Assignment_1 Machine Learning

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

Ans: Machine learning is the study of computer algorithms that can improve automatically through experience and by the use of data.

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

Ans: Recommendation Systems in ecommerce, House Price Prediction, Credit card fraud detection,

Image and Video recognition.

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

Ans: labelled training set have X and Y features both. We are telling ML model to train itself, according to X and Y both. Every X has corresponding Y value. The Machine Learning algorithm then finds relationships between the given parameters, establishing a cause and effect relationship between the variables in the dataset.

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

Ans: Fraudulent Transactions and House Price Prediction.

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

Ans: Common unsupervised tasks include clustering, customer segmentation, Recommender system, anomaly detection.

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

Ans: I would use a reinforcement learning approach. Reinforcement learning is a system where an "agent" observes the environment, selects and performs actions, then receives a reward or punishment based on the result of the action. Over time the agent learns by itself what is the most productive strategy.

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

Ans: I would use clustering techniques like K-means algorithm.

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

Ans: I would consider it as a supervised learning algorithm.

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

Ans:- An online learning system learns from new data on-the-fly. As a result, the system is trained incrementally either by using one example at a time or using a mini-batch approach. This keeps each learning step cheap and memory efficient.

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

Ans:- Out-of-core learning is used when a dataset is too large to fit into a computer's memory. The algorithm loads part of the data, runs a training step, then repeats the process until it has run on all the data.

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

Ans:- Instance-based learning algorithms use a measure of similarity to generalize to new cases. In an instance-based learning system, the algorithm learns the examples by heart, then uses the similarity measure to generalize.

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

Ans:- A model parameter is a variable of the selected model which can be estimated by fitting the given data to the model. In machine learning, a hyperparameter is a parameter whose value is used to control the learning process.

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, accuracy or similar to determine what the parameter value has to be in order to minimize the function. Methods to make predictions are Linear Regression, Logistic Regression, Naïve Bayes, Support Vector Machines,etc.

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

Ans: Not enough training data, Irrelevant Features, Overfitting and Underfitting.

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: Model Overfitting, Unrepresentative Data Sample are the main causes for the model which performs well on the training data but fails to generalize the results to new situations. Three different options to solve this problem are: Reduce the network's capacity by removing layers or reducing the

number of elements in the hidden layers, Apply regularization, which comes down to adding a cost to the loss function for large weights, Use Dropout layers, which will randomly remove certain features by setting them to zero.

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:- – Validation set: A set of examples used to tune the parameters of a classifier, for example to choose the number of hidden units in a neural network.

Test set: A set of examples used only to assess the performance of a fully-specified classifier.

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).