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

Ans:

**As the term implies machines learn and then make predictions like humans. The way we make predictions based on the past experiences and observations, similarly machine learning algorithms are used to make predictions. We use historical data to feed to ML models to train them and once they are trained then we just need to input data in order to get predictions.**

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

Ans:

**Fraud Detection**: this ML application is used to detect frauds even before they occur or in real time as soon as they occur so frauds can be prevented

**Medical Diagnosis:** ML is used here to make predictions if a person has disease or not based on symptoms and medical history. For example, a ML model can be deployed to find if a person is diabetic or not

**Image Recognition**: ML models can be used to recognise images and do face tagging on it.

**Speech Recognition:** ML models are also used to convert speech into text in real time.

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

Ans: **it means the data set in which the values for the target variables exist. This training data set is used to train ML models. Model tries to find hidden pattern in the training data set in order to learn relationship between different independent features and dependent features. In order to achieve all these, we are required to have a data set in which the target variable’s values exist.**

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

Ans: **in the supervised algorithms the two most essential tasks are training the model and testing the model.**

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

Ans:

**Anomaly detection:**  credit card fraud is the example of anomaly detection where any unusual activity is detected and blocked.

**Clustering**: a grouping of a set of items. For example, google news shows a group of latest news or amazon shopping shows the a set of items associated with a particular item.

**Visualization**: unsupervised algorithms can help quickly plot graphs or charts based on unlabelled and complex data.

**Dimensionality reduction**: it is used to remove those features from data set which don’t contribute to the prediction of a variable.

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

Ans: **reinforcement learning because the terrains are unfamiliar for the robot and robot will have to learn by making mistakes and correcting them.**

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

Ans: **classification algorithms can be used to categorise data into different groups.**

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

Ans: **it is a unsupervised learning problem as we don’t have a labelled data.**

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

Ans: **online learning system is a ML method in which the models are not trained on the entire training data set and the data is fed into the model in a sequential manner and the model gets better at making prediction with every cycle of data. It is also good for those scenarios where it is computationally infeasible to train the model with the entire data set.**

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

Ans: **out-of-core algorithms are used to process data which is too large to fit into a computer’s main memory at once. In this case, these algorithms exploit external storage. Core learning is just opposite of out-of-core algorithms and it uses the data which can be fit into a computer’s memory.**

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

Ans: **instance based learning**

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

Ans: **model parameter is the one which is used by the model to make predictions such as coefficient in linear regression model and the hyperparameters are the ones which can be adjusted by users such as number of clusters in K-means.**

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: **it looks for optimal value of parameters in a model. They make predictions and achieve success based on 3 steps in model based machine learning, namely:**

1. **Describe the model**
2. **Condition on observed data**
3. **Perform inference**

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

Ans: **Inadequate training data**

**Poor quality of data**

**Inadequate computational power to build a deep leaning model**

**Overfitting or underfitting of a model**

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: **the model could be overfitted where the model performs well on the training data set but can’t perform well on a different/unseen data set. Overfitting can be caused by inaccurate data or inadequate data and complexity of the model.**

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

Ans: **Test data set is used to determine any model’s efficacy in predicting outcomes. We check the predicted values of the model against the observed values of the test data set to determine how accurately the model has predicted the values.**

17.What is a validation set's purpose?

Ans: **the purpose of the validation set it to avoid overfitting in the model.**

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

Ans: **more often we have to use different models on training data set to find the best model for any given data set. In that case, we may have to use train-dev set in order to rank the models in terms of their accuracy. In order to use the train-dev kit we split the data into three parts usually 70-20-10% being train-dev-test data set.**

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

Ans: **test set is unseen data which is used to determine the efficacy of the model if we tune hyperparameters with test data then you are exposing the unseen data to the model.**