1. Machine learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed. Machine learning focuses on developing computer programs that can access data and use it to learn for themselves.
2. 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. The training set is **used to train the algorithm, and then you use the trained model on the test set to predict the response variable values that are already known**. The final step is to compare the predicted responses against the actual (observed) responses to see how close they are.
4. The two most common supervised learning tasks are **regression and classification**. In a regression problem we our prediciton is a scalar value. When we're trying to solve a classification problem, our output is either 1 or 0. Supervised learning (SL) is the machine learning task of **learning a function that maps an input to an output based on example input-output pairs**. It infers a function from labeled training data consisting of a set of training examples.

**5.**

**Unsupervised machine learning** is the process of inferring underlying hidden patterns from historical data. Within such an approach, a machine learning model tries to find any similarities, differences, patterns, and structure in data by itself. No prior human intervention is needed.

Let’s get back to our example of a child’s experiential learning.

Picture a toddler. The child knows what the family cat looks like (provided they have one) but has no idea that there are a lot of other cats in the world that are all different. The thing is, if the kid sees another cat, he or she will still be able to recognize it as a cat through a set of features such as two ears, four legs, a tail, fur, whiskers, etc.

In machine learning, this kind of prediction is called unsupervised learning. But when parents *tell* the child that the new animal is a cat – drumroll – that’s considered supervised learning.

Unsupervised learning finds a myriad of real-life applications, including:

Data exploration,customer segmentation,recommender systems, target marketing campaigns, and data preparation and visualization, etc.

Common unsupervised tasks include **clustering, visualization, dimensionality reduction, and association rule learning**.

6.

The best Machine Learning algorithm to allow a robot to walk in unknown terrain is **Reinforced Learning**, where the robot can learn from response of the terrain to optimize itself. I would use a reinforcement learning approach. Reinforcement learning is a system where an "agent" observes the environment, selects and performs actions, then recieves a reward or punishment based on the result of the action. Over time the agent learns by itself what is the most productive strategy.

**Autonomous learning**, which is a variant of self-supervised learning involving deep learning and unsupervised methods, has also been applied to robot and control tasks.

7.

We will use the **k-means clustering algorithm** to derive the optimum number of clusters and understand the underlying customer segments based on the data provided.

If you don't know how to define the groups, then you can use a **clustering algorithm (unsupervised learning)** to segment your customers into clusters of similar customers.

8.

Spam detection is a **supervised machine learning problem**. This means you must provide your machine learning 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.

Online learning is **education that takes place over the Internet**. It is often referred to as “e- learning” among other terms. However, online learning is just one type of “distance learning” - the umbrella term for any learning that takes place across distance and not in a traditional classroom.

10.

Out-of-core learning refers to a set of algorithms working with data that cannot fit into the memory of a single computer, but that can easily fit into some data storage such as a local hard disk or web repository. Your available RAM, the core memory on your single machine, may indeed range from a few gigabytes (sometimes 2 GB, more commonly 4 GB, but we assume that you have 2 GB at maximum) up to 256 GB on large server machines. Large servers are like the ones you can get on cloud computing services such as Amazon **Elastic Compute Cloud** (**EC2**), whereas your storage capabilities can easily exceed terabytes of capacity using just an external drive (most likely about 1 TB but it can reach up to 4 TB).As machine learning is based on globally reducing a cost function, many algorithms initially have been thought to work using all the available data and having access to it at each iteration of the optimization process. This is particularly true for all algorithms based on statistical.

11.

Learning algorithm that relies on a similarity measure to make predictions is **instance-based algorithm**.

The algorithm learns the examples by heart, then uses the similarity measure to generalize.

12.

Model Parameters: These are the parameters in the model that must be determined using the training data set. These are the fitted parameters.

Hyperparameters: These are adjustable parameters that must be tuned in order to obtain a model with optimal performance.

### 13) What do model based learning algorithms search for? What is the most common strategy they use to succeed? How do they 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 predicitons 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 4 of the main challenges in Machine Learning?

* Not gathering enough data, or sampling noise. Sampling noise means we'll have non-representative data as a result 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 actually training an algorithm to learn about "YouTube cat videos."
* Your dataset is full of missing values, outliers, and noise (poor measurments).
* 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 simplyfying 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) If your model performs great on the training data but generalizes poorly to new instances, what is happening? Can you name 3 possible solutions?

This is a case where the model is overfitting the training data. To couteract 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 is a test set and why would you want to use it?

When we want to know how well our model generalizes to new cases we prefer to use a test set instead of actually 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 the purpose of a validation set?

Let's say we have a linear model and we want to perform some hyperparameter tuning to reduce the generalization error. One way to do this 100 different models with 100 different hyperparameter values using the training set and finding the generalization error with the test set. You find the best hyperparameter value gives you 5% generalization error.

So you launch the model into production and find you're seeing 15% generalization error. This isn't going as expected. What happened?

The problem is that for each iteration of hyperparameter tuning, you measured the generalization error then updated the model using the same test set. In other words, your produced the best generalization error for the test set. The test set no longer represents cases the model hasn't seen before.

A common solution to this problem is to have a second holdout set called the validation set. You train multiple models with various hyperparameters using the training set, you select the model and hyperparameters that perform best on the validation set, and when you are happy about your model you run a single final test against the test set to get an estimate of the generalization error.

### 18) What can go wrong if you tune hyperparameters using the test set?

Your model will not be generalizable to new examples.

### 19) What is cross-validation and why would you prefer it to a validation set?

Cross-validation helps us compare models without wasting too much training data in the validation set.