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

Answer: Machine learning can be thought of as the process of teaching computers and improve the learning process for them without being explicitly programmed over time based on the experience that a computer comes across it’s learning process.

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

Answer: 4 distinct types of issues where machine learning shines are:

1. Classification: this helps in classification of an entity into a certain group which are available in a list of already available groups.
2. Regression: in this scenario, we simply predict a numerical value based on patterns from other observations learnt in the past via computer.
3. Clustering: in this case, we simply group list of points continuously and it is an ongoing process.
4. Recommendation : based on choices made in the past , computer suggests the same or similar thing to the thing that we have already used up.

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

Answer: A labelled training set is an criterion or a column which helps in training a particular machine learning model by learning by taking this criterion into consideration and put the upcoming non training dataset into appropriate group based on closest similar training set.

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

Answer: Classification and regression are 2 most important tasks that are supervised.

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

Answer: Clustering , recommendation systems ,PCA and neural networks related tasks are completely unsupervised .

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

Answer: Reinforcement learning model is the base in this case scenario.

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

Answer: We will use clustering algorithms like Hierarchical or DBSCAN to do customers into different segments.

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

Answer: Spam detection is supervised as we are simply classifying mails as spam or non-spam via different supervised machine learning algorithms.

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

Answer: Online learning system is a subdivision of ML where model will change from second to second . It adapts to changes no matter how drastic the change is and it learns with every upcoming experience.

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

Answer: Out of core/online learning helps in integration of ML models in iOS applications

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

Answer: Recommender systems works on the similarity measure such as user-user similarity and item-item similarity.

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

Answer: Following are the differences between parameters and hyperparameters:

1. Parameters are used for making predictions and hyperparameters are required for estimating model parameters.
2. Parameters are not set manually whereas hyperparameters are set manually.
3. Parameters are specific to a model whereas hyperparameters are a list of values to be tested for these model specific 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?

Answer: Model based learning algorithms look for patterns amongst all the data rows available, find a pattern from the same and then predict the best possible result for a new data point based on it’s past learning experience. The most popular method by which they achieve success is by minimizing a cost function which helps in knowing how bad/good out model is performing on the new dataset. To make a prediction for a new data point, they feed the entire data point to a function in created by model from it’s past experiences which uses minimized parameter theta obtained by running the cos function continuously.

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

Answer: Following are the main challenges in machine learning:

1. Not collecting enough data to build ML model on top of it or it will memorize the data not find the patterns.
2. Missing values in the dataset.
3. Overfitting model on training data and bad accuracy on test or cross validation data.
4. Model underfitting on train data.

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 our ML model performs well on training data but fails to generalize on test data, we call it overfitting model.

It may happen due to presence of lot of outliers, less amount of total data in general and sometimes even removing features may also help as more features can lead to curse of dimensionality.

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

Answer: Once a training of the model is done on train set, we check how our trained model performs the prediction on the unknown dataset with different and if the predictions is correct or not .This is simply done as need to be sure of our dataset is performing well before exposing this model to real world for making predictions.

17.What is a validation set's purpose?

Answer: Validation is a small part of training data where we might consider that small part might be considered as test set for hyperparamter tuning purposes so that our trained model has seen all kinds of scenario to find a pattern and then prediction on test dataset is done. This helps in getting the best model in terms of generalization error and best values for model parameters.

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

Answer: Our model will never generalize well to unseen data.