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

Ans=>Machine Learning Means learning parameter using mathematical operation. When we import linear regression we actually give y=mx+c to our system and then our system try to find best value of m and c (in order to minimize loss Yloss=Ypredicted-yactual).Here our system learn value of m and c.

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

Ans=>1🡪Image Recognition: Image recognition is one of the most common applications of machine learning. ...

1. Speech Recognition. ...
2. Traffic prediction: ...
3. Product recommendations: ...
4. Self-driving cars: ...
5. Email Spam and Malware Filtering: ...
6. Virtual Personal Assistant: ...
7. Online Fraud Detection:

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

Ans-> For example, a data label might indicate **whether a photo contains a horse or a cow**, which words were uttered in an audio recording, what type of action is being performed in a video, what the topic of a news article is, what the overall sentiment of a tweet is, or whether a dot in an X-ray is a tumor.

We use labelled data for supervised ml technique.

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

Ans-> The two most common supervised learning tasks are **regression and classification**

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

Ans-> Association Rules(PCA,SVD), Clustering, Anamoly Detection,Visualization.

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

Ans-> 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.

Steps of reinforcement learning.

Input: The input should be an initial state from which the model will start

Output: There are many possible outputs as there are a variety of solutions to a particular problem

Training: The training is based upon the input, The model will return a state and the user will decide to reward or punish the model based on its output.

The model keeps continues to learn.

The best solution is decided based on the maximum reward

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

Ans->Unsupervised machine learning Algorithm. Ex- KMeans, DBSCAN, Hierarchical clustering.

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

Ans-> 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.What is the concept of an online learning system?

Ans->Data is dynamic. Its property changes for example pre COVID 19 and post COVID 19.

Traditional machine learning techniques run in batch mode. For example, supervised learning tasks where the complete training data is fed in advance to train a model by applying certain algorithms. Such an approach requires entire training data available prior to the learning task and the process is also in offline mode due to expensive training costs. Conventional techniques suffer from some critical drawbacks like low efficiency in both time and space cost; and poor scalability for large-scale applications because the model often has to retrain from scratch for new data.

On the other hand, online learning is a combination of different techniques of ML where data arrives in sequential order and the learner (algorithm/model) aims to learn and update the best predictor for future data at every step. Online learning is able to overcome drawbacks of offline learning like models can be updated instantly for any change in data. Therefore online learning is far more efficient and scalable for large-scale learning tasks in real-world data, analytics, and various applications where data is not only large in size but also arrives at high velocity.

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

Ans-> 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?

Ans->Instance Based learning.

The Machine Learning systems which are categorized as instance-based learning are the systems that learn the training examples by heart and then generalizes to new instances based on some similarity measure. It is called instance-based because it builds the hypotheses from the training instances. It is also known as memory-based learning or lazy-learning. The time complexity of this algorithm depends upon the size of training data. The worst-case time complexity of this algorithm is O (n), where n is the number of training instances.

For example, If we were to create a spam filter with an instance-based learning algorithm, instead of just flagging emails that are already marked as spam emails, our spam filter would be programmed to also flag emails that are very similar to them. This requires a measure of resemblance between two emails. A similarity measure between two emails could be the same sender or the repetitive use of the same keywords or something else.

Some of the instance-based learning algorithms are :

K Nearest Neighbor (KNN)

Self-Organizing Map (SOM)

Learning Vector Quantization (LVQ)

Locally Weighted Learning (LWL)

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

Ans->for example (Elastic net regression y=mx+c),

Parameter=m and c(model will find this value by minimizing loss function)

Hyper parameter=l1 ratio,alpha(by tuning on training data we found this value)

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. In order 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?

Ans-> 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?

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). 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.)

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.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 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 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 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 a validation set's purpose?

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

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.

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

Your model will not be generalizable to new examples.