**ASSIGNMENT-1**

1. **What does one mean by the term "machine learning"?**

Ans:- Machine learning is a subfield of artificial intelligence, which is broadly defined as the capability of a machine to imitate intelligent human behavior. Artificial intelligence systems are used to perform complex tasks in a way that is similar to how humans solve problems.

1. **Can you think of 4 distinct types of issues where it shines?**

Ans:- Machine learning is great for problems whose solution requires a great deal of work or a long list of rules, complex problems that are hard to get a solution of using a traditional method, fluctuating environments and getting insights about complex problems and large data.

1. **What is a labeled training set, and how does it work?**

Ans:- The training set is used to train the algorithm, and then you use the trained model on the test set to predict the response variable values that are already known. The final step is to compare the predicted responses against the actual (observed) responses to see how close they are.

1. **What are the two most important tasks that are supervised?**

Ans:- Two most common supervised tasks are classification and regression.

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

Ans:- Common unsupervised tasks include clustering, visualization, dimensionality reduction, and association rule learning

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

Ans:- The best Machine Learning algorithm to allow a robot to walk in unknown terrain is Reinforced Learning, where the robot can learn from response of the terrain to optimize itself.

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

Ans:- In a business context: Clustering algorithm is a technique that assists customer segmentation which is a process of classifying similar customers into the same segment. Clustering algorithm helps to better understand customers, in terms of both static demographics and dynamic behaviours.

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

Ans:- Spam detection is a supervised machine learning problem. This means you must provide your machine learning model with a set of examples of spam and ham messages and let it find the relevant patterns that separate the two different categories.

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

Ans:- Online learning is education that takes place over the Internet. It is often referred to as “e- learning” among other terms. However, online learning is just one type of “distance learning” - the umbrella term for any learning that takes place across distance and not in a traditional classroom

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

Ans:- It is a way to train your model on data that cannot fit your core memory.” Out-of-core learning refers to the machine learning algorithms working with data that cannot fit into a single machine's memory but can easily fit into some data storage, such as a local hard disk or web repository

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

Ans:- Learning algorithm that relies on a similarity measure to make predictions is instance- based algorithm.

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

Ans:-  Model parameters are internal to the model and estimated from data automatically, whereas Hyperparameters are set manually and are used in the optimization of the model and help in estimating the model parameters.

**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?**

Ans:- Model based learning algorithm 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.

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

Ans:- Four main challenges in Machine Learning include overfitting the data (using a model too complicated), underfitting the data (using a simple model), lacking in data and nonrepresentative data.

**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?**

Ans:- If a model has been trained too well on training data, it will be unable to generalize. It will make inaccurate predictions when given new data, making the model useless even though it is able to make accurate predictions for the training data. This is called overfitting

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

Ans:- A test set in machine learning is a secondary (or tertiary) data set that is used to test a machine learning program after it has been trained on an initial training data set.

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

Ans:- 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?**

Ans:- The goal of dev-set is to rank the models in term of their accuracy and helps us decide which model to proceed further with. Using Dev set we rank all our models in terms of their accuracy and pick the best performing model

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

Ans:- If you use this data to choose hyperparameters, you actually give the model a chance to "see" the test data and to develop a bias towards this test data. Therefore, you actually lose the possibility to find out how good your model would actually be on unseen data (because it has already seen the test data)