

# ML\_ASSISMENT\_01

## 1: What does one mean by the term "machine learning"?

**Sol:** Machine learning is a field of artificial Intelligence or study of computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

## 2: Can you think of 4 distinct types of issues where it shines?

**Sol:**

- a. Analyzing images of products on a production line to auto classifying them.
- b. Analyzing the quantum applications which normally usually take many years.
- c. Medical Science and analyzes DNA structure.
- d. Space Exploration and analysis of subatomic particles

## 3. What is a labeled training set, and how does it work?

**Sol:** In machine learning, a labeled training set is a collection of input data samples, where each sample is associated with one or more output labels that serve as the ground truth for the problem at hand. Also called supervised learning.

Using a labeled training set is to enable a machine learning model to learn how to map inputs to outputs by learning from examples to make accurate predictions on new (unseen data)

## 4. What are the two most important tasks that are supervised?

**Sol:**

- a. Predicting a target numeric Value (Example price of car using features also called predictor).
- b. Classification (Spam filter is one of the example)

**5.Can you think of four examples of unsupervised tasks?**

**Sol:** a.Visualization Algorithm (example feeding unlabelled complex data and get 2D and 3D representation of the data )

- b. Anamoly detection (catching abnormal card transaction or catching Manufacturing defects )
- c. Finding/Exploring the relationship between the complex data (example supply and Demand)
- d.Clustering the data(Group segregation)

**6.State the machine learning model that would be best to make a robot walk through various unfamiliar terrains?**

**Sol:** Reinforcement Learning the learning system called agent ,where agent observe the environment, select and perform action .

It must learn by itself what is the best strategy.

**7.Which algorithm will you use to divide your customers into different groups?**

**Sol:** clustering algorithm.

**8.Will you consider the problem of spam detection to be a supervised or unsupervised learning problem?**

**Sol:** Unsupervised learning.

**9.What is the concept of an online learning system?**

**Sol:** In online learning we train the system incrementally by feeding it data instances sequentially, either individually or in small groups called mini batches.

In this system learn fast as the data arrives continuously(STOCK PRICES)and needs to change rapidly and autonomously.

**10.What is out-of-core learning, and how does it differ from core learning?**

**Sol:** Training the system on huge datasets that cannot fit in one machine's main memory. It loads part of the data and runs a training step on that data. (it is usally done in offline mode) .

**11.What kind of learning algorithm makes predictions using a similarity measure?**

**Sol:** Instance Based learning.

In Instance Based learning the system learns the example by heart and then generalizes to new cases by using a similarity measure to compare them to the learned examples. (spams mail system compare the similar words)

**12.What's the difference between a model parameter and a hyperparameter in a learning algorithm?**  
**Sol:**

Parameters	Hyperparameters
Parameters are the configuration model, which are internal to the model.	Hyperparameters are the explicitly specified parameters that control the training process.
Parameters are essential for making predictions.	Hyperparameters are essential for optimizing the model.
These are specified or estimated while training the model.	These are set before the beginning of the training of the model.
It is internal to the model.	These are external to the model.
These are learned & set by the model by itself.	These are set manually by a machine learning engineer/practitioner.
These are dependent on the dataset, which is used for training.	These are independent of the dataset.
The values of parameters can be estimated by the optimization algorithms, such as Gradient Descent.	The values of hyperparameters can be estimated by hyperparameter tuning.
The final parameters estimated after training decide the model performance on unseen data.	The selected or fine-tuned hyperparameters decide the quality of the model.
Some examples of model parameters are Weights in an ANN, Support vectors in SVM, Coefficients in Linear Regression or Logistic Regression.	Some examples of model hyperparameters are the learning rate for training a neural network, K in the KNN algorithm, etc.

**13.What are the criteria that model-based learning algorithms look for? What is the most popular method they use to achieve success? What method do they use to make predictions?**

**Sol:**Model based learning algorithms search for the optimal value of parameters in a model that will give the best results for the new instances. We often use a cost function or similar to determine what the parameter value has to be in order to minimize the function. The model makes predictions by using the value of the new instance and the parameters in its function.

**14.Can you name four of the most important Machine Learning challenges?**

**Sol:** a. Insufficient quantity of training data  
b. Non representative training data  
c. Poor quality Data  
d. Irrelevant feature

**15.What happens if the model performs well on the training data but fails to generalize the results to new situations? Can you think of three different options?**

**Sol:** Overfitting happens when the model is too complex relative to the amount and noisiness of the training data.

- a.By reducing the number of attributes in the training data or by constraining the model.
- b. Gather more training Data
- c. Reduce the noise in the training data(Removing the outliers)

**16.What exactly is a test set, and why would you need one?**

**Sol:** Test set is the testing dataset upon which the trained model checks for checking how accurately it is predicting the unknown data.

**17.What is a validation set's purpose?**

**Sol:** A validation set is a set of data used to train artificial intelligence (AI) with the goal of finding and optimizing the best model to solve a given problem. Validation sets are also known as dev sets.

**18.What precisely is the train-dev kit, when will you need it, how do you put it to use?**

**Sol:**The goal of dev-set is to rank the models in terms of their accuracy and help us decide which model to proceed further with. Using the Dev set we rank all our models in terms of their accuracy and pick the best performing model. i.e. dev set ranks models similar to a search engine like google rank pages and then pick the top model and hence act as a filter to remove bad models.

**19.What could go wrong if you use the test set to tune hyperparameters?**

**Sol:**

The model will not generalize well to unseen data because it overfits the test set. Tuning model hyper-parameters to a test set means that the hyper-parameters may overfit to that test set. If the same test set is used to estimate performance, it will produce an overestimate.