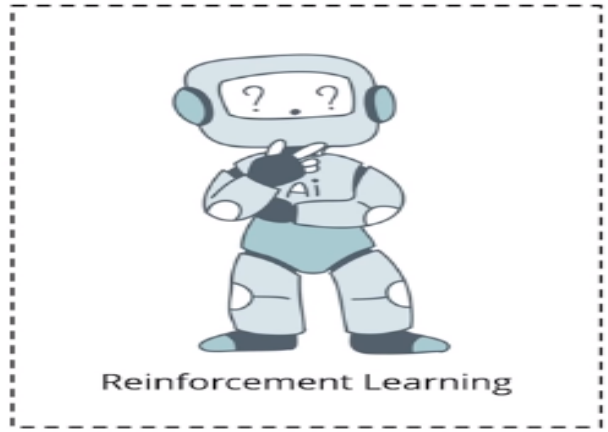


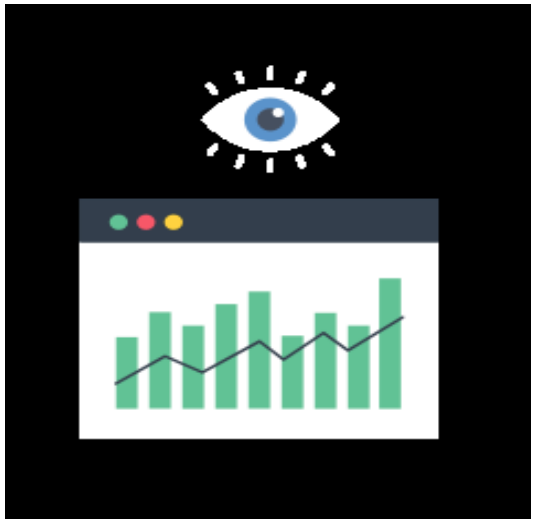
Machine learning algorithms

Machine learning algorithms has been broadly categorized into three categories

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

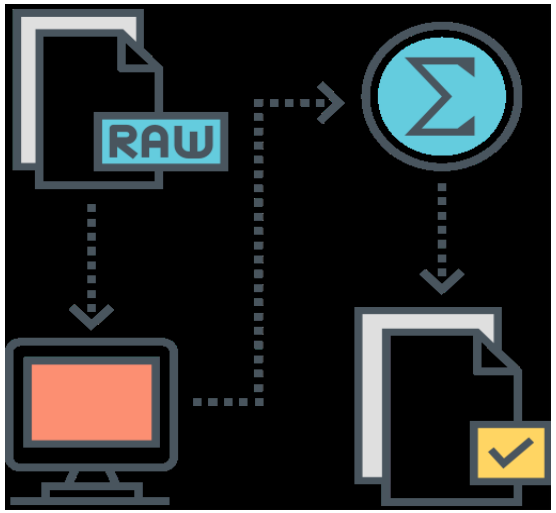
Categories of learning





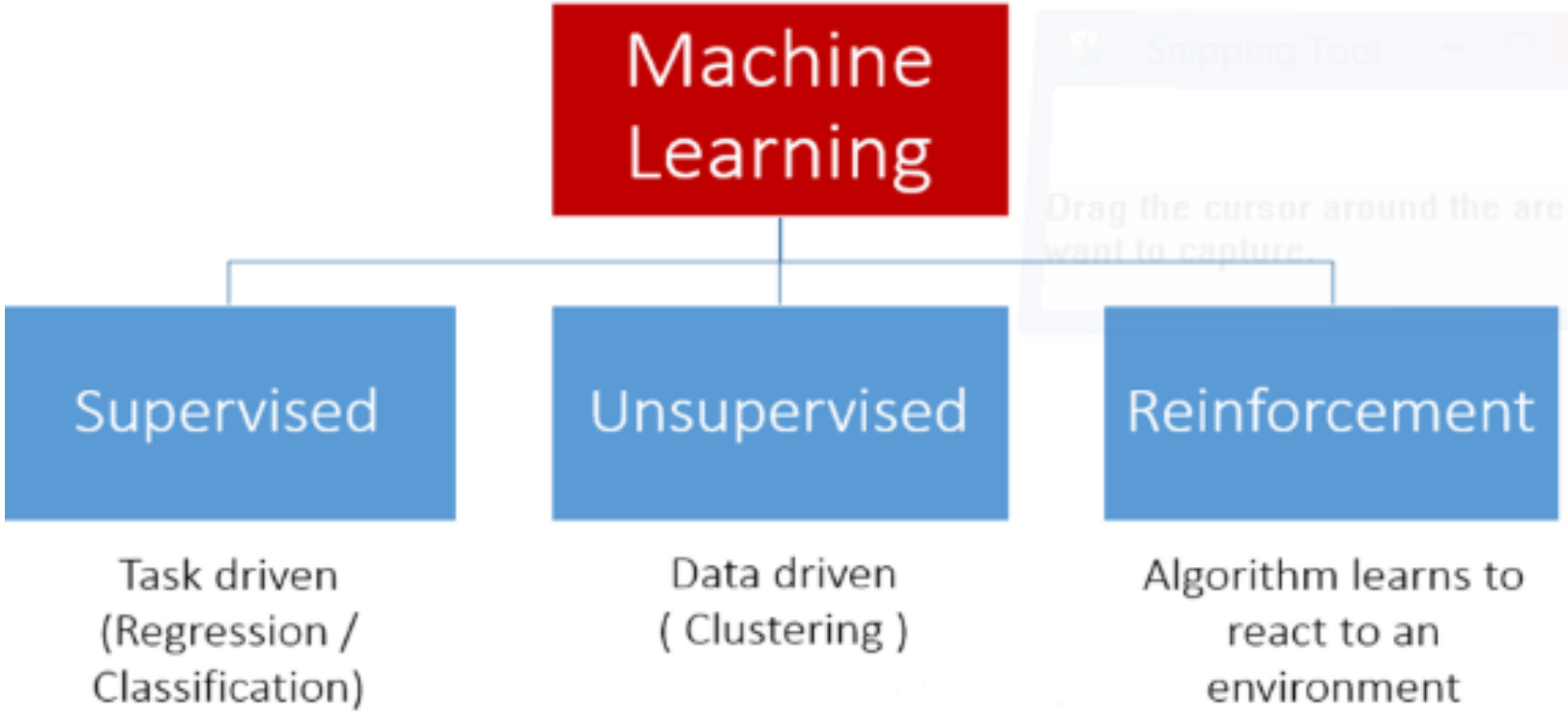
Supervised Learning

Unsupervised Learning

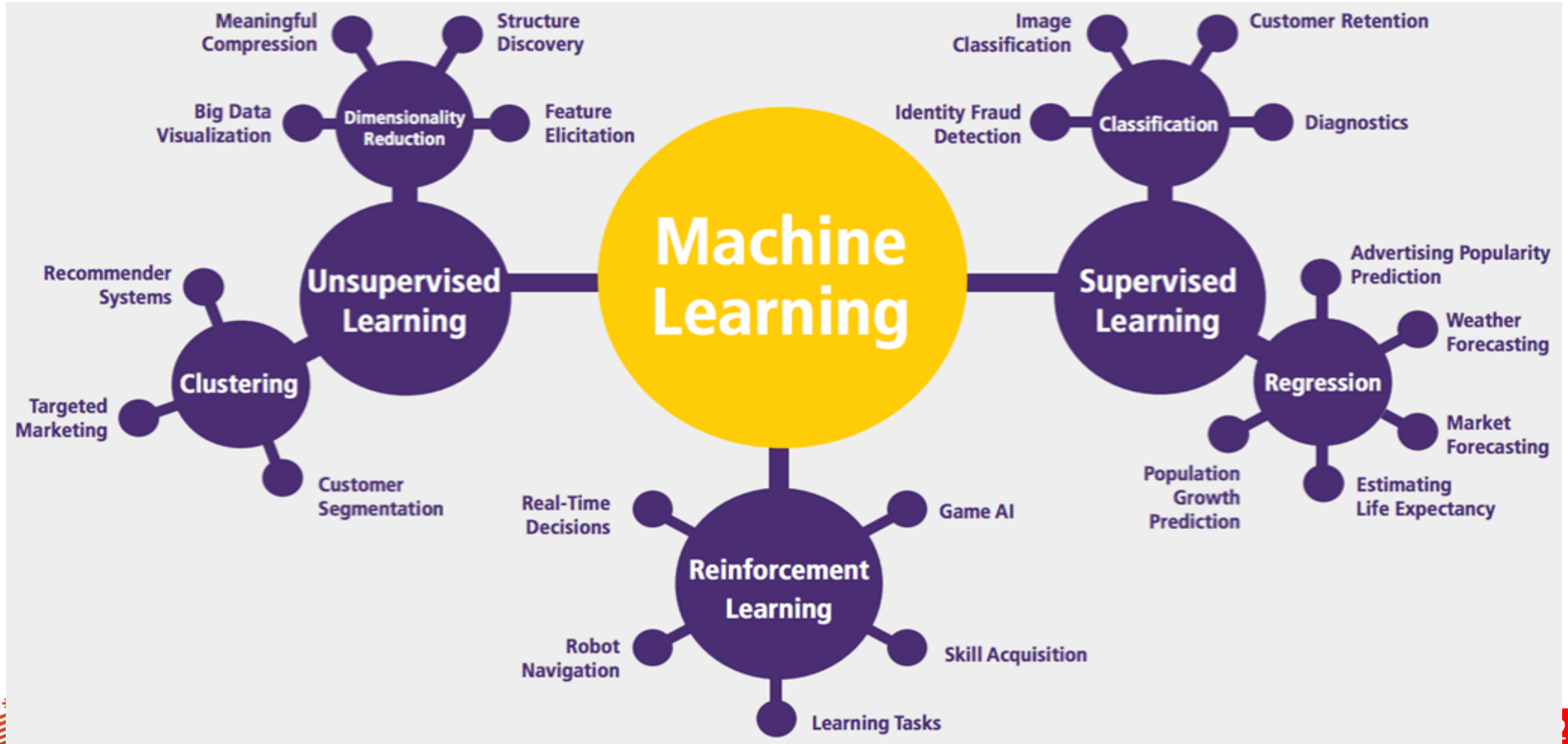


Reinforcement Learning

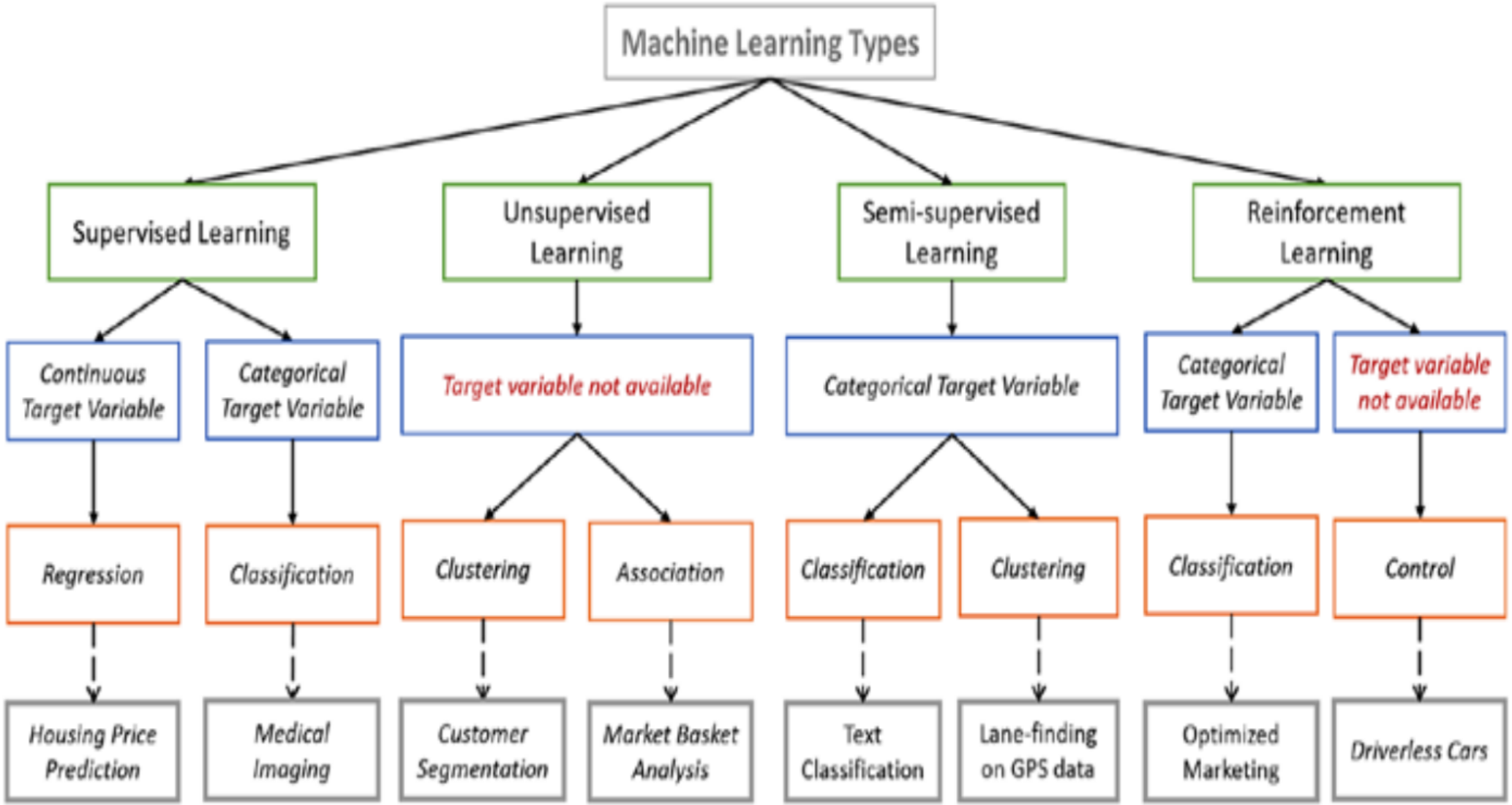
Types of Machine Learning Algorithms



Machine Learning Algorithms



Classification and Clustering Algorithms



Definition

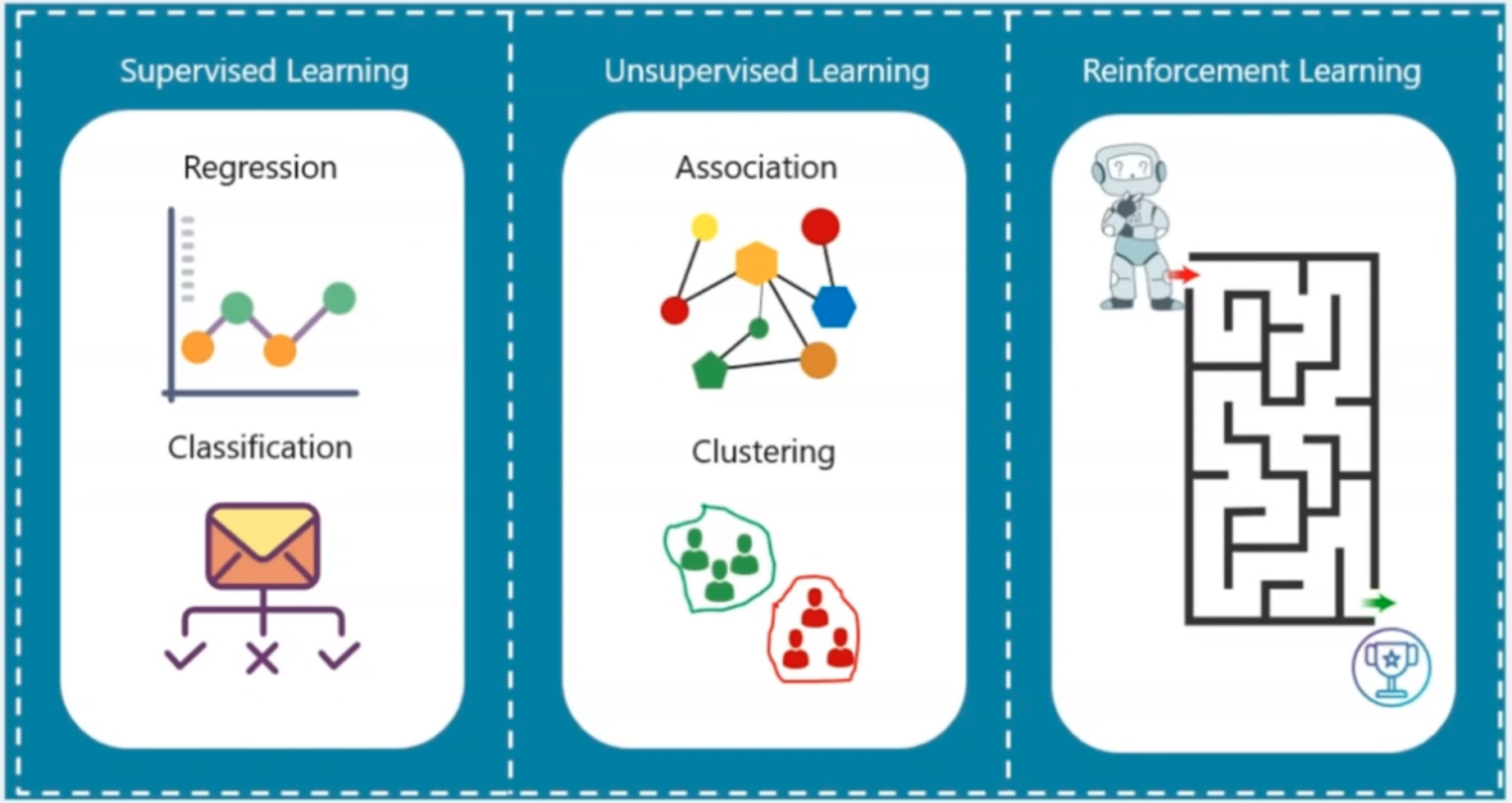
- Definition
- Type of Problems
- Type of data
- Training
- Aim
- Approach
- Output Feedback
- Popular Algorithms
- Applications

Supervised learning is a method in which we teach the machine using labelled data

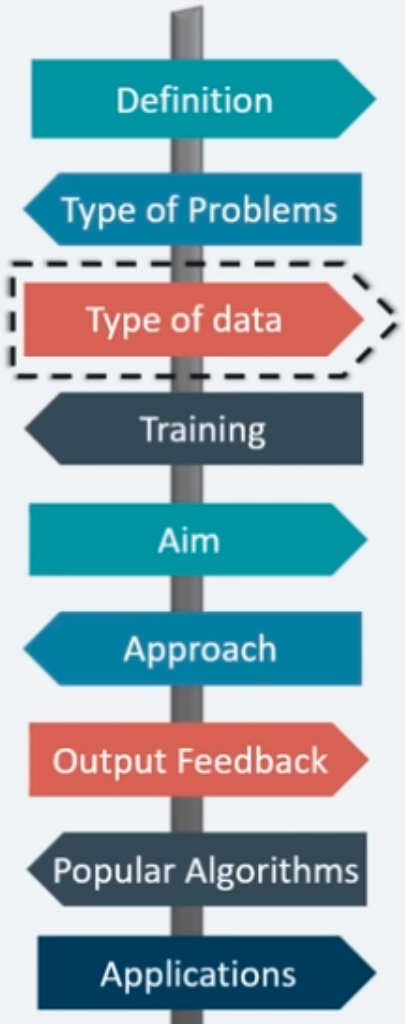
In unsupervised learning the machine is trained on unlabelled data without any guidance

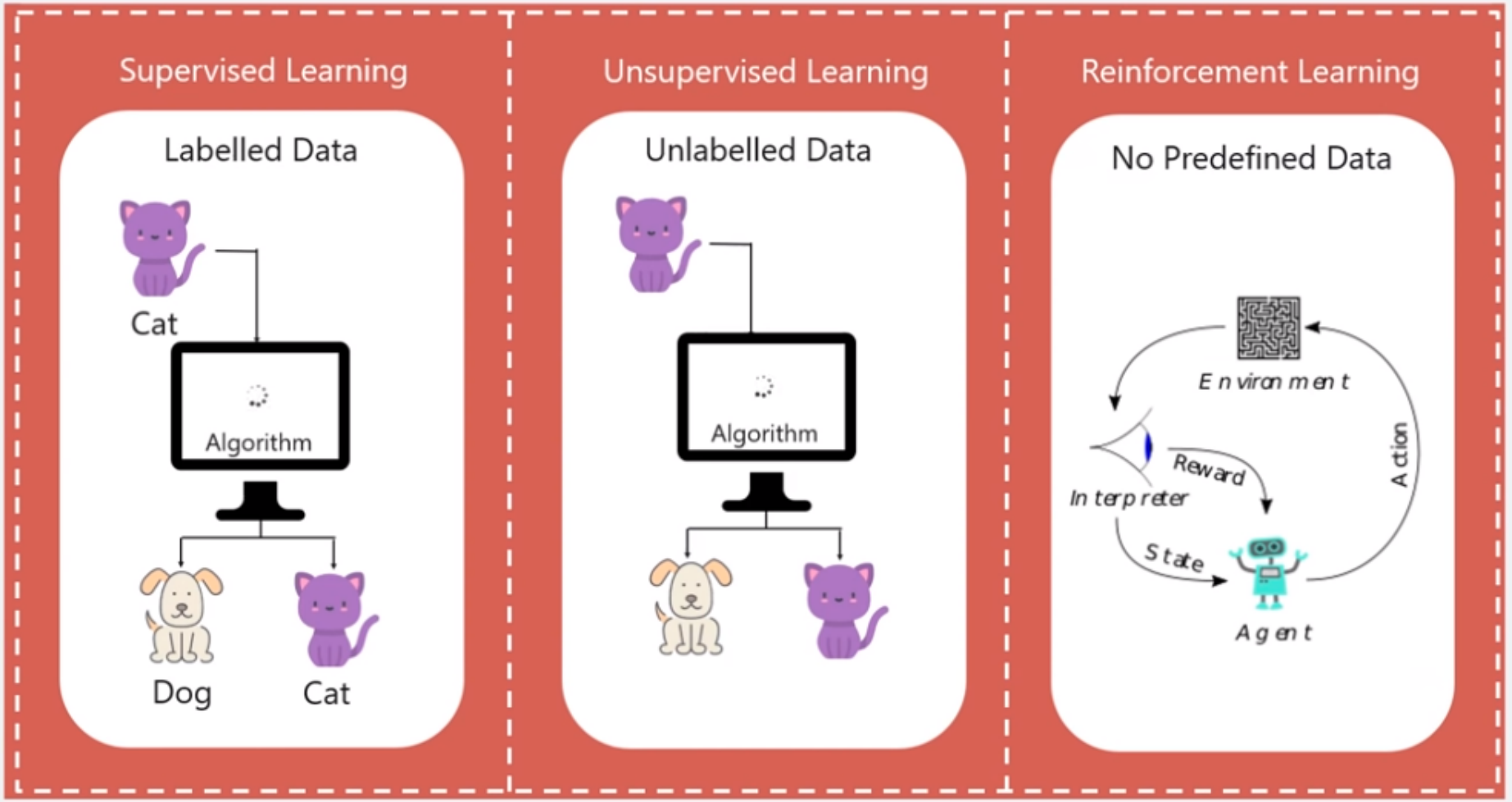
In Reinforcement learning an agent interacts with its environment by producing actions & discovers errors or rewards

Problem Type



Type of data





Training

- Definition
- Type of Problems
- Type of data
- Training
- Aim
- Approach
- Output Feedback
- Popular Algorithms
- Applications

Supervised Learning

External supervision

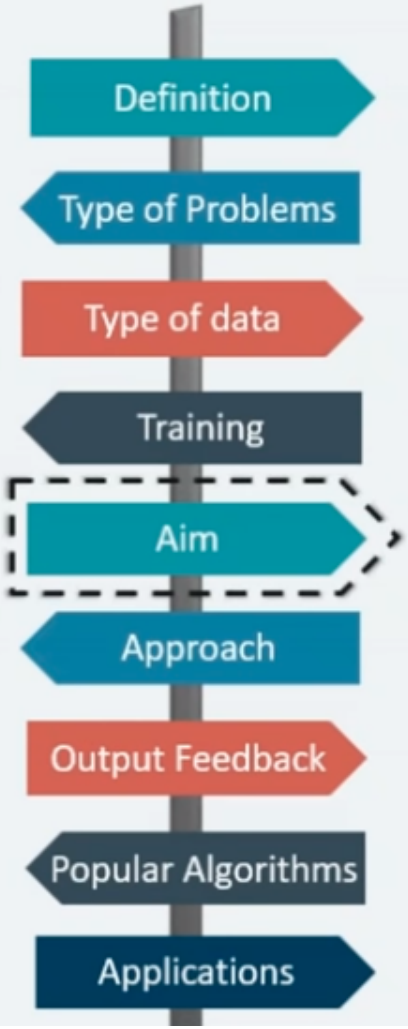
Unsupervised Learning

No supervision

Reinforcement Learning

No supervision

Aim



Supervised Learning

Forecast outcomes

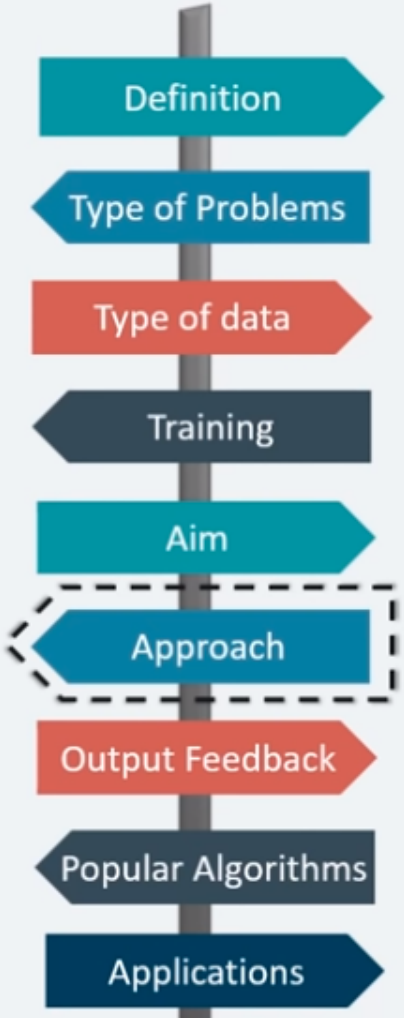
Unsupervised Learning

Discover underlying patterns

Reinforcement Learning

Learn series of action

Approach



Supervised Learning

Map labelled input to known output

Labelled Input

Training

Algorithm

Known Output

Unsupervised Learning

Understand patterns and discover output

Unlabelled Input

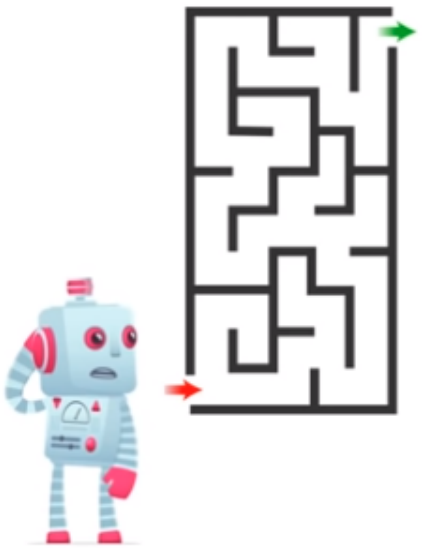
Explore patterns & trends

Algorithm

Output

Reinforcement Learning

Follow Trail and Error method



Popular Algorithms

Supervised Learning	Unsupervised Learning	Reinforcement Learning
Linear Regression	K- Means	Q- Learning
Logistic Regression	Apriori	SARSA
Support Vector Machine	C- Means	
K Nearest Neighbour		
Random Forest		

Applications

- Definition
- Type of Problems
- Type of data
- Training
- Aim
- Approach
- Output Feedback
- Popular Algorithms
- Applications

Supervised Learning

Risk Evaluation

Forecast Sales

Unsupervised Learning

Recommendation Systems

Anomaly Detection

Reinforcement Learning

Self driving cars

Gaming

Supervised Learning

- An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output.
- For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans.
- *Supervised learning is used when the output data is known*. The algorithm will predict new data.
- There are two categories of supervised learning:

1. **Regression problems** – Used to predict future values and the model is trained with the historical data.
    1. E.g., Predicting the future price of a house.
2. **Classification problems** – Various labels train the algorithm to identify items within a specific category.
    1. E.g., Dog or cat(as mentioned in the above example), Apple or an orange, Beer or wine or water.

Supervised Learning

- In a **regression** problem, we are trying to predict results within a continuous output, meaning that we are trying to map input variables to some continuous function.
- In a **classification** problem, we are instead trying to predict results in a discrete output. In other words, we are trying to map input variables into discrete categories.
- **Example 1:**
  - Given data about the size of houses on the real estate market, try to predict their price. Price as a function of size is a continuous output, so this is a regression problem.
  - We could turn this example into a classification problem by instead making our output about whether the house "sells for more or less than the asking price." Here we are classifying the houses based on price into two discrete categories.
  - Regression - Given a picture of a person, we have to predict their age on the basis of the given picture
  - Classification - Given a patient with a tumor, we have to predict whether the tumor is malignant or benign.

Supervised Learning

- In Supervised Learning, we have two sets of variables. One is called the target variable, or labels (the variable we want to predict) and features(variables that help us to predict target variables).
- We show the program(model) the features and the label associated with these features and then the program is able to find the underlying pattern in the data.
- Take this example of the dataset where we want to predict the price of the house given its size. The price which is a target variable depends upon the size which is a feature.

Number of rooms	Price
1	$100
3	$300
5	$500

- we can say that the supervised learning model has a set of input variables (x), and an output variable (y).
- An algorithm identifies the mapping function between the input and output variables. The relationship is **y = f(x).**

Classification and Regression Examples

- Classification : Imagine you want to predict the gender of a customer for a commercial.
  - Gathering data on the height, weight, job, salary, purchasing basket, etc. from your customer database. You know the gender of each of your customer, it can only be male or female.
  - The objective of the classifier will be to assign a probability of being a male or a female (i.e., the label) based on the information (i.e., features you have collected).
- When the model learned how to recognize male or female, you can use new data to make a prediction.
- **Regression** : When the output is a continuous value, the task is a regression.
  - For instance, a financial analyst may need to forecast the value of a stock based on a range of feature like equity, previous stock performances, macroeconomics index. The system will be trained to estimate the price of the stocks with the lowest possible error.

Supervised Learning

Algorithm Name	Description	Type
Linear regression	Finds a way to correlate each feature to the output to help predict future values.	Regression
Logistic regression	Extension of linear regression that's used for classific ation tasks. The output variable 3is binary (e.g., only black or white) rather than continuous (e.g., an infinit e list of potential colors)	Classification
Decision tree	Highly interpretable classification or regression mode l that splits data-feature values into branches at decisi on nodes (e.g., if a feature is a color, each possible co lor becomes a new branch) until a final decision outp ut is made	Regression Classificati on
Naive Bayes	The Bayesian method is a classification method that makes use of the Bayesian theorem. The theorem upd ates the prior knowledge of an event with the indepen dent probability of each feature that can affect the ev ent.	Regression Classificati on
Support vector mach ine	Support Vector Machine, or SVM, is typically used f or the classification task. SVM algorithm finds a hyp	Regression (not very co mmon) Classification

Supervised Learning

Algorithm Name	Description	Type
	erplane that optimally divided the classes. It is best u sed with a non-linear solver.	
Random forest	The algorithm is built upon a decision tree to improv e the accuracy drastically. Random forest generates many times simple decision trees and uses the 'majori ty vote' method to decide on which label to return. Fo r the classification task, the final prediction will be th e one with the most vote; while for the regression tas k, the average prediction of all the trees is the final pr ediction.	Regression Classificati on
AdaBoost	Classification or regression technique that uses a mul titude of models to come up with a decision but weig hs them based on their accuracy in predicting the out come	Regression Classificati on
Gradient-boosting tr ees	Gradient-boosting trees is a state-of-the-art classificat ion/regression technique. It is focusing on the error c ommitted by the previous trees and tries to correct it.	Regression Classificati on

Unsupervised Learning

- This approach is the one where we have no target variables, and we have only the input variable(features) at hand. The algorithm learns by itself and discovers an impressive structure in the data.
- The goal is to decode the underlying distribution in the data to gain more knowledge about the data. e.g., explores customer demographic data to identify patterns)
- You can use it when you do not know how to classify the data, and you want the algorithm to find patterns and classify the data for you
- We can group the unsupervised learning problems as:

1. **Clustering:** This means bundling the input variables with the same characteristics together. E.g., grouping users based on search history
2. **Association:** Here, we discover the rules that govern meaningful associations among the data set. E.g., People who watch ‘X’ will also watch ‘Y’.

Unsupervised Learning

- Unsupervised learning allows us to approach the problems with little or no idea what our results should look like. We can derive structure from data where we don't necessarily know the effect of the variables.
- We can derive this structure by clustering the data based on relationships among the variables in the data.
- With unsupervised learning there is no feedback based on the prediction results.
- **Example:**
- Clustering: Take a collection of 100,000 different genes, and find a way to automatically group these genes into groups that are somehow similar or related by different variables, such as lifespan, location, roles, and so on.

Unsupervised Learning

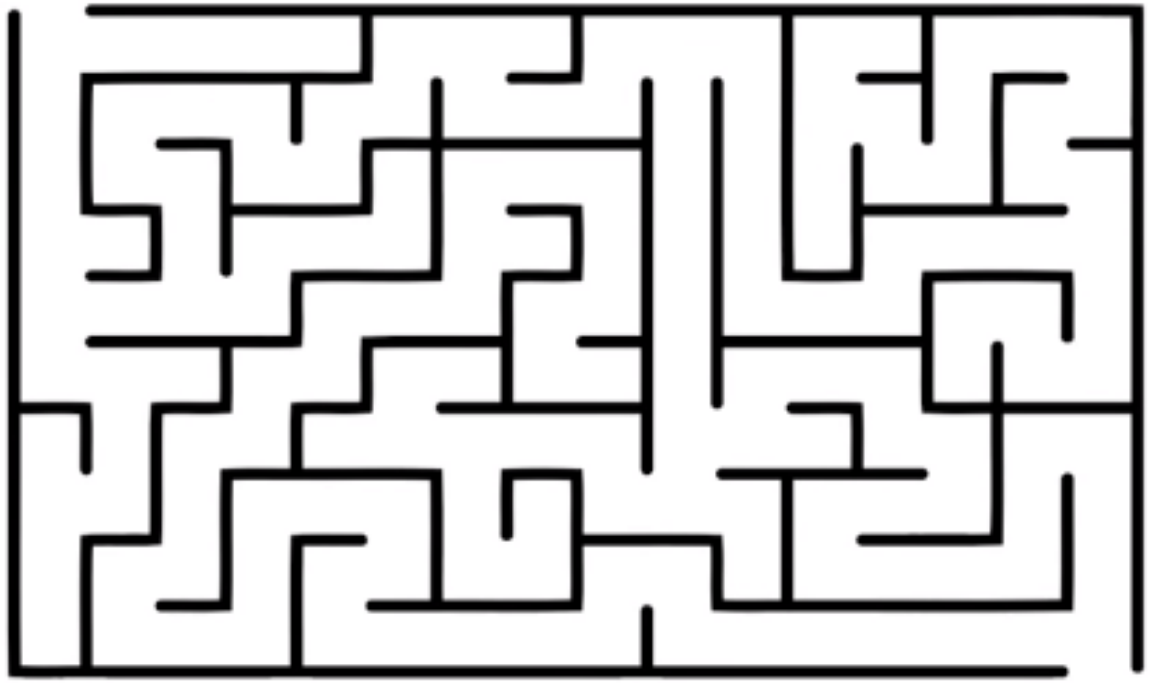
Algorithm	Description	Type
K-means clustering	Puts data into some groups (k) that each contains data with similar characteristics (as determined by the model, not in advance by humans)	Clustering
Gaussian mixture model	A generalization of k-means clustering that provides more flexibility in the size and shape of groups (clusters)	Clustering
Hierarchical clustering	Splits clusters along a hierarchical tree to form a classification system. Can be used for Cluster loyalty-card customer	Clustering
Recommender system	Help to define the relevant data for making a recommendation.	Clustering
PCA/T-SNE	Mostly used to decrease the dimensionality of the data. The algorithms reduce the number of features to 3 or 4 vectors with the highest variances.	Dimension Reduction

Reinforcement Learning

- machine learning models are trained to make a series of decisions based on the rewards and feedback they receive for their actions. The machine learns to achieve a goal in complex and uncertain situations and is rewarded each time it achieves it during the learning period.
- Reinforcement learning is different from supervised learning in the sense that if there is no answer available, so the reinforcement agent decides the steps to perform a task. The machine learns from its own experiences when there is no training data set present.

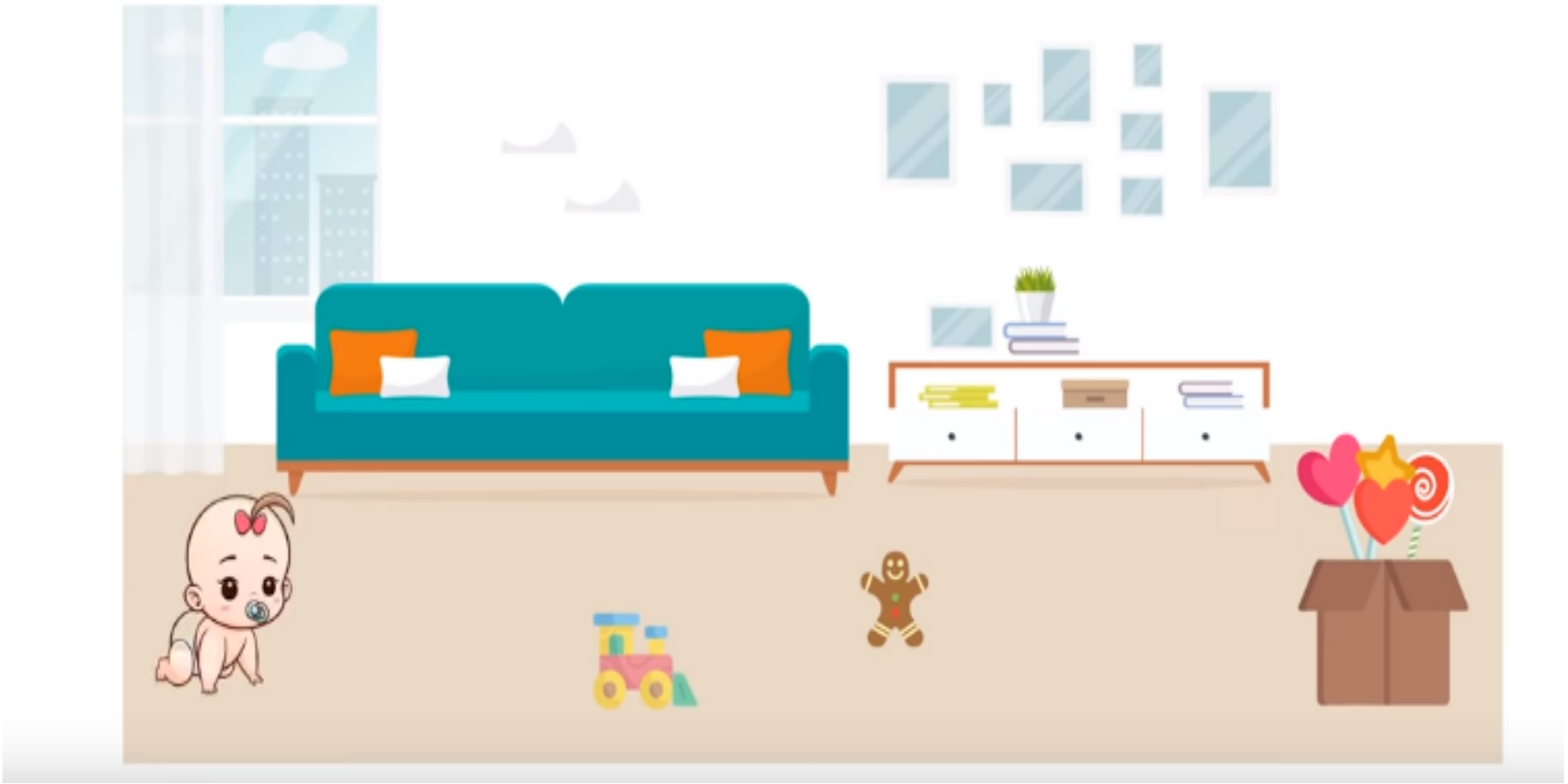
What is Reinforcement Learning(RL)?

Reinforcement learning is a type of Machine Learning where an agent learns to behave in a environment by performing actions and seeing the results



Reinforcement Learning analogy

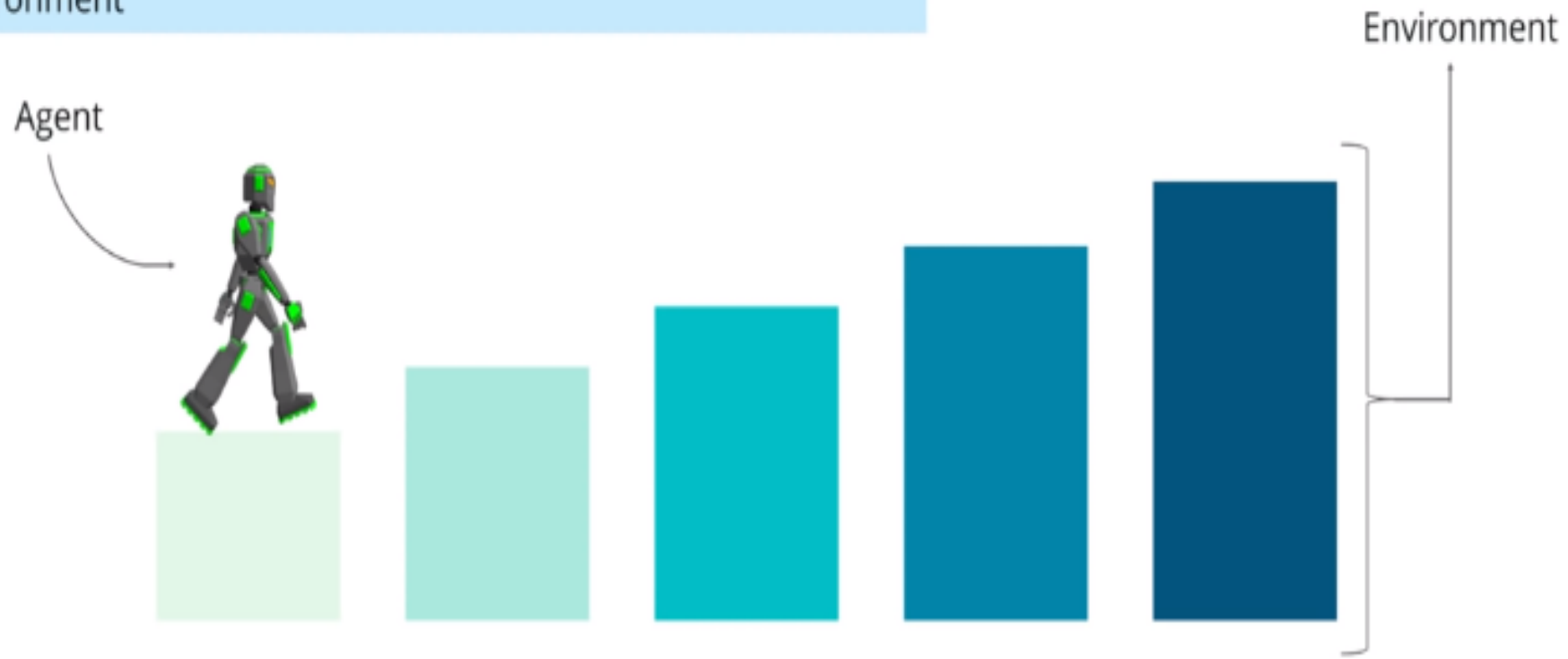
Scenario 2: Baby starts crawling but falls due to some hurdle in between



Reinforcement Learning process

Reinforcement Learning system is comprised of two main components:

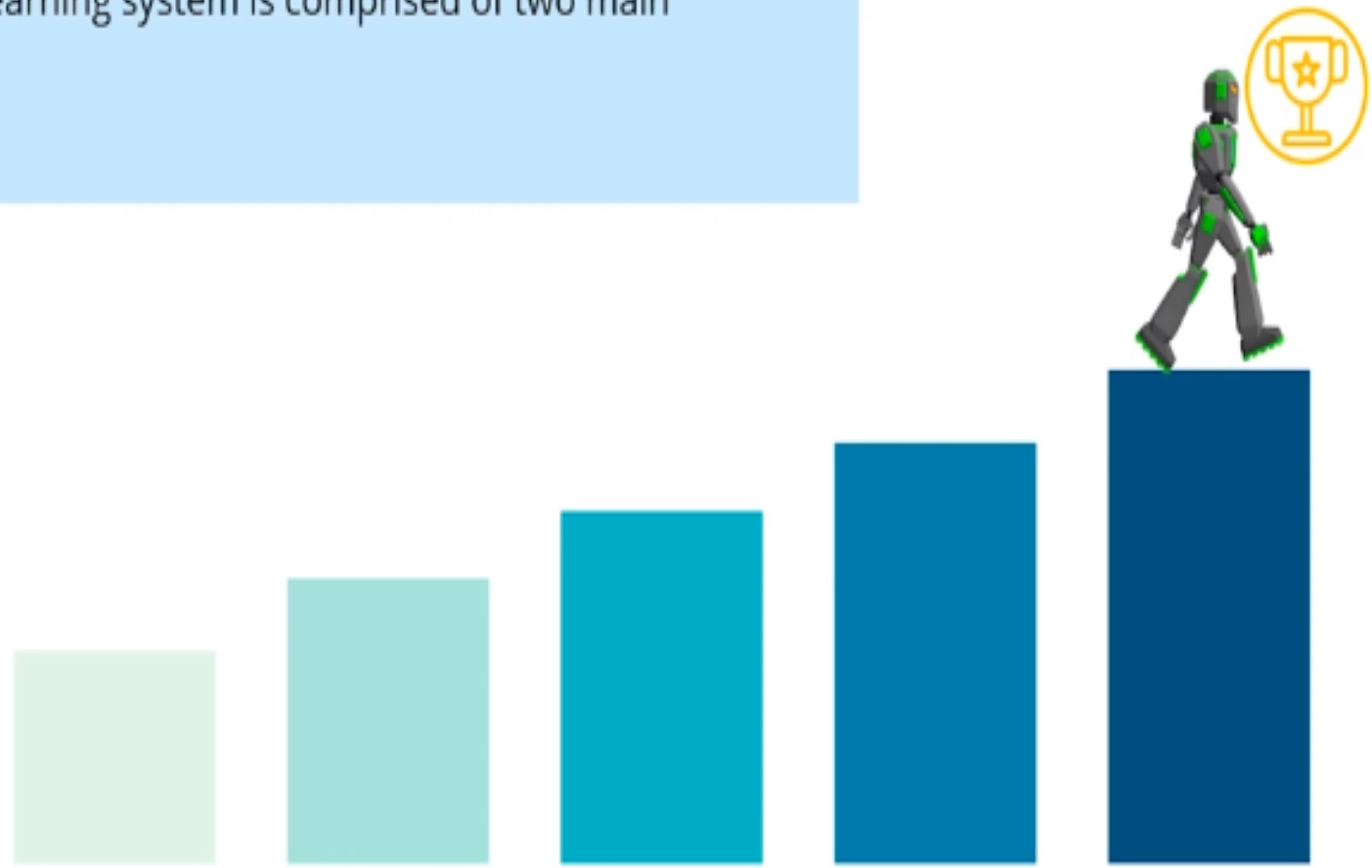
- Agent
- Environment



Reinforcement Learning process

Reinforcement Learning system is comprised of two main components:

- Agent
- Environment



Reinforcement Learning counter strike example

1. The RL Agent (Player1) collects state S^0 from the environment
2. Based on the state S^0, the RL agent takes an action A^0, initially the action is random
3. The environment is now in a new state S^1
4. RL agent now gets a reward R^1 from the environment
5. The RL loop goes on until the RL agent is dead or reaches the destination

Reinforcement Learning definition's

Agent: The RL algorithm that learns from trial and error

Environment: The world through which the agent moves

Action (A): All the possible steps that the agent can take

State (S): Current condition returned by the environment

Reinforcement Learning definitions

Reward (R): An instant return from the environment to appraise the last action

Policy (π): The approach that the agent uses to determine the next action based on the current state

Value (V): The expected long-term return with discount, as opposed to the short-term reward R

Action-value (Q): This similar to Value, except, it takes an extra parameter, the current action (A)

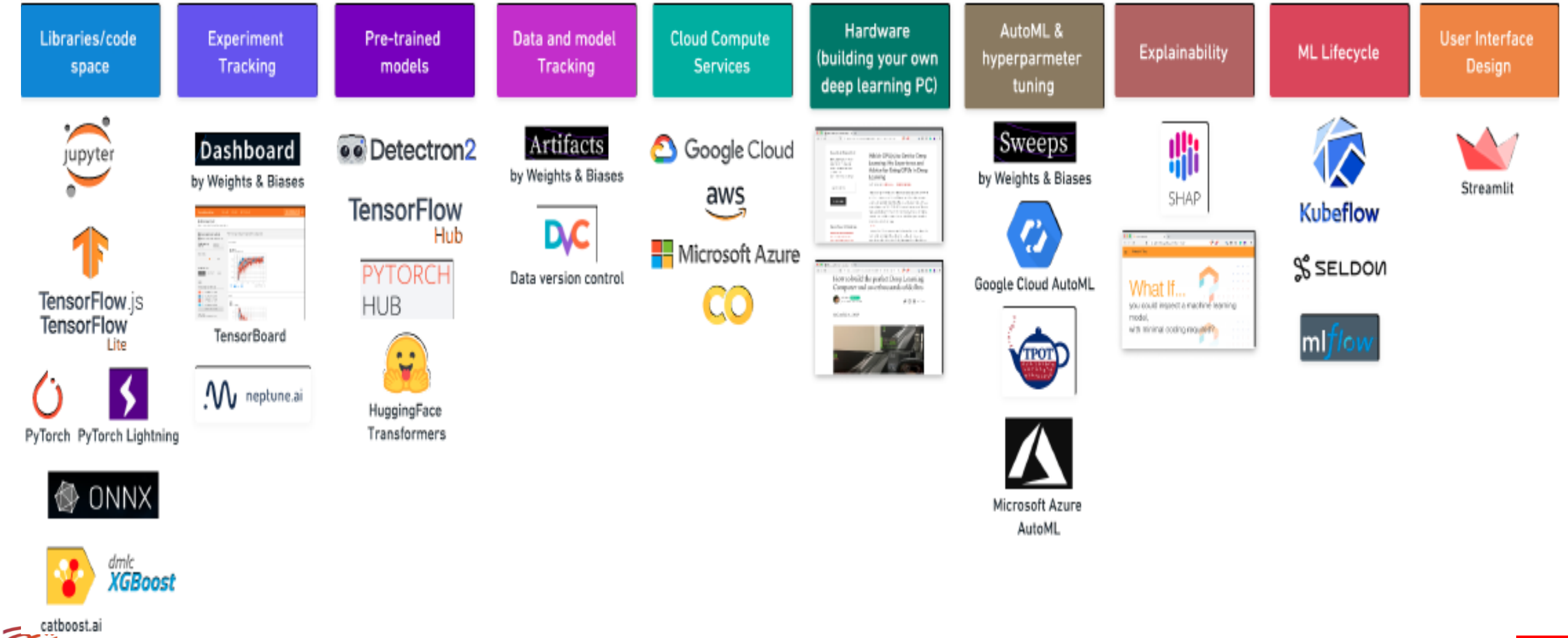
Application of Machine Learning

- **Augmentation**: Machine learning, which assists humans with their day-to-day tasks, personally or commercially without having complete control of the output. Such machine learning is used in different ways such as Virtual Assistant, Data analysis, software solutions. The primary user is to reduce errors due to human bias.
- **Automation**: Machine learning, which works entirely autonomously in any field without the need for any human intervention. For example, robots performing the essential process steps in manufacturing plants.
- **Finance Industry** Machine learning is growing in popularity in the finance industry. Banks are mainly using ML to find patterns inside the data but also to prevent fraud.
- **Government organization** The government makes use of ML to manage public safety and utilities. Take the example of China with the massive face recognition. The government uses Artificial intelligence to prevent jaywalker.
- **Healthcare industry** Healthcare was one of the first industry to use machine learning with image detection.
- **Marketing** Broad use of AI is done in marketing thanks to abundant access to data. Before the age of mass data, researchers develop advanced mathematical tools like Bayesian analysis to estimate the value of a customer. With the boom of data, marketing department relies on AI to optimize the customer relationship and marketing campaign.

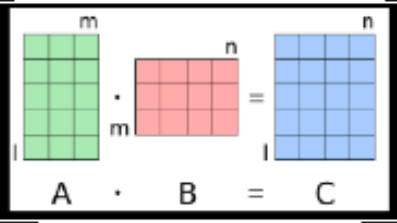
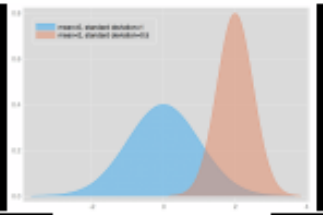
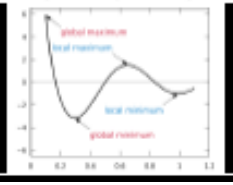
Application of Machine Learning

- Prediction — Machine learning can also be used in the prediction systems. Considering the loan example, to compute the probability of a fault, the system will need to classify the available data in groups.
- Image recognition — Machine learning can be used for face detection in an image as well. There is a separate category for each person in a database of several people.
- Speech Recognition — It is the translation of spoken words into the text. It is used in voice searches and more. Voice user interfaces include voice dialing, call routing, and appliance control. It can also be used a simple data entry and the preparation of structured documents.
- Medical diagnoses — ML is trained to recognize cancerous tissues.
- Financial industry and trading — companies use ML in fraud investigations and credit checks.

Machine Learning Tools



Machine Learning Mathematics

Linear Algebra	$\begin{aligned} 2x + y - z &= 8 \\ -3x - y + 2z &= -11 \\ -2x + y + 2z &= -3 \end{aligned}$
Matrix Manipulation	
Multivariate Calculus	$f(x,y) = \frac{x^2 y}{x^4 + y^2}$
The Chain Rule	$\frac{d}{dx}\left[f\left(g(x)\right)\right] = f'\left(g(x)\right)g'(x)$
Probability + Distributions	
Optimization	

Accuracy, Precision and Recall

- **To measure the performance of the system.**
- When the system correctly classifies a tumor as being malignant, the prediction is called a **true positive**.
- When the system incorrectly classifies a benign tumor as being malignant, the prediction is a **false positive**.
- Similarly, a **false negative** is an incorrect prediction that the tumor is benign, and a **true negative** is a correct prediction that a tumor is benign.
- These four outcomes can be used to calculate several common measures of classification performance, like accuracy, precision, recall and so on.

Accuracy

- **Accuracy is calculated with the following formula –**

$$ACC = {(TP + TN)}/{(TP + TN + FP + FN)}$$

Where,

– TP is the number of true positives

– TN is the number of true negatives

– FP is the number of false positives

– FN is the number of false negatives

Precision and Recall

- **Precision** is the fraction of the tumors that were predicted to be malignant that are actually malignant. Precision is calculated with the following formula –

$$PREC = {}^{TP}/_{(TP + FP)}$$

- **Recall** is the fraction of malignant tumors that the system identified. Recall is calculated with the following formula –

$$R = {}^{TP}/_{(TP + FN)}$$

- **F-measure:** uses Harmonic Mean in place of Arithmetic Mean as it punishes the extreme values more. It will always be nearer to the smaller value of Precision or Recall.

$$F = (2 * R * PREC)/(R + PREC)$$

Confusion Matrix

- A **confusion matrix** is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.
- The **confusion matrix** itself is relatively simple to understand, but the related terminology can be **confusing**.

		Predicted class: *P*	Predicted class: *N*
Actual Class	*P*	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Confusion Matrix

- A **confusion matrix** is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.
- The **confusion matrix** itself is relatively simple to understand, but the related terminology can be **confusing**.

		Predicted class: *P*	Predicted class: *N*
Actual Class	*P*	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Reference

a. Essential Reading

1. Class Notes
2. Tom M. Mitchell, Machine Learning, India Edition 2013, McGraw Hill Education.

b. Recommended Reading

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, h The Elements of Statistical Learning, 2nd edition, springer series in statistics.
2. Ethem Alpaydın, Introduction to machine learning, second edition, MIT press.