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

* Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.
* In classic terms, machine learning is a type of artificial intelligence that enables self-learning from data and then applies that learning without the need for human intervention.

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

* Machine learning algorithms have had good results on problems.
* Such has spam detection in email, cancer diagnosis, fraudulent credit card transactions, and automatically driving vehicles.

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

* A labeled training set is a collection of data where one of the features of the data indicates the class the training example belongs to.
* A labeled training set is used in supervised learning algorithms.

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

* The two most common supervised learning tasks are regression and classification.
* In a regression problem we our prediction is a scalar value.
* When we're trying to solve a classification problem, our output is either 1 or 0.

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

* Common 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?

* I would use a reinforcement learning approach.
* Reinforcement learning is a system where an "agent" observes the environment, selects and performs actions, then receives a reward or punishment based on the result of the action.
* Over time the agent learns by itself what is the most productive strategy.

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

* I would use some sort of clustering algorithm that can find the decision boundaries in the groups automatically.
* This is an unsupervised approach.
* If I already knew the categories of my customers, then I would choose a supervised approach and go with a classification algorithm.

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

* I would frame it as a supervised learning problem because humans have a general idea about what spam is and what it isn't.
* We can use this notion to create a labeled dataset for an algorithm to learn from.

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

* An online learning system learns from new data on-the-fly.
* As a result, the system is trained incrementally either by using one example at a time or using a mini-batch approach.
* This keeps each learning step cheap and memory efficient.

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

* Out-of-core learning is used when a dataset is too large to fit into a computer's memory.
* The algorithm loads part of the data, runs a training step, then repeats the process until it has run on all the data.

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

* Instance-based learning algorithms use a measure of similarity to generalize to new cases.
* In an instance-based learning system, the algorithm learns the examples by heart, then uses the similarity measure to generalize.

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

* A hyperparameter is a parameter of the learning algorithm, not the model.
* For example, in a simple linear regression problem our model is parameterized by theta which is a vector of weights.
* To find the best values for theta we have a cost function which is run repeatedly by the gradient descent algorithm.
* Gradient descent has a hyperparameter called alpha which is the learning rate of the algorithm.

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?

* The goal for a model-based algorithm is to be able to generalize to new examples.
* To do this, model-based algorithms search for optimal values for the model's parameters, often called theta.
* This searching, or "learning", is what machine learning is all about.
* Model-based system learn 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, a new instance's features are fed into a hypothesis function which uses the minimized theta found by repeatedly running the cost function.

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

* Not gathering enough data, or sampling noise. Sampling noise means we'll have non-representative data because of chance.
* Using a dataset that is not representative of the cases you want to generalize to. This is called sampling bias. For example, if you want to train an algorithm with "cat videos", and all your videos are from YouTube, you're training an algorithm to learn about "YouTube cat videos."
* Your dataset is full of missing values, outliers, and noise (poor measurements).
* The features in your dataset are irrelevant. Garbage in, garbage out.
* Feature selection - choose the most relevant features from your dataset
* Feature extraction - combine features in your dataset to generate a new, more useful feature
* When your model performs well on the training data, but not on test data, you've over fit your model. Models that suffer from overfitting do not generalize well to new examples. Overfitting happens when the model is too complex relative to the amount and noisiness of the data.
* Try simplifying the model by reducing the number of features in the data or constraining the parameters by reducing the degrees of freedom.
* Gather more training data.
* Reduce noise in the training data by fixing errors and removing outliers.
* When your model is too simple to learn the underlying structure of the data you've underfit your model.
* Select a more powerful model with more parameters
* Use feature engineering to feed better features to the model
* Reduce the constraints of the model (increase degrees of freedom, reduce regularization parameter, etc.)

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?

* This is a case where the model is overfitting the training data.
* To counteract overfitting, we can reduce the complexity of the model by removing features or constraining the parameters.
* We could gather more data. Finally, we can reduce noisiness in the data by fixing errors and removing outliers.

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

* When we want to know how well our model generalizes to new cases, we prefer to use a test set instead of deploying the system.
* To build the test set we split the training data (50-50, 60-40, 80-20 are common splits) into a training set and test set.
* Our model is training with the training set.
* Then we use the model to run predictions on the test set.
* Our error rate on the test set is called the generalization error or out-of-sample error.
* This error tells us how well our model performs on examples it has never seen before.
* If the training error is low, but the generalization error is high, it means we're overfitting our model.

17.What is a validation set's purpose?

* A validation set is a set of data used to train artificial intelligence (AI) with the goal of finding and optimizing the best model to solve a given problem.
* Validation sets are also known as dev sets.

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

* Training Data Learning algorithm like gradient descent use training data iteratively to learn the parameters of the model.
* In the training process, data is slowly memorized into the parametric aspect of the model with the goal of generalizing this model to unseen data.
* Training process emits the parameters of a model and hence the sole purpose of training data is to decide about which parameters to pick given huge options to choose from.
* To solve any machine learning problem, we have more than one idea in our mind to try e.g. different model architectures, add regularization or not etc. We use training data to emit parameters for each of the model i.e. If we have 3 models in mind to try,
* we use training data to give parameters of each of the model choice but now we need to decide which model is good one out of all these choices

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

* If you use this data to choose hyperparameters, you give the model a chance to "see" the test data and to develop a bias towards this test data.
* Therefore, you lose the possibility to find out how good your model would be on unseen data (because it has already seen the test data).