**1. How you can define Machine Learning?**

Machine learning is the concept that a computer program can learn and adapt to new data without human intervention. Machine learning is a field of artificial intelligence (AI) that keeps a computer’s built-in algorithms current regardless of changes in the worldwide economy.

Or

Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions

**2. What do you understand Labelled training dataset?**

Labelled training set is a set of training data which has a solution to the problem or task (a.k.a. label).Labelled data is a designation for pieces of data that have been tagged with one or more labels identifying certain properties or characteristics, or classifications or contained objects. Labels make that data specifically useful in certain types of machine learning known as supervised machine learning setups. Labelled dataset is the process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to provide context so that a machine learning model can learn from it.

**3. What are 2 most common supervised ML tasks you have performed so far?**

Two most common supervised tasks are classification and regression

**4. What kind of Machine learning algorithm would you used to walk robot in various unknown area?**

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.

**5. What kind of ML algo you can use to segment your user into multiple groups?**

The best algorithm to segment customers into multiple groups is either supervised learning (if the groups have known labels) or unsupervised learning (if there are no group labels).

**6. What type of learning algo realised on similarity measure to make a prediction?**

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

**7. What is an online learning system?**

Online learning system is a learning system in which the machine learns as data is given in small streams continuously. In computer science, online machine learning is a method of machine learning in which data becomes available in a sequential order and is used to update the best predictor for future data at each step, as opposed to batch learning techniques which generate the best predictor by learning on the entire training data set at once. Online learning is a common technique used in areas of machine learning where it is computationally infeasible to train over the entire dataset, requiring the need of out-of-core algorithms. It is also used in situations where it is necessary for the algorithm to dynamically adapt to new patterns in the data, or when the data itself is generated as a function of time, e.g., stock price prediction. Online learning algorithms may be prone to catastrophic interference, a problem that can be addressed by incremental learning approaches.

**8. What is out of core learning?**

Out-of-core learning system is a system that can handle data that cannot fit into your computer memory. It uses online learning system to feed data in small bits. 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 yourstorage 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

**9. Can you name couple of ml challenges that you have faced?**

Four main challenges in Machine Learning include overfitting the data (using a model too complicated), underfitting the data (using a simple model), lacking in data and nonrepresentative data.

**10. Can you please give 1 example of hyperparameter tuning wrt some classification algorithm?**

Machine learning algorithms have hyperparameters that allow you to tailor the behavior of the algorithm to your specific dataset.

[Hyperparameters](https://machinelearningmastery.com/difference-between-a-parameter-and-a-hyperparameter/) are different from parameters, which are the internal coefficients or weights for a model found by the learning algorithm. Unlike parameters, hyperparameters are specified by the practitioner when configuring the model.

Typically, it is challenging to know what values to use for the hyperparameters of a given algorithm on a given dataset, therefore it is common to use random or grid search strategies for different hyperparameter values.

The more hyperparameters of an algorithm that you need to tune, the slower the tuning process. Therefore, it is desirable to select a minimum subset of model hyperparameters to search or tune. Not all model hyperparameters are equally important. Some hyperparameters have an outsized effect on the behavior, and in turn, the performance of a machine learning algorithm. As a machine learning practitioner, you must know which hyperparameters to focus on to get a good result quickly.

## Logistic Regression

Logistic regression does not really have any critical hyperparameters to tune.

Sometimes, you can see useful differences in performance or convergence with different solvers (solver).

* **solver** in [‘newton-cg’, ‘lbfgs’, ‘liblinear’, ‘sag’, ‘saga’]

Regularization (penalty) can sometimes be helpful.

* **penalty** in [‘none’, ‘l1’, ‘l2’, ‘elasticnet’]

**Note**: not all solvers support all regularization terms.

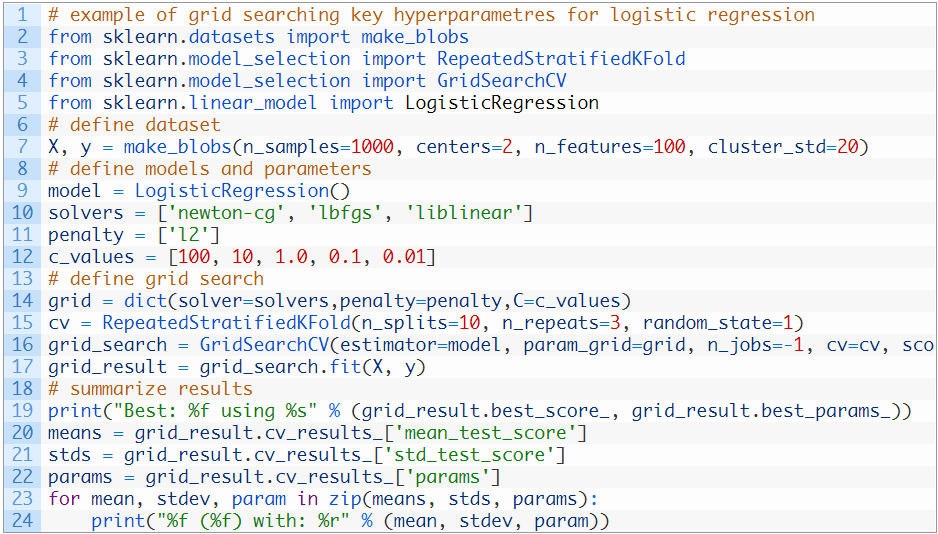
The C parameter controls the penality strength, which can also be effective.

* **C** in [100, 10, 1.0, 0.1, 0.01]

For the full list of hyperparameters, see:

* [sklearn.linear\_model.LogisticRegression API](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html).

The example below demonstrates grid searching the key hyperparameters for LogisticRegression on a synthetic binary classification dataset.



**11. What is out of bag evaluation? What do you understand by hard & soft voting classifier?**

In random forests, there is no need for cross-validation or a separate test set to get an unbiased estimate of the test set error. It is estimated internally, during the run, as follows:

Each tree is constructed using a different bootstrap sample from theoriginal data. About one-third of the cases are left out of the bootstrap sample and not used in the construction of the kth tree.

Put each case left out in the construction of the kth tree down the kth tree to get a classification. In this way, a test set classification is obtained for each case in about one-third of the trees. At the end of the run, take j to be the class that got most of the votes every time case n was oob. The proportion of times that j is not equal to the true class of n averaged over all cases is the oob error estimate. This has proven to be unbiased in many tests.

In classification, a hard voting ensemble involves summing the votes for crisp class labels from other models and predicting the class with the most votes. A soft voting ensemble involves summing the predicted probabilities for class labels and predicting the class label with the largest sum probability.

**14. Let’s Suppose I have trained 5 diff model with same training dataset & all of them have achieved 95%precision. Is there any chance that you can combine all these models to get better result? If yes, How? If no, Why?**

Yes, we can use K fold

Cross-validation is another method to estimate the skill of a method on unseen data. Like using a train-test split.Cross-validation systematically creates and evaluates multiple models on multiple subsets of the dataset. This, in turn, provides a population of performance measures.

We can calculate the mean of these measures to get an idea of how well the procedure performs on average. We can calculate the standard deviation of these measures to get an idea of how much the skill of the procedure is expected to vary in practice.

This is also helpful for providing a more nuanced comparison of one procedure to another when you are trying to choose which algorithm and data preparation procedures to use. Also, this information is invaluable as you can use the mean and spread to give a confidence interval on the expected performance on a machine learning procedure in practice. Both train-test splits and k-fold cross validation are examples of resampling methods.

**15. What do you understand by Gradient decent? How will you explain Gradient decent to a kid?**

Gradient Descent is an optimization algorithm used for minimizing the cost function in various machine learning algorithms. It is basically used for updating the parameters of the learning model.

**Types of gradient Descent:**

1. **Batch Gradient Descent:**This is a type of gradient descent which processes all the training examples for each iteration of gradient descent. But if the number of training examples is large, then batch gradient descent is computationally very expensive. Hence if the number of training examples is large, then batch gradient descent is not preferred. Instead, we prefer to use stochastic gradient descent or mini-batch gradient descent.
2. **Stochastic Gradient Descent:** This is a type of gradient descent which processes 1 training example per iteration. Hence, the parameters are being updated even after one iteration in which only a single example has been processed. Hence this is quite faster than batch gradient descent. But again, when the number of training examples is large, even then it processes only one example which can be additional overhead for the system as the number of iterations will be quite large.
3. **Mini Batch gradient descent:** This is a type of gradient descent which works faster than both batch gradient descent and stochastic gradient descent. Here *b* examples where*b<m* are processed per iteration. So even if the number of training examples is large, it is processed in batches of b training examples in one go. Thus, it works for larger training examples and that too with lesser number of iterations.

An easy explanation is as follows:

Imagine that you were in the hills, and had to find the lowest valley.

Do this repeatedly:

Start from any point on any hill

Look in all four directions (ahead, behind, left, right) to determine where you might be able to descend (rather than ascend)

Take a step in that direction

Return to step (b) above

If you have reached a point where taking a step in any direction doesn’t make a difference, you’re at a minimum (you’ve reached the valley)

Caveat: When you’re at a valley, from where you can see some other cavern or valley, you’re likely to be at a “local minimum”

Notes:

How big each step you take down the hills are - that represents your learning rate. Too big a step, and you bounce between hills, and too small a step, and you take forever to descend into the valley

Gradient descent algorithms do much the same things, but in an arbitrary number of dimensions. The problem I’ve described above is in three dimensions of space, with steps taken over time. In mathematical functions, you can define arbitrarily large spaces where gradient descent can take place.

**16.Can you please explain diff between regression & classification?**

Classification and Regression are two major prediction problems which are usually dealt with Data mining and machine learning.

**Classification** is the process of finding or discovering a model or function which helps in separating the data into multiple categorical classes i.e. discrete values. In classification, data is categorized under different labels according to some parameters given in input and then the labels are predicted for the data.   
The derived mapping function could be demonstrated in the form of “IF-THEN” rules. The classification process deal with the problems where the data can be divided into binary or multiple discrete labels.

Let’s take an example, suppose we want to predict the possibility of the wining of match by Team A on the basis of some parameters recorded earlier. Then there would be two labels Yes and No.

**Regression** is the process of finding a model or function for distinguishing the data into continuous real values instead of using classes or discrete values. It can also identify the distribution movement depending on the historical data. Because a regression predictive model predicts a quantity, therefore, the skill of the model must be reported as an error in those predictions   
Let’s take a similar example in regression also, where we are finding the possibility of rain in some particular regions with the help of some parameters recorded earlier. Then there is a probability associated with the rain.

**17.Explain a clustering algorithm of your choice.**

K-Means Clustering

K-Means is by far the most popular clustering algorithm given that it is very easy to understand and apply to a wide range of data science and machine learning problems. Here’s how you can apply the K-Means algorithm to your clustering problem.

The first step is to select a number of clusters randomly, each of which is represented by a variable ‘k’. Next, each cluster is assigned a centroid, i.e., the centre of that particular cluster. It is important to define the centroids as far off from each other as possible to reduce variation. After all the centroids are defined, each data point is assigned to the cluster whose centroid is at the closest distance.

Once all data points are assigned to respective clusters, the centroid is again assigned for each cluster. Once again, all data points are rearranged in specific clusters based on their distance from the newly defined centroids. This process is repeated until the centroids stop moving from their positions.

K-Means algorithm works wonders in grouping new data. Some of the practical applications of this algorithm are in sensor measurements, audio detection, and image segmentation.

**18.How you can explain ML, DL, NLP, Computer vision & reinforcement learning with example in your own term**

 (AI) is the domain of producing intelligent machines. ML refers to systems that can assimilate from experience (training data) and Deep Learning (DL) states to systems that learn from experience on large data sets. ML can be considered as a subset of AI. Deep Learning (DL) is ML but useful to large data sets. The figure below roughly encapsulates the relation between AI, ML, and DL: In summary, DL is a subset of ML & both were the subsets of AI.

Additional Information: ASR (Automatic Speech Recognition) & NLP (Natural Language Processing) fall under AI and overlay with ML & DL as ML is often utilized for NLP and ASR tasks.

NLP enables computers to understand natural language as humans do. Whether the language is spoken or written, natural language processing uses artificial intelligence to take real-world input, process it, and make sense of it in a way a computer can understand. Just as humans have different sensors -- such as ears to hear and eyes to see -- computers have programs to read and microphones to collect audio. And just as humans have a brain to process that input, computers have a program to process their respective inputs. At some point in processing, the input is converted to code that the computer can understand.

Computer vision:

Computer Vision, often abbreviated as CV, is defined as a field of study that seeks to develop techniques to help computers “see” and understand the content of digital images such as photographs and videos.The problem of computer vision appears simple because it is trivially solved by people, even very young children. Nevertheless, it largely remains an unsolved problem based both on the limited understanding of biological vision and because of the complexity of vision perception in a dynamic and nearly infinitely varying physical world.

**Reinforcement Learning:**  
The model learns through a trial-and-error method. This kind of learning involves an agent that will interact with the environment to create actions and then discover errors or rewards of that action.

Machine Learning involves algorithms that learn from patterns of data and then apply it to decision making. Deep Learning, on the other hand, is able to learn through processing data on its own and is quite similar to the human brain where it identifies something, analyse it, and makes a decision.

The key differences are as follow:

The manner in which data is presented to the system.

Machine learning algorithms always require structured data and deep learning networks rely on layers of artificial neural networks.

**19. How you can explain semi-supervised ML in your own way with example?**

Semi-supervised machine learning is a combination of supervised and unsupervised machine learning methods.

With more common supervised machine learning methods, you train a machine learning algorithm on a “labeled” dataset in which each record includes the outcome information. This allows the algorithm to deduce patterns and identify relationships between your target variable and the rest of the dataset based on information it already has. In contrast, unsupervised machine learning algorithms learn from a dataset without the outcome variable. In semi-supervised learning, an algorithm learns from a dataset that includes both labeled and unlabeled data, usually mostly unlabeled.

Examples of semi supervised Ml

Speech Analysis: Since labeling of audio files is a very intensive task, Semi-Supervised learning is a very natural approach to solve this problem.

Internet Content Classification: Labeling each webpage is an impractical and unfeasible process and thus uses Semi-Supervised learning algorithms. Even the Google search algorithm uses a variant of Semi-Supervised learning to rank the relevance of a webpage for a given query.

Protein Sequence Classification: Since DNA strands are typically very large in size, the rise of Semi-Supervised learning has been imminent in this field.

**20. What is difference between abstraction & generalization in your own word.**

Abstraction is the process of removing details of objects. We step back from concrete objects to consider a number of objects with identical properties. So a concrete object can be looked at as a “superset” of a more abstract object.

A generalization, then, is the formulation of general concepts from specific instances by abstracting common properties. A concrete object can be looked at as a “subset” of a more generalized object.

In other words:

1. For any two concepts A and B, A is an **abstraction** of B if and only if:
   * Every instance of concept B is also an instance of concept A
2. For any two concepts A and B, A is a **generalization** of B if and only if:
   * Every instance of concept B is also an instance of concept A
   * There are instances of concept A which are not instances of concept B