ML_Assignment_01

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

Answer: Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.Why do we use machine learning? Simply put, machine learning allows the user to feed a computer algorithm an immense amount of data and have the computer analyze and make data-driven recommendations and decisions based on only the input data. The three machine learning types are supervised, unsupervised, and reinforcement learning.

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

Answer: Machine learning algorithms have had good results on problems such has spam detection in email, cancer diagnosis, fraudulent credit card transactions, and automatically driving vehicles.

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

Answer: Data labeling (or data annotation) is the process of adding target attributes to training data and labeling them so that a machine learning model can learn what predictions it is expected to make. This process is one of the stages in preparing data for supervised machine learning.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 training set is a portion of a data set used to fit (train) a model for prediction or classification of values that are known in the training set, but unknown in other (future) data. The training set is used in conjunction with validation and/or test sets that are used to evaluate different models.

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

Answer:The two most common supervised tasks are regression and classification. Common unsupervised tasks include clustering, visualization, dimensionality reduction, and association rule learning.

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

Answer:unsupervised tasks include clustering, visualization, dimensionality reduction, and association rule learning.

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

unfamiliar terrains?

Answer:Precise machine learning processes are being used to train robots and improve accuracy. Artificial intelligence teaches functions like spatial relations, grasping objects, computer vision, motion control, etc., in robots to make them understand and work on unseen data and situations.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

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

Answer:We will use the k-means clustering algorithm to derive the optimum number of clusters and understand the underlying customer segments based on the data provided.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 behaviors.

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

problem?

Answer: 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.

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

Answer: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.

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

Answer:The term out-of-core typically refers to processing data that is too large to fit into a computer's main memory. Typically, when a dataset fits neatly into a computer's main memory, randomly accessing sections of data has a (relatively) small performance penalty.Out-of-core leanring refers to the machine learning algorithms working with data cannot fit into the memory of a single machine, but that can easily fit into some data storage such as local hard disk or web repository.

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

Answer:Learning algorithm that relies on a similarity measure to make predictions is instance-based algorithm.There are two major types of prediction algorithms, classification and regression. Classification refers to predicting a discrete value such as a label, while regression refers to predicting a continuous number such as a price.

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

algorithm?

    Answer:In a machine learning model, there are 2 types of parameters:

Model Parameters: These are the parameters in the model that must be determined using the training data set. These are the fitted parameters. Hyperparameters: These are adjustable parameters that must be tuned in order to obtain a model with optimal performance. For example, suppose you want to build a simple linear regression model using an m-dimensional training data set. Then your model can be written as:

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?

Answer:Knowledge of Data: The data's structure and complexity help dictate the right algorithm. Accuracy Requirements: Different questions demand different degrees of accuracy, which influences algorithm selection. Processing Speed: Algorithm choice may depend on the time constraints in place for a given analysis 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?

Answer: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.Offline Learning & Deployment of the model. Machine Learning engineering follows these steps while building an application 1) Data collection 2) Data cleaning 3) Feature engineering 4) Analyzing patterns 5) Training the model and Optimization 6) Deployment.

```
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?


Answer::
Your model is overfitting your training data when you see that the model
performs well on the training data but does not perform well on the evaluation
data. This is because the model is memorizing the data it has seen and is
unable to generalize to unseen examples.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?

Answer: You may already be familiar with training and test sets. When training ML and DL models, you often split the entire dataset into training and test sets. This is because you need a separate test set to evaluate your model on unseen data to increase the generalizing capability of the model.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 validatio sets purpose?

Answer: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. A supervised AI is trained on a corpus of training data.

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

Answer::

The dataset used to fine tune the parameters of network to prevent overfitting.Training Dataset: The sample of data used to fit the model. The actual dataset that we use to train the model (weights and biases in the case of a Neural Network). The model sees and learns from this data.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?

Answer: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).

In [ ]: