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

Ans: Machine learning is a subset of AI and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving accuracy/.

Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects

ML builds mathematical models and makes predictions using historical data or information.

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

Ans: Machine learning algorithms have a good results on issues such as

1. spam detection in email.
2. cancer diagnosis
3. fraudulent credit card transactions.
4. automatically driving vehicles.

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

Ans. Labeled training set is a set in which meaningful and informative labels are provided.

Label depends on the context of the problem we are trying to solve.Normally if we are trying to predict a feature based on the others, that feature is the Label.

ex. we have information about pets. This is data. A table of information is the data we have. Each row is a data point which represents a different pet. Each pet is described then, by certain features. features are simply the columns of the table. Then the features here may be size, name, type, color, weight etc. This is what describes our data. Some features are special,though, and we call them labels.Now predict the type of pet we have(ex. dog or cat),based on information on that pet then that is the label. If we are trying to predict if the pet is sick or healthy based on symptoms and other information then that is the label.

A supervised learning model, which applies algorithms to map one input to one output. This requires a labeled set of data where models can learn from to make correct decisions. Data labelling typically starts by asking humans to make judgements about a given piece of unlabelled data.

ex. Labelers may be asked to tag all the images in a dataset where “does the photo contain a bird” is true. The tagging can be as roughly as a simple yes/no or as singular as identifying the specific pixels in the image associated with the bird. The machine learning models use human-provided labels to learn the underlying patterns in a process called “model training”. The result is a trained model that can be used to make predictions on the new data.

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

Ans: Two most common supervised tasks are Classification and Regression.

Regression is the problem of estimating or predicting a continuous quantity.

Classification deals with assigning observations into discrete categories.

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

Ans: Common four 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?

Ans: The models based on Reinforcement learning algorithms which use dynamic programming techniques are the best.

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

Ans: Clustering Algorithms

8.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 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?

Ans: It refers to instruction that is delivered electronically through various multimedia and internet platforms and applications.Its is used interchangeably with other terms such as web-based learning, e-learning, computer-assisted instruction and internet-based learning.

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

Ans: It is a way to train your model on that 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 fir into some data storage, such as a local 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 an instance-based algorithm. The system learns the examples by heart, then generalizes to new cases by using a similarity measure to compare them to the learned examples (or a subset of them).

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

Ans: in summary, model parameters are estimated from data automatically and model hyperparameters are set manually and are used in processes to help estimate 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 algorithms look for optimal values of parameters in a model that will give the best results for the new instances. They learn by minimizing a cost function.

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

Ans: Challenges:

1. Overfitting the training data. (Poor quality data)
2. Underfitting the training data (irrelevant data)
3. Lacking in data ( Not enough data)
4. Non-Representative data. (irrelevant features)

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: This is overfitting of model. due to low bias and high variance. Predictions are inconsistent and accurate on average. This can happen when the model uses a large number of parameters.

to solve the overfitting problem:

1. Get more and clean data.
2. implement a simpler model
3. reduce the parameters. ( redundant features) use regularization to apply penalty to input parameters with the larger coefficients.

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

Ans: A test set in an AIML is a secondary data set that is used to test a machine learning program after it has trained on an initial training data set.

17.What is a validation set's purpose?

Ans: A validation set’s purpose is to train AI with the goal of finding and optimizing the best model to solve a given problem. Validation sets are also known as Development (dev) sets.

The validation test is useful for hyperparameter tuning or selecting the best model out of different models.

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

Ans: The train-dev kit is a training and validation dataset. The training dataset is used to train a model with given parameters. The dev-set is to rank the models in terms of their performance parameter and helps us to decide which model to proceed further with.

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

Ans: The model will overfit and will not perform well on the new dataset. The problem of data leakage will happen exposing the model to the new data set and will fine tune itself. The model will not generalize to unseen data as the model will overfit the test set. If the same test set is used to estimate performance, it will produce an overestimate.