**Supervised and Unsupervised learning**

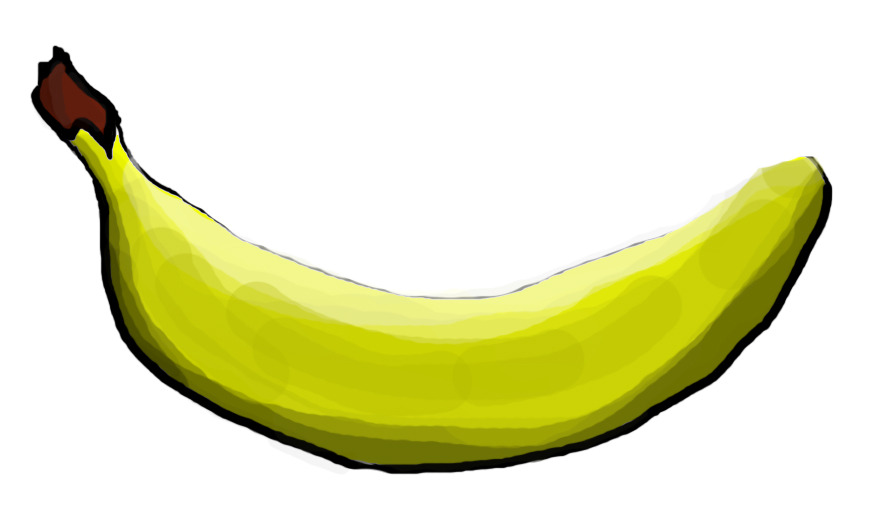
Supervised learning, as the name indicates, has the presence of a supervisor as a teacher. Basically, supervised learning is when we teach or train the machine using data that is well labelled. Which means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that the supervised learning algorithm analyses the training data (set of training examples) and produces a correct outcome from labelled data.

**For instance**, suppose you are given a basket filled with different kinds of fruits. Now the first step is to train the machine with all the different fruits one by one like this:



* If the shape of the object is rounded and has a depression at the top, is red in colour, then it will be labelled as –**Apple**.
* If the shape of the object is a long curving cylinder having Green-Yellow colour, then it will be labelled as –**Banana**.

Now suppose after training the data, you have given a new separate fruit, say Banana from the basket, and asked to identify it. 



Since the machine has already learned the things from previous data and this time has to use it wisely. It will first classify the fruit with its shape and colour and would confirm the fruit name as BANANA and put it in the Banana category. Thus, the machine learns the things from training data (basket containing fruits) and then applies the knowledge to test data (new fruit).

Supervised learning is classified into two categories of algorithms:

* **Classification**: A classification problem is when the output variable is a category, such as “Red” or “blue”, “disease” or “no disease”.
* **Regression**: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Supervised learning deals with or learns with “labelled” data. This implies that some data is already tagged with the correct answer.

**Types:-**

* Regression
* Logistic Regression
* Classification
* Naive Bayes Classifiers
* K-NN (k nearest neighbours)
* Decision Trees
* Support Vector Machine

**Advantages:-**

* Supervised learning allows collecting data and produces data output from previous experiences.
* Helps to optimize performance criteria with the help of experience.
* Supervised machine learning helps to solve various types of real-world computation problems.

**Disadvantages:-**

* Classifying big data can be challenging.
* Training for supervised learning needs a lot of computation time. So, it requires a lot of time.



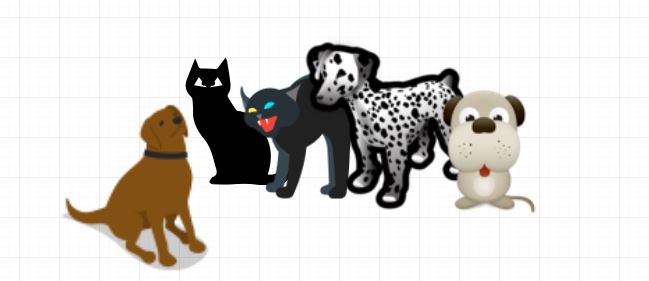
*Steps*

**Unsupervised learning**

Unsupervised learning is the training of a machine using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore, the machine is restricted to find the hidden structure in unlabelled data by itself.

**For instance**, suppose it is given an image having both dogs and cats which it has never seen. 



Thus, the machine has no idea about the features of dogs and cats so we can’t categorize it as ‘dogs and cats ‘. But it can categorize them according to their similarities, patterns, and differences, i.e., we can easily categorize the above picture into two parts. The first may contain all pics having **dogs** in them and the second part may contain all pics having **cats** in them. Here you didn’t learn anything before, which means no training data or examples.

It allows the model to work on its own to discover patterns and information that was previously undetected. It mainly deals with unlabelled data.

Unsupervised learning is classified into two categories of algorithms:

* **Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.
* **Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Types of Unsupervised Learning: -

**Clustering**

1. Exclusive (partitioning)
2. Agglomerative
3. Overlapping
4. Probabilistic

**Clustering Types: -**

1. Hierarchical clustering
2. K-means clustering
3. Principal Component Analysis
4. Singular Value Decomposition
5. Independent Component Analysis

**Supervised vs. Unsupervised Machine Learning**

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Supervised machine learning** | **Unsupervised machine learning** |
| Input Data | Algorithms are trained using labelled data. | Algorithms are used against data that is not labelled |
| Computational Complexity | Simpler method | Computationally complex |
| Accuracy | Highly accurate | Less accurate |
| No. of classes | No. of classes is known | No. of classes is not known |
| Data Analysis | Uses offline analysis | Uses real-time analysis of data |
| Algorithms used | Linear and Logistics regression, Random Forest,  Support Vector Machine, Neural Network, etc. | K-Means clustering, Hierarchical clustering,  Apriori algorithm, etc. |