

# Test-1

## 1. What is difference between Machine learning & Deep learning?

In my opinion if we consider Machine learning as a State of country then the deep learning is the capital of that state. Deep learning is a subset of Machine learning which is a subset of AI.

### Machine learning:

- It uses automated algorithms to predict future decision and models using the input data.
- Various variables are examined present in the dataset.
- There is smart dataset available for computation.
- The output is generally classified value.

### Deep learning:

- It uses neural network to interpret the data features by processing data through various stages.
- The algorithms are implementation become self driven for further process.
- There is generally huge dataset for computation.
- The output can be in various forms of image, text etc.

## 2. Describe the concept of Supervised learning and unsupervised learning?

**Supervised Learning:** In this the system is presented with a given set of inputs and their respective outputs. The goal of the program is to learn from the inputs in order to reproduce the outputs.

The majority of practical machine learning uses supervised learning. Supervised learning is where you have input variables (X) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.  $Y = f(X)$

The goal is to approximate the mapping function so well that when you have new input data (X) that you can predict the output variables (Y) for that data. It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Supervised learning problems can be further grouped into regression and classification problems.

**Unsupervised learning:** In this, there is no target variable in the case of unsupervised learning. The computer is left on its own to find patterns within the data.

Unsupervised learning is where you only have input data (X) and no corresponding output variables. The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data. These are called unsupervised learning because unlike supervised learning above there is no correct answers and there is no teacher. Algorithms are left to their own devices to discover and present the interesting structure in the data. Unsupervised learning problems can be further grouped into clustering and association problems.

## 3. Select 2 examples of supervised learning and unsupervised learning and explain the concept of those two.

### Supervised learning:

**1. Decision trees:** A simple predictive model maps the outcomes of an item to the input data. It is a popular predictive modeling technique, which is used commonly.

Decision tree models are basically of two types:

- **Classification trees:** These refer to dependent variables that take a finite value. In these tree structures, branches represent the rules of the features that lead to the class labels, and leaves represent the class labels of the outcome.

- **Regression trees:** When dependent variables takes continuous values, then they're called regression trees.

2. **Logistic regression:** Logistic regression is another supervised learning technique, which is basically a probabilistic classification model. It is mainly used in predicting a binary predictor, such as whether a customer is going to churn or if a credit card transaction is fraudulent. Logistic regression uses logistics. A logistic function is a very useful function that can take any value from a negative infinity to a positive infinity, and output values from 0 to 1. Hence, it is interpretable as a probability.

**Example:**

1. Let's assume that the training data is a set of text that represents different news articles. These news articles can be related to sports, international, national, and various other categories of news. These categories will act as our labels. From this training data, we'll derive feature vectors where each word could be a vector or certain vectors could be derived from the text. For example, the number of instances of the word "Football" could be a vector, or the number of instances of the word "Prime Minister" could be a vector as well.

2. Even in classification of any product involves the supervised learning. Firstly, the feature vectors are retrieved based on input data.

In these examples, the feature vectors and labels are fed to the Machine Learning Algorithm, which learns from the data. Once the model is trained, it is then used on the new data where the features are again extracted and then inputted to the model, which generates the target data.

**Unsupervised learning:**

**k-means clustering:** The k-means clustering is an unsupervised learning technique that helps in partitioning data of n observations into K buckets of similar observations. The clustering algorithm is called so because it operates by computing the mean of the features which refer to the dependent variables based on which we cluster things, such as segmenting of customers based on an average transaction amount and the average number of products purchased in a quarter of a year. This mean value then becomes the center of a cluster. The number K refers to the number of clusters, that is, the technique consisting of computing a K number of means, leading to the clustering of the data around these k-means.

**Hierarchical clustering:** Hierarchical clustering is an unsupervised learning technique where a hierarchy of clusters is built out of observations. This clustering groups data at various levels of a cluster tree or dendrogram. It is not a single set of clusters, but a hierarchy of multiple levels where clusters at a particular level are joined as clusters on the next level. This allows you to decide the level of clustering that is most suitable.

**Example:**

1. Let's take the example of images that will act as our training and input datasets. The images contain the faces of a human being, horses, and insects. From these images, features are extracted, which will help identify the group that the images belong to. These features are then inputted to the unsupervised machine learning algorithm. The algorithm will find patterns within the data and help in bucketing these images to the respective group. This same algorithm can then be used for new images and helps in bucketing the images into the required buckets.

2. Identifying the crying or smiling people in the photographs is also a type of unsupervised learning problem.

#### 4. Give two example of artificial neural network and its application, when two use, advantages and disadvantages.

**Types:**

1. **Feedforward Artificial Neural Network:** This is the "basic" one, which is used to extract information from the input for a classification or regression task. Ex. If you want to *predict whether someone is overweight* (toy example), knowing the sex, age, height, weight and body fat of the person, this is the neural network you want to use.

**Advantages:**

1. The number of hidden layers will help you in modeling more and more complex functions.
2. By increasing the hidden layers, you will start getting more accurate model.

**Disadvantages:**

1. Having a dense feedforward network for a simple function will lead to high variance. High variance slows a network convergence to a solution.

**2. Convolutional Neural Network:** For most of the computer vision work, this is the best suited neural network. It extracts features of the image through convolutions, which start at a low level (edges) and the more you advance through the network, the higher level the features (shapes, faces etc.). It also uses shift-invariance through pooling, which allow features to be detected whether ever they are in the image. If you want to *detect whether a given image contains a cat*, this is you guy.

**Advantages:**

1. CNNs outperform NNs on conventional image recognition tasks and many other tasks.
2. For a completely new task / problem CNNs are very good feature extractors. This means that you can extract useful attributes from an already trained CNN with its trained weights by feeding your data on each level and tune the CNN a bit for the specific task.

**Disadvantages:**

1. Need a large dataset for training and computations.
2. A convolution is a significantly slower operation.

**Applications:**

**Handwriting Recognition:** The idea of Handwriting recognition has become very important. This is because handheld devices like the Palm Pilot are becoming very popular. Hence, we can use Neural networks to recognize handwritten characters.

The bitmap pattern of the handwritten character is input, with the correct letter or digit as the desired output. Such programs need the user to train the network by providing the program with their handwritten patterns.

The two common applications of handwriting recognition are:

- Optical character recognition for data entry
- Validation of signatures on bank cheque

Feed-forward networks have the following characteristics:

- a. First, they arrange perceptrons in layers, with the first layer taking in inputs and the last layer producing outputs. The middle layers have no connection with the external world, and hence we call them hidden layers.
- b. Each perceptron in one layer is connected to every perceptron on the next layer. Hence information is “fed forward” from one layer to the next in a continuous manner. This explains why we call these networks feed-forward networks.
- c. There is no connection among perceptrons in the same layer.

**Stock Exchange Prediction** – The day-to-day business of the stock market is very complicated. Many factors weigh in whether a given stock will go up or down on any given day. Thus, Neural networks can examine a lot of information in a fast manner and sort it all out. So we can use them to predict stock prices since.

For large business companies, making predictions for stock exchange is common. This is by using parameters, such as current trends, political situation, public view, and economists’ advice.

We can also use neural networks in currency prediction, business failure prediction, debt risk assessment, and credit approval.

## **5. What is data scientist for you? Is it only about machine learning or deep learning? or something else.**

To me Data scientist is a person who understand their "Data". If a person understand the data they are dealing with they can easily analyze, manipulate and visualize your data. The algorithms are just the learning part and with constant practice anyone can master them. So, primarily understanding your data is the job of data scientist.

Data Scientist is not only about machine learning or deep learning, they just provide the parameters, functions and platform to do computation on your data to obtain the desired results. A data scientist needs to learn about the content of the data then only computations can be used using machine learning and deep learning.

Data scientists are a new breed of analytical data expert who have the technical skills to solve complex problems – and the curiosity to explore what problems need to be solved.