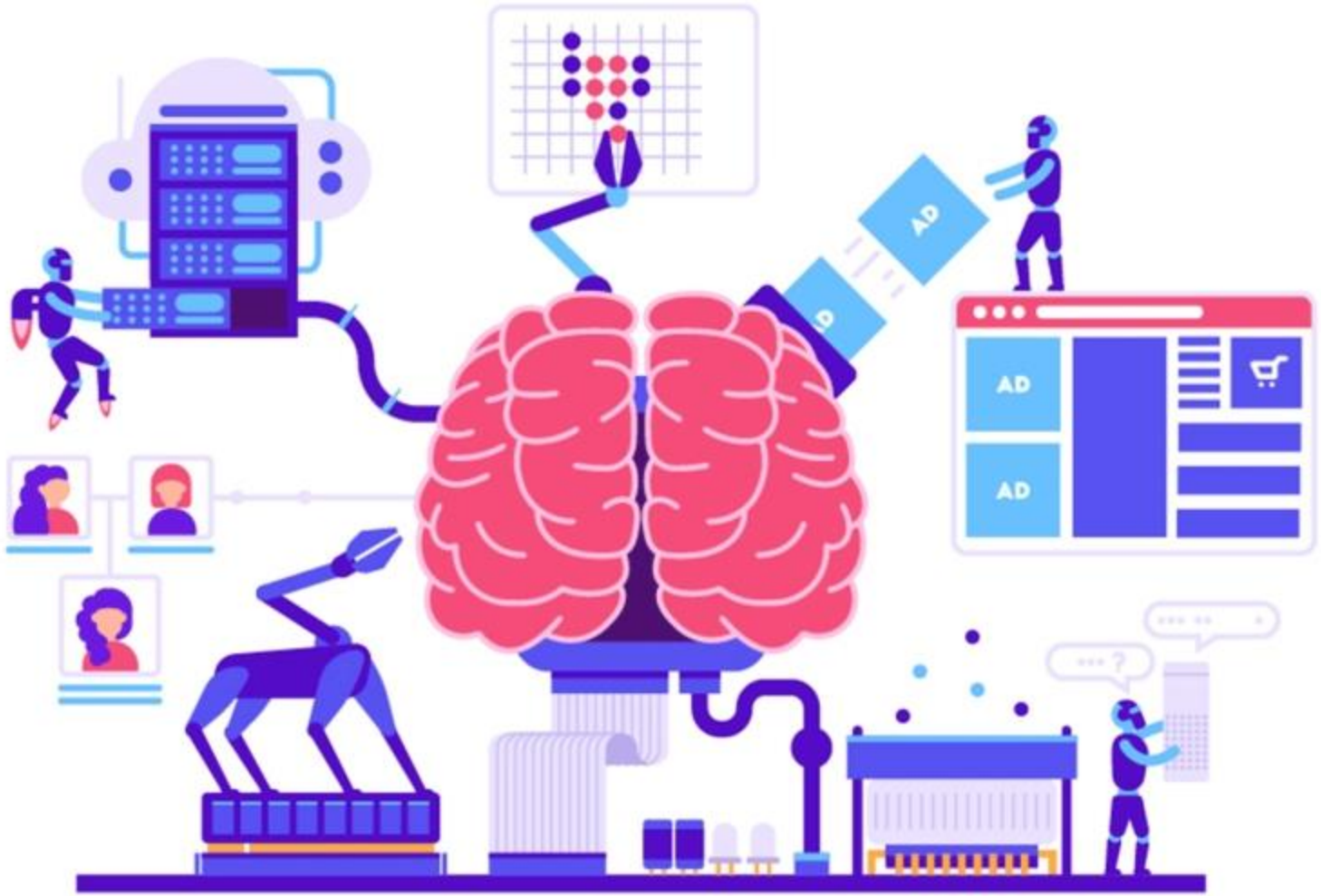
Introduction to Data Science

Machine learning (ML)



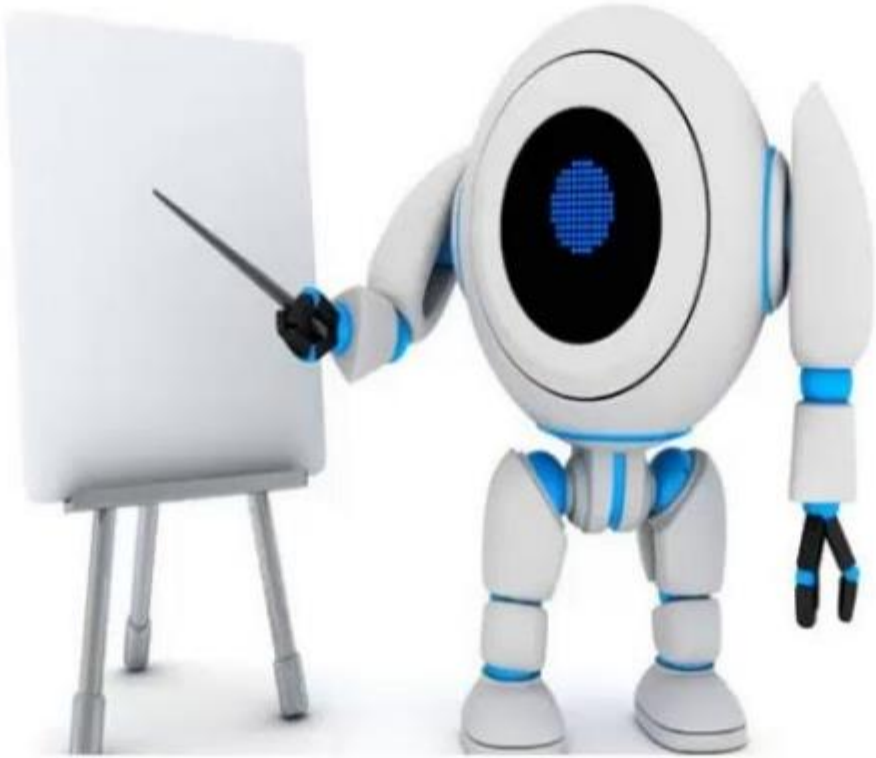
At the end of this session, you should be able to answer the following questions:

1. Define machine learning
2. List down different types of machine learning
3. Explain supervised learning
4. List down the most common algorithms in supervised learning
5. Explain unsupervised learning
6. List down the most common algorithms in unsupervised learning

What is machine learning?

Learn From Experience

Data
Learn From ~~Experience~~



Follow Instructions

What is machine learning?

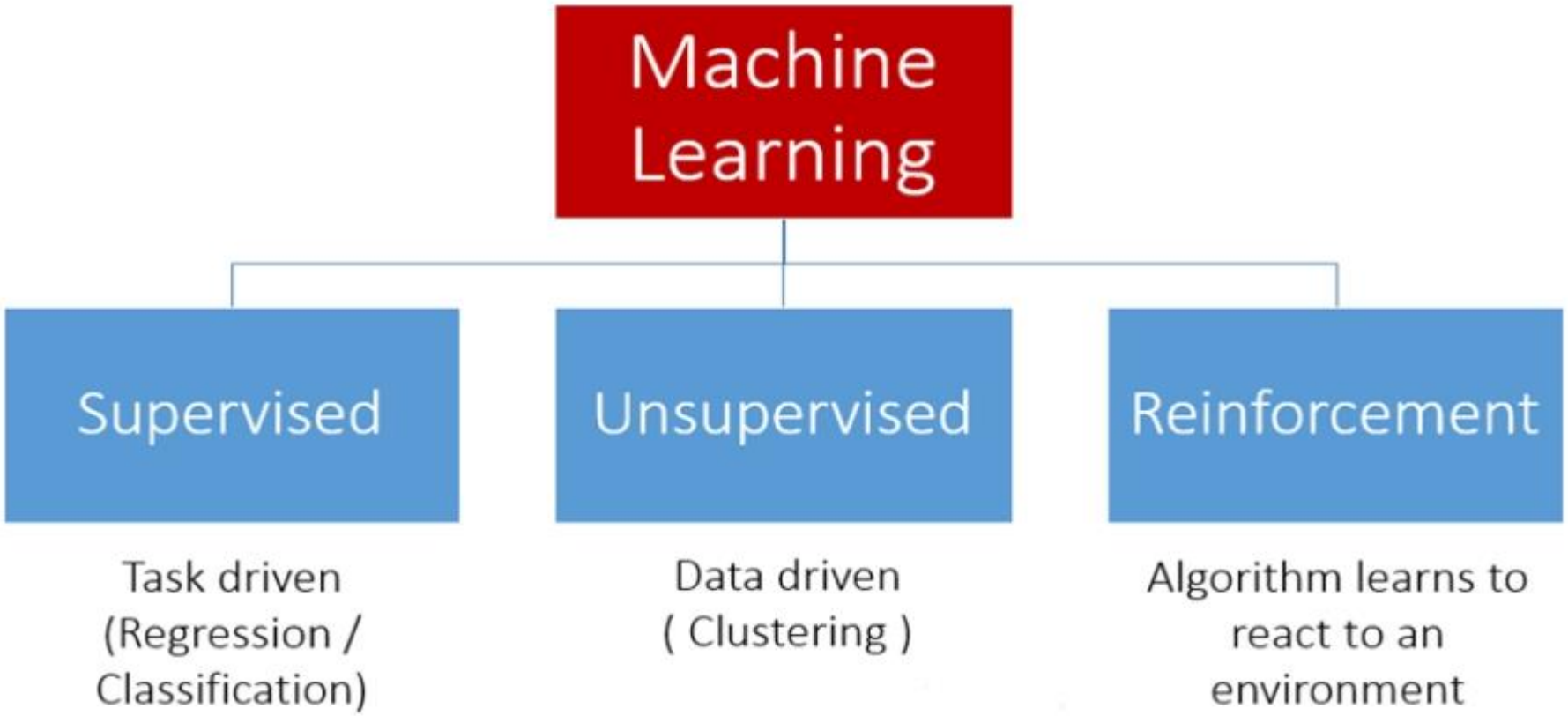
- An application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

- In other words, ability of IT systems to independently find solutions to problems by recognizing patterns in databases.

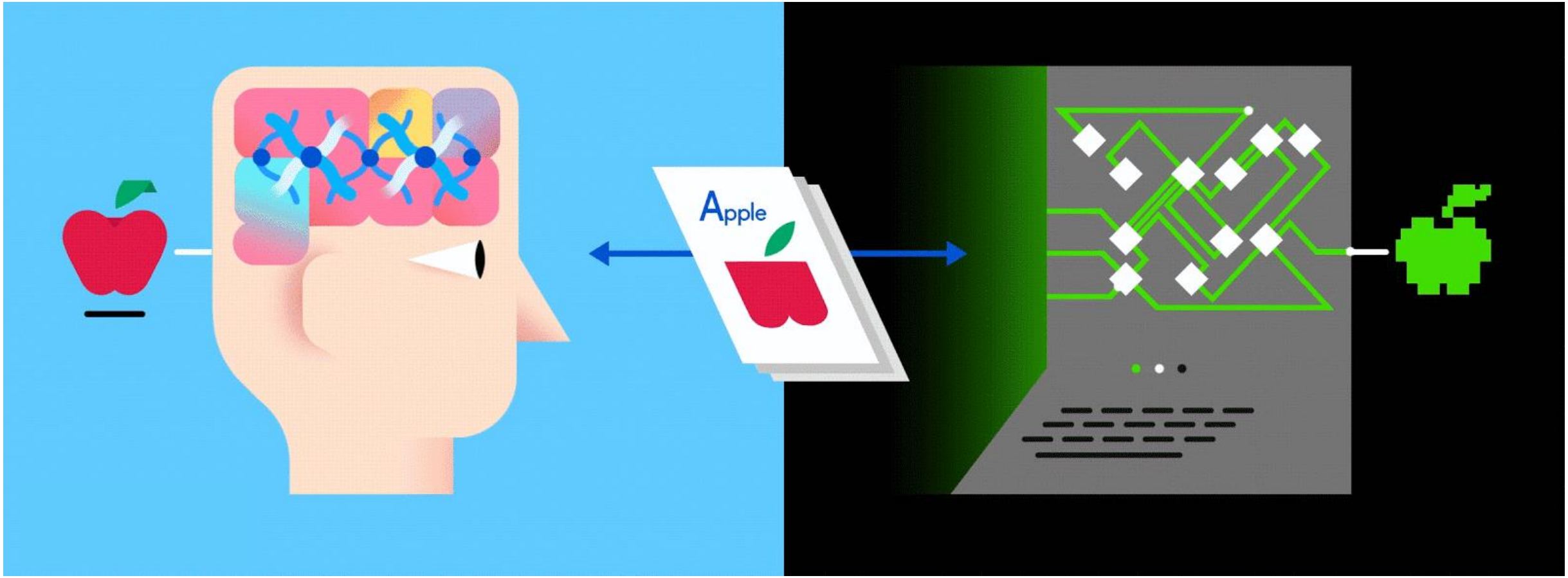
- In order to enable the software to independently generate solutions, we need to:

1) Fit in dataset
2) Select required algorithms
3) Specify pattern of data (ex: select variables, clusters, training and testing data)

Types of machine learning



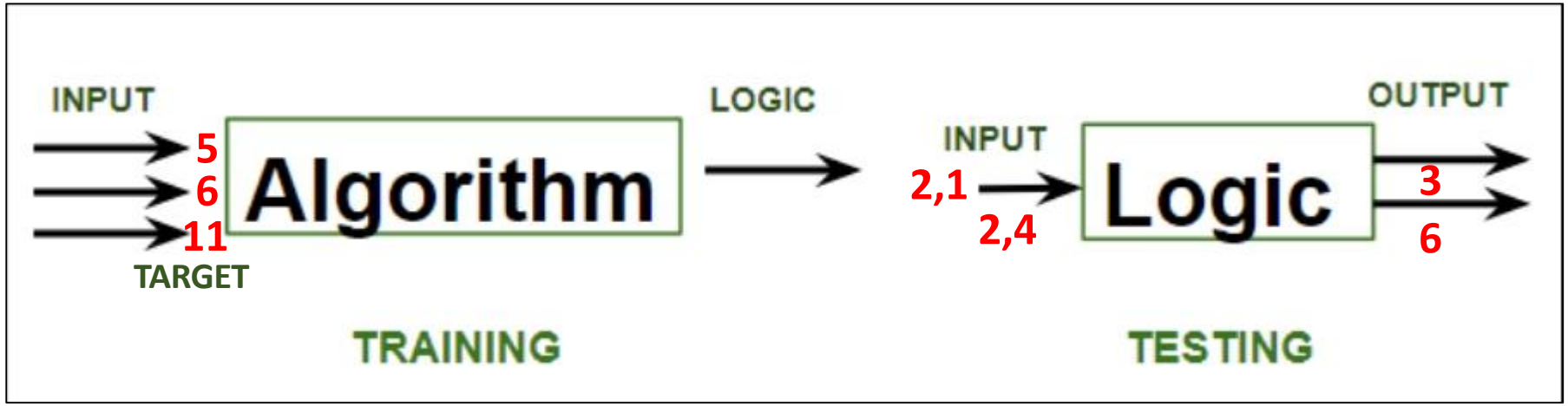
Supervised learning



Supervised learning

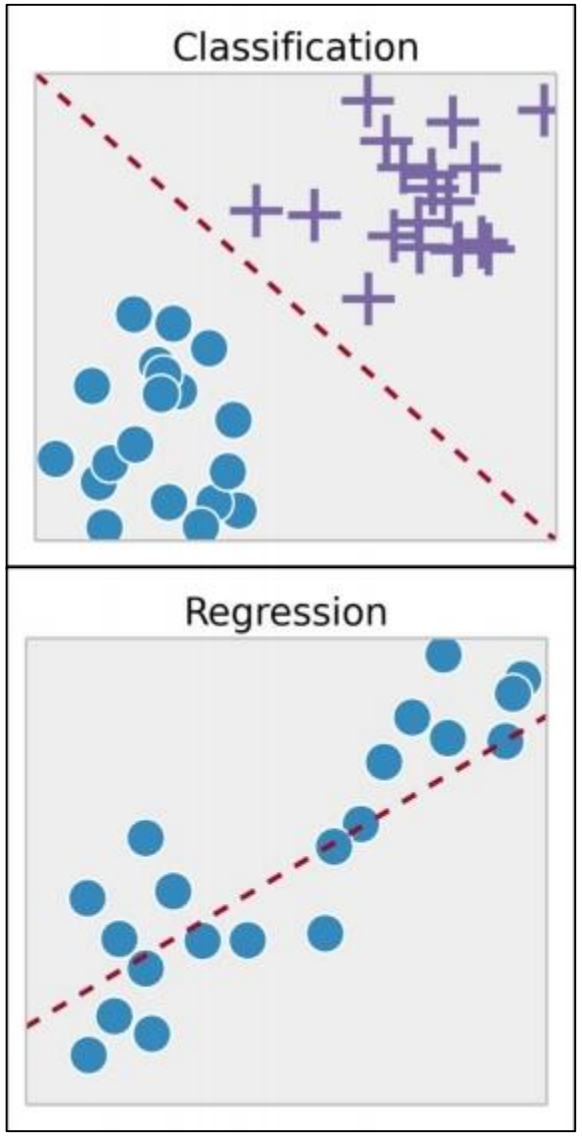
- Provided with inputs and targets (class labels)

- Already know what the correct output should look like, having the idea that there is a relationship between the input and the output.



- Train the model with large training data (inputs & targets)
- Predict the output using new data/ testing data

Types of supervised learning

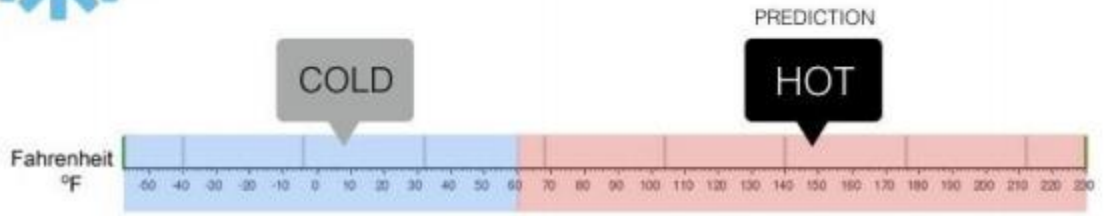


Regression

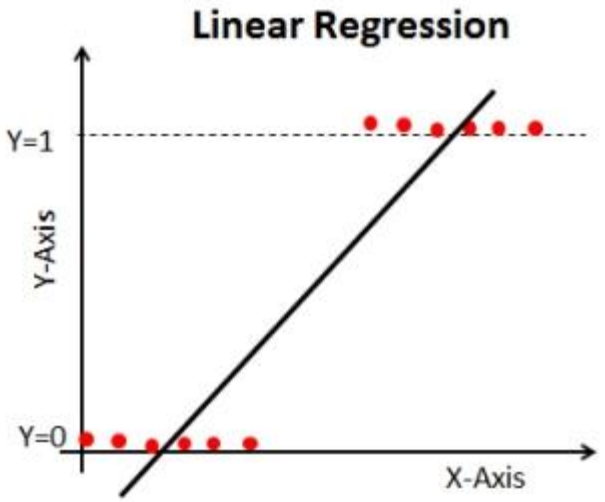
What is the temperature going to be tomorrow?

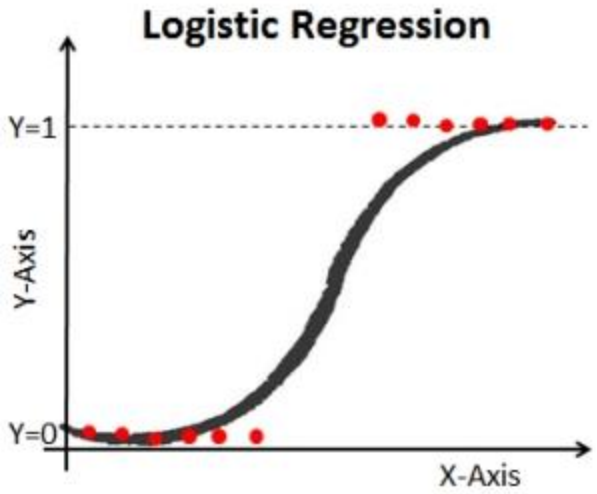
Classification

Will it be Cold or Hot tomorrow?



Regression





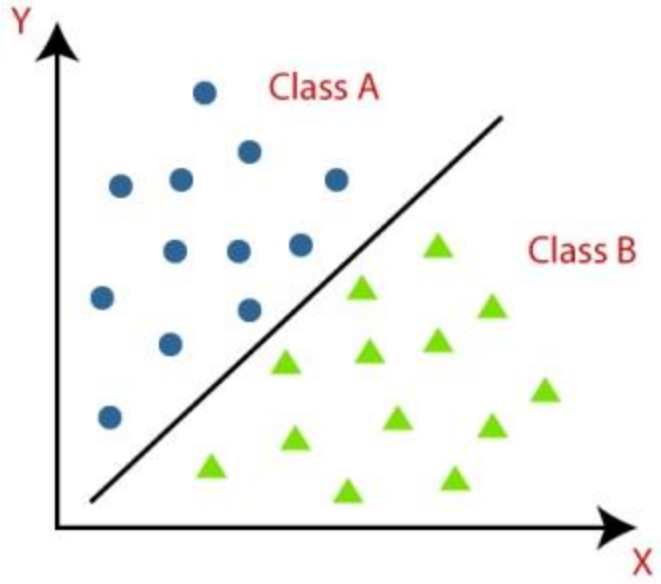
- In regression problem, we are trying to predict results within a continuous output, meaning that we are trying to map input variables to some continuous function.

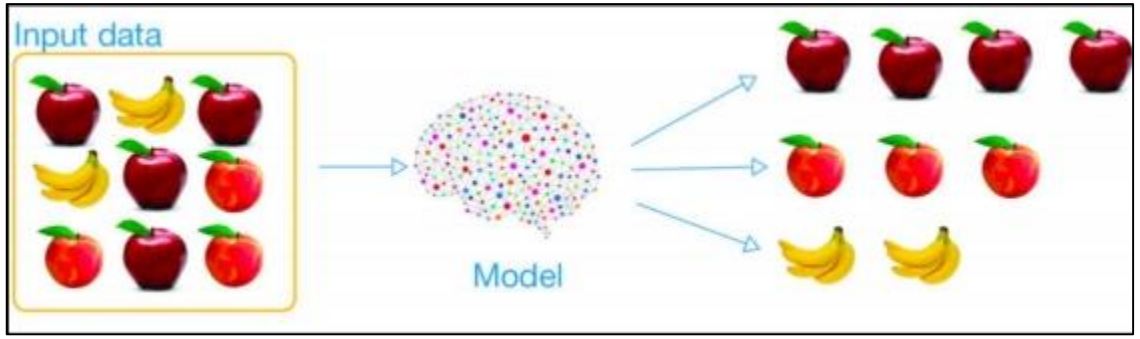
- Example 1 : given data about the size of houses on the real estate market, try to predict their price.

- Example 2: given a picture of a person, we have to predict their age or gender.

Linear Regression	Logistic Regression
Target is an interval variable.	Target is a discrete (binary or ordinal) variable.

Classification



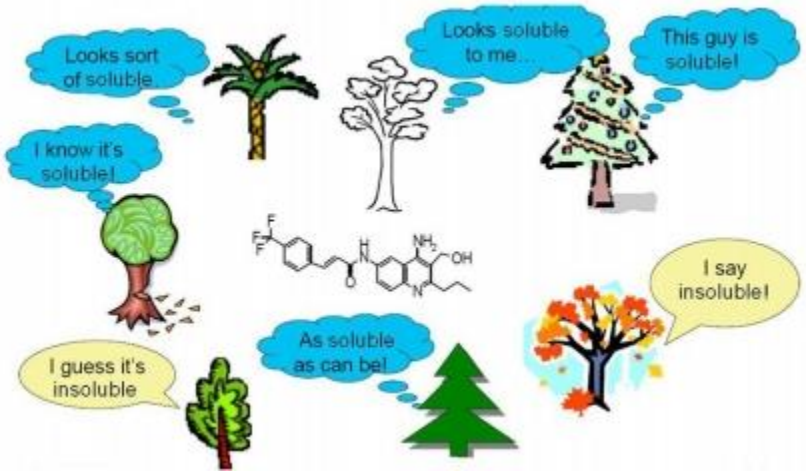


- Finding the category of the input variable, or in more academic terms, mapping input variables into discrete categories.
- like , yes or no, 0 or 1, true or false.
- Example 1: from the example of house price given above, if we change the output to “Sells for more or less than asking price,” then it is a classification problem. (Binary classification)
- Example 2: given a patient with tumour , we have to predict whether the tumour is malignant or benign . (Binary classification)
- Example 3: is this patient in cancer stage 1, 2, 3 or 4? (multi class classification)

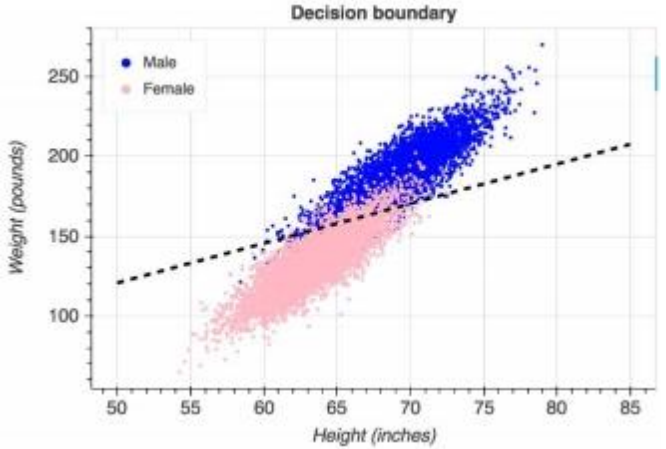
Most common algorithms in supervised learning

Decision tree

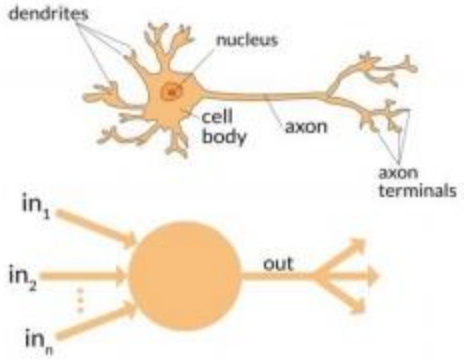
Random forest



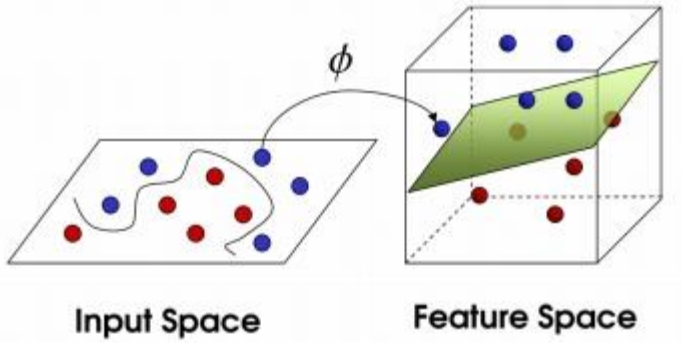
Logistic regression



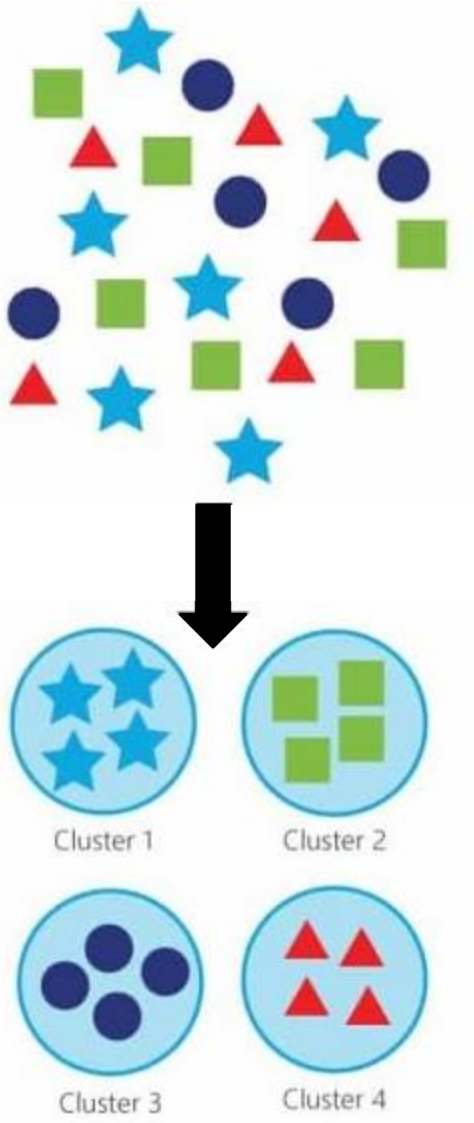
Artificial neural networks



Support vector machine



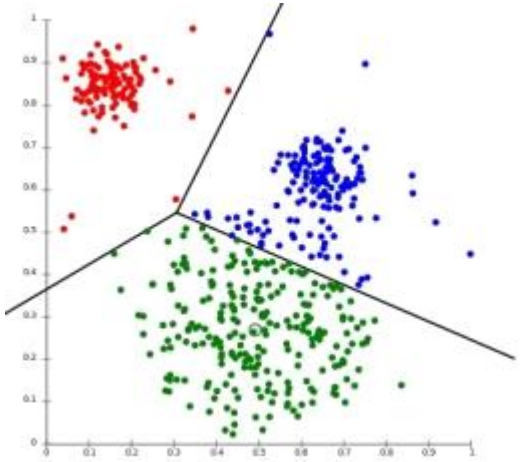
Unsupervised learning/ Clustering



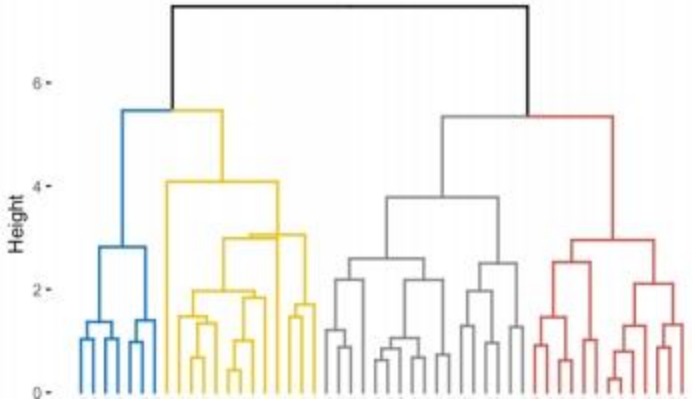
- Provided with inputs only.
- To approach problems with little or no idea what the results should look like.
- Can derive structure from data where we don't necessarily know the effect of the variables.
- Derive this structure by clustering the data based on relationships among the variables in the data. There
- is no feedback based on the prediction results. For
- example, take a collection of 1,000,000 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, functions, and so on.

Algorithms used in unsupervised learning

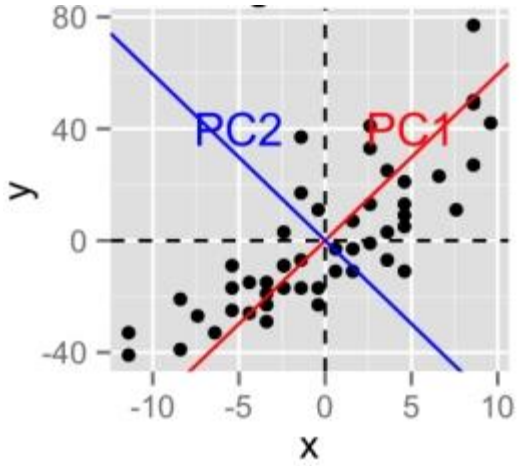
k-means clustering



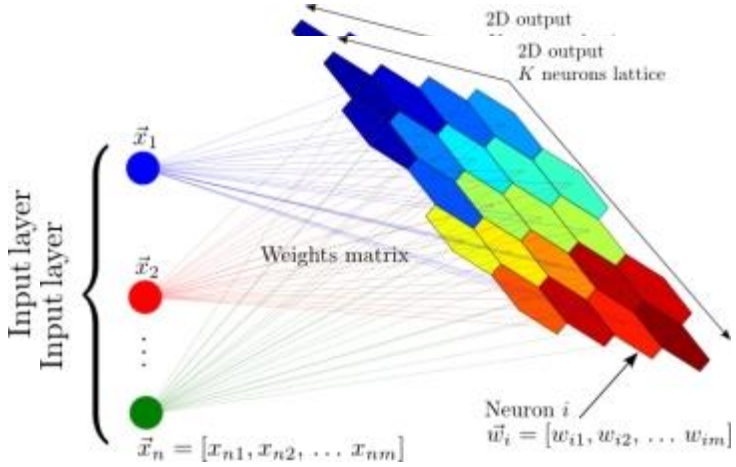
Hierarchical clustering



Principal Component Analysis

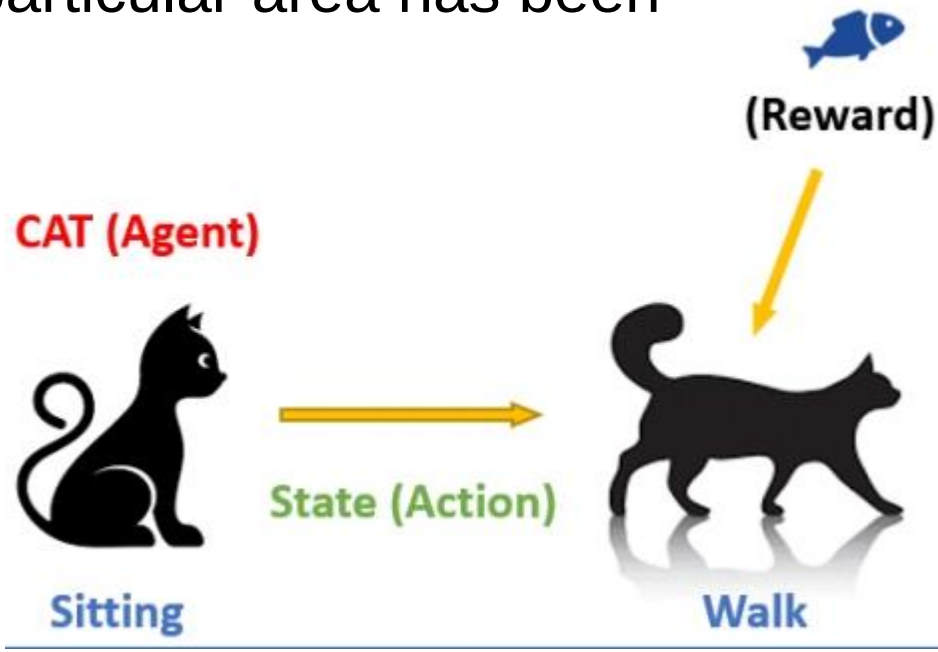
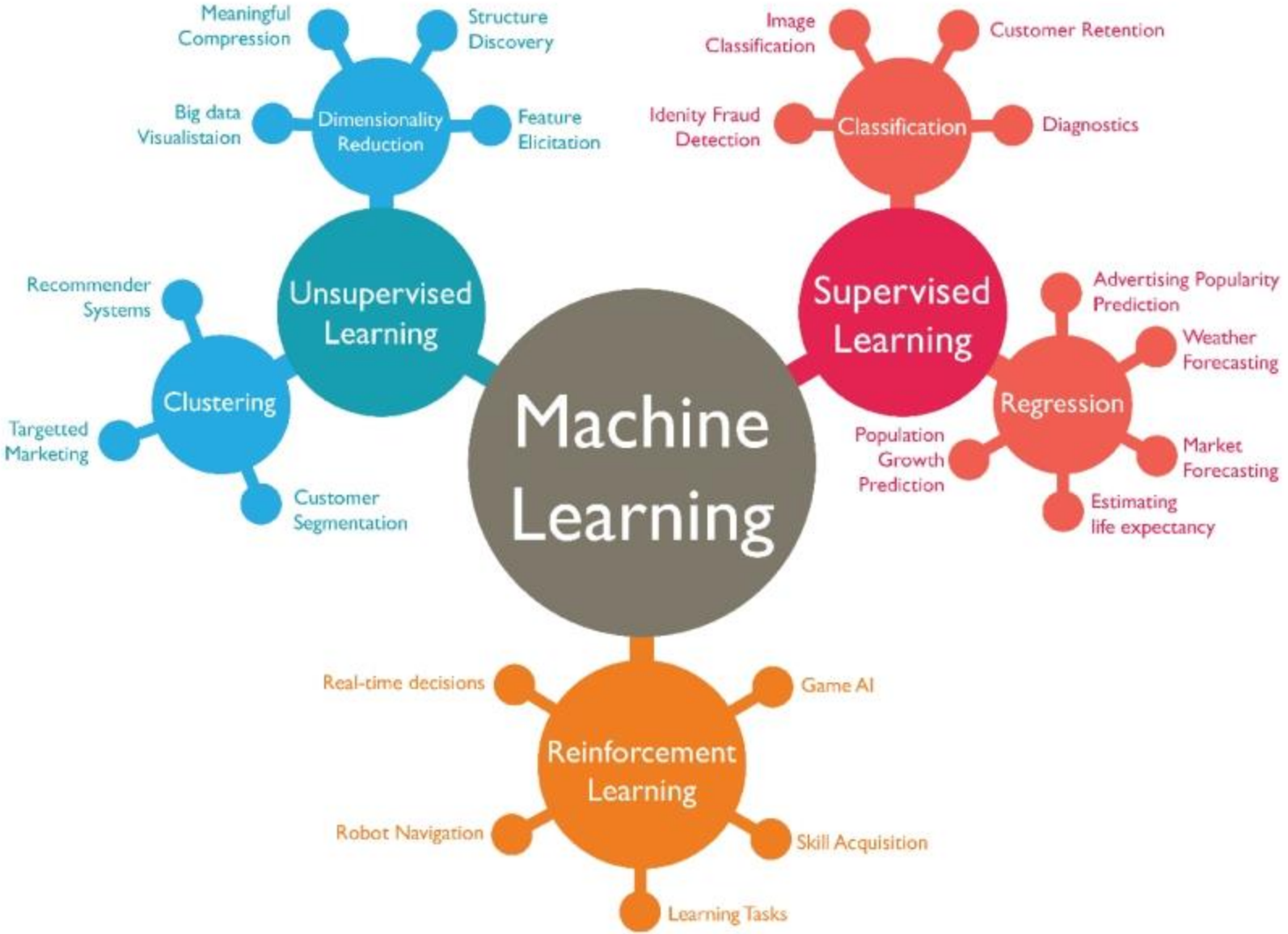


Self organising map



Reinforcement learning

- Reinforcement Learning is, when exposed to an environment, how the machine train itself using trial and error.
- Machine mainly learns from past experiences and tries to perform best possible solution to a certain problem.
- In past couple of years, a lot of improvements in this particular area has been seen.
- Through the interaction with the environment, an agent can then use reinforcement learning to learn a series of actions that maximizes this reward via an exploratory trial-and-error approach or deliberative planning.



Introduction to Data Science : *Machine learning*