**Interview Questions on Machine Learning**

Q1) What are the basic differences between Machine Learning and Deep Learning

|  |  |  |
| --- | --- | --- |
| **Machine Learning Vs Deep Learning** | | |
|  | **Machine Learning** | **Deep Learning** |
| **Definition** | Sub-discipline of AI | Subset of machine learning |
| **Data** | Parses the data | Creates an artificial neural network |
| **Accuracy** | Requires manual intervention means decreased accuracy | Self-learning capabilities mean higher accuracy |
| **Interpretability** | Machine Learning is Faster | 10 Times Faster than ML |
| **Output** | ML models produce a numerical output | DL algorithms can range from an image to text or even an audio |
| **Data dependencies** | High | Low |
| **Hardware dependencies** | Can work on low-end machines. | Heavily depend on high-end machines |
| **Future** | Effective with image recognition and face recognition in mobiles | Not much effective due to data processing limitations |

**Q2) What is the difference between Bias and Variance?**

**Bias**:  
Bias can be defined as a situation where an error has occurred due to use of assumptions in the [learning algorithm](https://mindmajix.com/top-10-machine-learning-algorithms).

**Variance:**  
Variance is an error caused because of the complexity of the algorithm that is been used to analyze the data.

**Q3) What is the difference between supervised and unsupervised machine learning?**

A Supervised learning is a process where it requires training labeled data.  When it comes to Unsupervised learning it doesn’t require data labeling.

**Q4) How is KNN different from K-means clustering?**

KNN stands for K- Nearest Neighbours, it is classified as a supervised algorithm.  
K-means is an unsupervised cluster algorithm.

**Q5) Comparision between Machine Learning and Big Data**

|  |  |  |
| --- | --- | --- |
| **Machine Learning Vs Big Data** | | |
| **Feature** | **Machine Learning** | **Big Data** |
| Data Use | Technology that helps in reducing human intervention. | Data research, especially when working with huge data. |
| Operations | Existing data helps to tech machine what can be done further | Design patterns with analytics on existing data in terms of decision making. |
| Pattern Recognition | Similar to Big Data, existing data helps in pattern recognition. | Sequence and classification analysis helps in pattern recognition. |
| Data Volume | Best performance, while working with small-datasets. | Datasets helps in understanding and solving problems associated with large data volumes. |
| Application | Read existing data to predict future information. | Storing and analysing patterns within huge data volumes. |

**Q6) Explain what is precision and Recall?**

Subscribe to our youtube channel to get new updates..!

**Recall:**  
It is known as a true positive rate. The number of positives that your model has claimed compared to the actual defined number of positives available throughout the data.

**Precision:**  
It is also known as a positive predicted value. This is more based on the prediction. It is a measure of a number of accurate positives that the model claims when compared to the number of positives it actually claims.

**Q7) What is your favorite algorithm and also explain the algorithm in briefly in a minute?**

This type of questions is very common and asked by the interviewers to understand the candidate skills and assess how well he can communicate complex theories in the simplest language.

This one is a tough question and usually, individuals are not at all prepared for this situation so please be prepared and have a choice of algorithms and make sure you practice a lot before going into any sort of interviews.

**Q8) What is the difference between Type 1 and Type 2 errors?**

Type 1 error is classified as a false positive. I.e. This error claims that something has happened but the fact is nothing has happened. It is like a false fire alarm. The alarm rings but there is no fire.

Type 2 error is classified as a false negative. I.e. This error claims that nothing has happened but the fact is that actually, something happened at the instance.

The best way to differentiate a type 1 vs type 2 error is:  
Calling a man to be pregnant- This is Type 1 example  
Calling pregnant women and telling that she isn’t carrying any baby- This is type 2 example  
  
**Q9) Define what is Fourier Transform in a single sentence?**

A process of decomposing generic functions into a superposition of symmetric functions is considered to be a Fourier Transform.

**Q10) What is deep learning?**

[Deep learning](https://mindmajix.com/machine-learning-vs-deep-learning) is a process where it is considered to be a subset of machine learning process.

**Q11) What is the F1 score?**

The F1 score is defined as a measure of a model’s performance.

**Q12) How is F1 score is used?**

The average of Precision and Recall of a model is nothing but F1 score measure. Based on the results, the F1 score is 1 then it is classified as best and 0 being the worst.

**Q13) How can you ensure that you are not overfitting with a particular model?**

In Machine Learning concepts, they are three main methods or processes to avoid overfitting:

Firstly, keep the model simple  
Must and should use cross validation techniques  
It is mandatory to use regularization techniques, for example, LASSO.

**Q14) How to handle or missing data in a dataset?**

An individual can easily find missing or corrupted data in a data set either by dropping the rows or columns. On contrary, they can decide to replace the data with another value.

In Pandas they are two ways to identify the missing data, these two methods are very useful.  
isnull() and dropna().

**Q15) Do you have any relevant experience on Spark or any of big data tools that are used for Machine Learning?**

Well, this sort of question is tricky to answer and the best way to respond back is, to be honest. Make sure you are familiar with Big data is and the different tools that are available. If you know about Spark then it is always good to talk about it and if you are unsure then it is best, to be honest and let the interviewer know about it.

So for this, you have to prepare what is Spark and its good to prepare other available Big data tools that are used for Machine learning.

**Q16) Pick an algorithm and write a Pseudocode for the same?**

This question depicts your understanding of the algorithm. This is something that one has to be very creative and also should have in-depth knowledge about the algorithms and first and foremost the individual should have a good understanding of the algorithms. Best way to answer this question would be start off with Web Sequence Diagrams.

**Q17) What is the difference between an array and Linked list?**

An array is an ordered fashion of collection of objects.  
A linked list is a series of objects that are processed in a sequential order.

**Q18): Define a hash table?**

They are generally used for database indexing.  
A hash table is nothing but a data structure that produces an associative array.

**Q19) Mention any one of the data visualization tools that you are familiar with?**

This is another question where one has to be completely honest and also giving out your personal experience with these type of tools are really important. Some of the data visualization tools are Tableau, Plot.ly, and matplotlib.

**Q20) What is your opinion on our current data process?**

This type of questions are asked and the individuals have to carefully listen to their use case and at the same time, the reply should be in a constructive and insightful manner. Based on your responses, the interviewer will have a chance to review and understand whether you are a value add to their team or not.

**Q21) Please let us know what was your last read book or learning paper on Machine Learning?**

This type of question is asked to see whether the individual has a keen interest towards learning and also he is up to the latest market standards. This is something that every candidate should be looking out for and it is vital for individuals to read through the latest publishings.

**Q22) What is your favorite use case for machine learning models?**

The decision tree is one of my favorite use case for machine learning models.

**Q23) Is rotation necessary in PCA?**

Yes, the rotation is definitely necessary because it maximizes the differences between the variance captured by the components.

**Q24) What happens if the components are not rotated in PCA?**

It is a straight effect. If the components are not rotated then it will diminish eventually and one has to use a lot of various components to explain the data set variance.

**Q25) Explain why Navie Bayes is so Naive?**

It is based on an assumption that all of the features in the data set are important, equal and independent.

**Q25) How Recall and True positive rate are related?**

The relation is  
True Positive Rate = Recall.

**Q27) Assume that you are working on a data set, explain how would you select important variables?**

The following are few methods can be used to select important variables:

1. Use of Lasso Regression method.
2. Using Random Forest, plot variable importance chart.
3. Using Linear regression.

**Q28) Explain how we can capture the correlation between continuous and categorical variable?**

Yes, it is possible by using ANCOVA technique. It stands for Analysis of Covariance.

It is used to calculate the association between continuous and categorical variables.

**Q29) Explain the concept of machine learning and assume that you are explaining this to a 5-year-old baby?**

Yes, the question itself is the answer.

Machine learning is exactly the same way how babies do their day to day activities, the way they walk or sleep etc. It is a common fact that babies cannot walk straight away and they fall and then they get up again and then try. This is the same thing when it comes to machine learning, it is all about how the algorithm is working and at the same time redefining every time to make sure the end result is as perfect as possible.

One has to take real time examples while explaining these questions.

**Q30) What is the difference between Machine learning and Data Mining?**

Data mining is about working on unstructured data and then extract it to a level where the interesting and unknown patterns are identified.

Machine learning is a process or a study whether it closely relates to design, development of the algorithms that provide an ability to the machines to capacity to learn.

**Q31) What is inductive machine learning?**

Inductive [machine learning is all about a process of learning by live examples](https://mindmajix.com/machine-learning-examples).

**Q32) Please state few popular Machine Learning algorithms?**

1. Nearest Neighbour
2. Neural Networks
3. Decision Trees etc
4. Support vector machines

**Q33) What are the different types of algorithm techniques are available in machine learning?**

Some of them are :

1. Supervised learning
2. Unsupervised learning
3. Semi-supervised learning
4. Transduction
5. Learning to learn

**Q34) What are the three stages to build the model in machine learning:**

1. Model building
2. Model testing
3. Applying the model