1. What does one mean by the term "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](https://www.investopedia.com/alternative-investments-4427781)

(AI) that keeps a computer’s built-in [algorithms](https://www.investopedia.com/terms/a/algorithm.asp) current regardless of changes in the worldwide economy.

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

Machine learning can be helpful in

1.Scam detection

2.Churn analysis

3.Image recognition

4.Stock Prediction

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

If we provide any tag or classes to raw data it is called as lab data in labelled training the sample input and output is provided by user or domain expert which is further divided into training and testing set

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

The two most important task that are supervised are

Fraud detection

Spam filter

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

Anomaly detection

Customer and market Segmentation

Recommender System

Target Marketing

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

Reinforcement Learning

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

We can use clustering algorithm to divide the customers into different groups like k mean clustering.

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

Spam detection is part of supervised learning as we are giving what email should be consider spam(Ham) or not Ham

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

online machine learning is a method of [machine learning](https://en.wikipedia.org/wiki/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

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

When we to train the model which more than the memory of Ram it is called as out of core

Learning here the we have streaming which can be divided into batch and then can be trained incrementally we use technique like sgd regressor.

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

Instance based learning are the systems that learn the training examples by heart and

Then generalizes to new instances by some similarity or measure .

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

**Model parameters are configuration variables that are internal to the model, and a model learns them on its own.** For example, W Weights or Coefficients of independent variables in the Linear regression model. Weights or Coefficients of independent variables SVM, weight, and biases of a neural network, cluster centroid in clustering.

Hyperparameters are those parameters that are explicitly defined by the user to control the learning process.

* These are usually defined manually by the machine learning engineer.
* One cannot know the exact best value for hyperparameters for the given problem. The best value can be determined either by the rule of thumb or by trial and error

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 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 four of the most important Machine Learning challenges?

1. Bad Quality of Data

2.Invalid features

3.Over fitting

4.Underfitting

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?

When the model perform wells on training data but performs badly on test data it is called overfitting

Three options to avoid overfitting are

- Gather more data for “training data”

- Reduce the noise level

- Select one with fewer parameters

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

Test set: The test dataset is the subset of the actual dataset, which is not yet used to train the model. The model is unaware of this dataset. So, by using the test dataset, we can compute the response of the created model on hidden data. We evaluate the model’s performance on the basis of the test dataset.

17.What is a validation set's purpose?

The validation dataset is a hybrid: it is training data that is used for testing, but it is not incl.uded in either the low-level training or the final testing. Early stopping is a technique in which the candidate models are iterations of the same network, and training stops when the error on the validation set develops, choosing the previous model – the one with the least error

We use several technique like k fold cross validation to achieve this

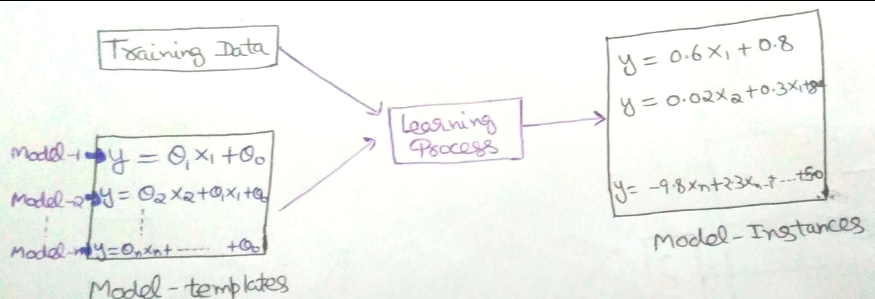
18.What precisely is the train-dev kit, when will you need it, how do you put it to use?

**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 make a decision about which parameters to pick given huge options to choose from.

To solve any machine learning problem we have more than one ideas 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.



The goal of **dev-set** is to rank the models in term of their accuracy and helps us decide which model to proceed further with. Using Dev set we rank all our models in terms of their accuracy and pick the best performing model. i.e. dev set ranks models similar to a search engine like google rank pages and then pick the top model and hence act as a filter to remove bad models.

**why do we need dev-set** ?

In the process of building models that learn from data, we need to find the best parameters of the model and best model out of all other available ones. If we don’t have dev data then we’ll train all the models and pick the model with the best performance on training data. By doing so, we are taking 2 decisions with a single process i.e.

1. Parameter choice
2. Model choice

While having dev set split, first training algorithm makes the choice for optimal parameters and then those parameters are used on dev data to help us find best model architecture as compared to both choices made together by learning algorithm itself.

Dev set helps us in reducing the complexity of diagnosis if things won’t go fine i.e. we’ll be able to assign error to either choice of parameters or picking up model architecture very concretely.

Lack of dev set and using only the training set doesn’t give you clue about which choice went wrong and luck rather than skill will be helpful to debug your learning algorithm there and to make decision further to improve the model accuracy.

**General piece of advice to avoid pitfalls:**

1. Choose dev/test from the distribution that reflect what data you expect to see in the future. If the dev/test set is made up of distribution different than real data distribution, your model might not work as expected.
2. You can have dev/test distribution different from training data distribution but then your diagnosis will become heavier because you will be dealing with train/test data mismatch problem as well.
3. Choose dev and test set from same distribution because it will help you with clear diagnosis when the model will not be working on the test set, which is you have overfitted the dev set and acquire more dev data.
4. Dev set should be large enough to rank the models and test data should be large enough to give you unbiased performance measures of your model.
5. You never make any decision in your learning process by looking at the test set. In fact, you even forget that you have any more data apart from train and dev set. Once all the decisions about learning algorithm are made you use test-set to only evaluate the performance of your model.

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

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