Task# 11:

1 .How would you define Machine Learning?

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.

2 .Can you name four types of problems where it shines?

- Spam detection
- Image recognition
- Predicting stock prices
- Recommender systems

3 .What is a labeled training set?

A labeled training set is a dataset that includes both the input data and the corresponding correct output.

4 .What are the two most common supervised tasks?

- Classification
- Regression

5 .Can you name four common unsupervised tasks?

- Clustering
- Anomaly detection
- Association
- Dimensionality reduction

6 .What type of Machine Learning algorithm would you use to allow a robot to walk in various unknown terrains?

A reinforcement learning algorithm.

7 .What type of algorithm would you use to segment your customers into multiple groups?

A clustering algorithm, such as k-means.

8 .Would you frame the problem of spam detection as a supervised learning problem or an unsupervised learning problem?

Supervised learning problem.

9 .What is an online learning system?

An online learning system learns incrementally by feeding it data instances sequentially.

10 .What is out-of-core learning?

Out-of-core learning is a technique used to handle data that doesn't fit into memory by loading it in small chunks.

11 .What type of learning algorithm relies on a similarity measure to make predictions?

Instance-based learning algorithm.

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

- Model parameter: Learned from the data (e.g., weights in a neural network).
- Hyperparameter: Set prior to the training process (e.g., learning rate).

13 .What do model-based learning algorithms search for? What is the most common strategy they use to succeed? How do they make predictions?

- Search for the optimal model parameters.
- Common strategy: Minimize the cost function.
- Predictions: Using the model to make predictions on new data.

14 .Can you name four of the main challenges in Machine Learning?

- Insufficient quantity of training data
- Non-representative training data
- Poor-quality data
- Overfitting and underfitting

15 .If your model performs great on the training data but generalizes poorly to new instances, what is happening? Can you name three possible solutions?

Overfitting.

Solutions:

- Simplify the model
- Use more training data
- Use regularization

16 .What is a test set and why would you want to use it?

A test set is a separate dataset used to evaluate the performance of the model. It helps ensure that the model generalizes well to new data.

17 .What is the purpose of a validation set?

A validation set is used to tune hyperparameters and evaluate the model's performance during the training phase.

18 .What can go wrong if you tune hyperparameters using the test set?

It can lead to overfitting the test set, making the model perform well on the test data but poorly on unseen data.

19 .What is repeated cross-validation and why would you prefer it to using a single validation set?

Repeated cross-validation involves running multiple rounds of cross-validation to reduce variability and get a more accurate measure of model performance.