

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

Ans Machine Learning is all about learning building systems that can learn from data. Learning means getting better at some task, given some performance.

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

Ans Machine learning is great for complex problems for which we have no algorithmic solution, to replace long lists of hand-tuned rules, to build systems that adapt to fluctuating environments, and finally to help humans learn (e.g. data mining).. Such as Email spam detection, Credit risk modelling, Cancer diagnosis, etc.

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

Ans A labeled training set is a training set that contains the desired solution for each instance.

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

Ans Two most important tasks that are supervised are regression and classification

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

Ans: Four examples of unsupervised tasks:

- Clustering
- Visualization
- Dimensionality
- Association Rule learning

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

Ans Reinforcement learning would be best to make a robot walk through various unfamiliar terrains.

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

Ans

- If do not know how to define groups then will use Clustering algorithm (unsupervised learning) to divide customers into different groups.
- If we know what groups we would like to have in that case we would use classification algorithm (Supervised learning) it will divide customers into these groups.

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

Ans Spam detection is a supervised learning problem: the algorithm is fed many emails along with their label such as spam or not spam.

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

Ans Online learning system is a learning system in which the machine learns as data is given in small streams continuously. It can learn incrementally, as opposed to a batch learning system. This makes it capable of adapting rapidly to both changing data and autonomous systems, and of training on very large quantities of data.

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

Ans : Out-of-core algorithms can handle vast quantities of data that cannot fit in a computer's main memory. An out-of-core learning algorithm chops the data into mini-batches and uses online learning techniques to learn from these mini-batches

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

Ans An instance-based learning system makes predictions using a similarity measure; then, when given a new instance, it uses a similarity measure to find the most similar learned instances and uses them to make predictions.

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

Ans A model has one or more model parameters that determine what it will predict given a new instance (e.g. the slope of a linear model). A learning algorithm tries to find optimal values for these parameters such that the model generalizes well to new instances. A hyperparameter is a parameter of the learning algorithm itself, not of the model (e.g. the amount of regularization to apply).

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

Model-based learning algorithms search for an optimal value for the model parameters such that the model will generalize well to new instances. We usually train such systems by minimizing a cost function that measures how bad the system is at making predictions on the training data, plus a penalty for model complexity if the model is regularized. To make predictions, we feed the new instance's features into the model's prediction function, using the parameter values found by the learning algorithm.

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

Ans

Some of the main challenges in Machine Learning are the lack of data, poor data quality, non-representative data, uninformative features, excessively simple models that under-fit the training data, and excessively complex models that overfit the data.

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?

Ans If a model performs great on the training data but fails to generalize to new instances, the model is likely overfitting the training data. To solve this, we can do any of the following three: get more data, implement a simpler model, or eliminate outliers or noise from the existing data set.

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

Ans Test set is a set that you test your model (fit using training data) to see how it performs. Test set is necessary so that you can determine how good (or bad) your model performs

17. What is a validation set's purpose?

Ans

A validation set is used to compare models. It makes it possible to select the best model and tune the hyperparameters.

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

Ans 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

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

Ans

If you tune hyperparameters using the test set, you risk overfitting the test set, and the generalization error you measure will be optimistic (you may launch a model that performs worse than you expect).