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

Answer. Machine Learning is a concept where a systems has the ability to learn and improve from past experiences without being explicitly programmed when sufficient amount of data is given to the system

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

Answer. Identifying Fraudulent Detection ,Recommendation System, Spam Email detection and  demand forecasting(Sales or anything ) is used in multiple industries

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

Answer. A labelled training set is a collection of data that is marked up or annotated with some information, to show the target class . Labeled training set is used mostly in training supervised learning algorithms

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

Answer. Regression and Classification

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

Answer. 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. Reinforcement Learning is an approach where an agents observes and learns about the environment and adjusts it actions through past experience in order to achieve best solution.

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

Answer. Clustering algorithms helps us to group similar type of data into clusters , in our we can divide the customers based on various factors

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

Answer. Supervised Learning will be the most effective as we have some idea about how spam email looks like if we feed data with both spam or not spam data to the email we will be able to solve it

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

Answer. An Online Learning System learns periodically or with certain interval of time on the go . This makes it capable of adapting rapidly to both changing data and autonomous systems, and of training on very large quantities of data.

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

Answer. Out-of-core learning can handle vast quantities of data that cannot fit in computers main memory so its chops data into mini-batches of data and it is then fed to the system or the algorithms

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

Answer. An instance-based learning system learns the training data by heart. When a new instance is given, it uses a similarity measure to find the most similar learned instances and uses them to make predictions.

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

Answer.  A hyperparameter is a parameter of the learning algorithm itself, not of the model. The algorithm uses these optimal parameters to find the generalized model. A model has one or more model parameters that determine what it will predict given a new instance (e.g. the slope of a linear model)

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. Model-based learning algorithms search for an optimal value for the model parameters such that the model will be generalized, when given a new instance it will be able to predict it correctly. It learns by minimizing a cost function that measures how bad the system is at making predictions on new data, plus a penalty for model complexity if the model is regularized. To make a prediction we fed new instances into the prediction function using all the parameter values found at the training stage

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

Answer.

1. Not gathering enough data or poor data quality to build model

2. full of missing values, outliers in data

3. The features in your dataset are irrelevant

4. Excessively complex models that overfit the data i.e When your model performs well on the training data, but not on test data, you've over-fitted your model

5. When your model is too simple to learn the underlying structure of the data you've under-fitted your 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?

Answer. When your model performs well on the training data, but not on test data, you've over-fitted your model to solve that we can gather as much data as possible, we can reduce the complexity of the model by removing features that are not contributing to solving our problem, reducing noise in data by fixing missing values and errors, and outliers

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

Answer. A test set is used for model generalization error to see that is our model performing well or not when introduced to new data

17.What is a validation set's purpose?

Answer. A validation set is used to compare models. It makes it possible to select the best model and tune the hyperparameters.

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

Answer. Cross-validation is a tool to compare models without needing a separate validation set. It is preferred over validation set because we can save from breaking of part of the training set to create a validation set, as having more data is valuable regardless.

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

Answer. The problem of Over-Fitting is possible and it will perform well on test data only and not a generalized model(worst case model)