

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

Ans: Machine learning is a subfield of artificial intelligence (AI) that focuses on enabling machines to learn from data and improve their performance on a task over time. In machine learning, algorithms are used to analyze and identify patterns in data, and these patterns are then used to make predictions or decisions about new data.

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

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

Ans: 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. What are the two most important tasks that are supervised?

Ans: Churn prediction, House price prediction

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

Ans: Clustering, Anomaly detection, Dimensionality reduction, Generative modeling

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

Ans: Reinforcement learning is a type of machine learning which allows the robot to learn from its experiences and adapt to different terrains, without the need for explicit programming of the walking movements.

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

Ans: K-means

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

Ans: Supervised

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

Ans: The online learning system is a type of machine learning system that learns incrementally from new incoming data over time, as opposed to being trained on a fixed dataset. In online learning, the machine learning model is updated continuously with new data as it becomes available, allowing the model to adapt and improve its predictions over time.

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

Ans: Out-of-core learning is a technique for training machine learning models on large datasets that cannot fit into the memory of a single machine. It involves processing the data in smaller chunks, typically using disk-based storage systems or distributed computing frameworks.

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

Ans: In nearest neighbor algorithms, predictions are made based on the similarity between the input data and the training data. The algorithm first identifies the k-nearest neighbors in the training data that are closest to the input data point based on a similarity measure, such as Euclidean distance or cosine similarity. Then, the algorithm makes a prediction based on the values of the k-nearest neighbors.

12.What's the difference between a model parameter and a hyperparameter in a learning 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: Model-based learning algorithms aim to build a model that captures the underlying patterns and relationships in the data, and use this model to make predictions on new data. There are several criteria that model-based learning algorithms look for:

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

Ans: Data quality, Overfitting and underfitting, Scalability and deployment

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 well on the training data but fails to generalize to new situations, it is likely overfitting to the training data. Overfitting occurs when a model is too complex and captures noise and patterns specific to the training data.

a)Reduce model complexity: If the model is too complex and is overfitting the training data, reducing the model complexity may help. This can be achieved by reducing the number of features used by the model, reducing the depth or width of the model, or using regularization techniques to constrain the model's weights.

b)Increase dataset size: If the model is overfitting because the training dataset is too small, collecting more data can help the model generalize better. More data can help the model capture a wider range of patterns and can reduce the risk of overfitting.

c)Use cross-validation: Cross-validation is a technique that involves partitioning the data into training and validation sets and training the model on the training set while evaluating its performance on the validation set. This can help identify overfitting and provide a more accurate estimate of the model's generalization performance.

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

Ans: A test set is a portion of the data that is held back from the model during training and is used to evaluate the performance of the model.

17.What is a validation set's purpose?

Ans: The purpose of a validation set in machine learning is to tune the hyperparameters of the model and to select the best model based on its performance on the validation set.

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

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