MACHINE LEARNING INTERVIEW QUESTIONS AND ANSWERS

1. What is machine learning?

   In answering this question, try to show you understand of the broad applications of machine learning, as well as how it fits into AI. Put it into your own words, but convey your understanding that machine learning is a form of AI that automates data analysis to enable computers to learn and adapt through experience to do specific tasks without explicit programming

2. What is candidate sampling in machine learning?

   A training-time optimization in which a probability is calculated for all the positive labels, using, for example, softmax, but only for a random sample of negative labels. For example, if we have an example labeled beagle and dog candidate sampling computes the predicted probabilities and corresponding loss terms for the beagle and dog class outputs in addition to a random subset of the remaining classes (cat, lollipop, fence).

3. Mention the difference between Data Mining and Machine learning?

   Machine learning relates to the study, design, and development of the algorithms that give computers the capability to learn without being explicitly programmed. While data mining can be defined as the process by which the unstructured data tries to extract knowledge or unknown interesting patterns. During this processing machine, learning algorithms are used.

4. What is A/B testing in Machine Learning?

   A statistical way of comparing two (or more) techniques, typically an incumbent against a new rival. A/B testing aims to determine not only which technique performs better but also to understand whether the difference is statistically significant. A/B testing usually considers only two techniques using one measurement, but it can be applied to any finite number of techniques and measures.

5. 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.

6. What is 'Overfitting' in Machine learning?

In machine learning, when a statistical model describes random error or noise instead of the underlying relationship 'overfitting' occurs. When a model is excessively complex, overfitting is normally observed, because of having too many parameters with respect to the number of training data types. The model exhibits poor performance which has been overfitted.

7. Why does overfitting happen?

The possibility of overfitting exists as the criteria used for training the model is not the same as the criteria used to judge the efficacy of a model.

8. How can you avoid overfitting?

By using a lot of data overfitting can be avoided, overfitting happens relatively as you have a small dataset, and you try to learn from it. But if you have a small database and you are forced to come with a model based on that. In such a situation, you can use a technique known as cross-validation. In this method the dataset splits into two section, testing and training datasets, the testing dataset will only test the model while, in the training dataset, the data points will come up with the model.

In this technique, a model is usually given a dataset of a known data on which training (training data set) is run and a dataset of unknown data against which the model is tested. The idea of cross-validation is to define a dataset to "test" the model in the training phase.

9. Explain Principal Component Analysis (PCA).

PCA is a dimensionality-reduction technique which mathematically transforms a set of correlated variables into a smaller set of uncorrelated variables called principal components.

10. What value do you optimize when using a support vector machine (SVM)?

For a linear function, SVM optimizes the product of input vectors as well as the

coefficients. In other words, the algorithm with the linear function can be restructured into a dot-product.

11. What is inductive machine learning?

The inductive machine learning involves the process of learning by examples, where a system, from a set of observed instances, tries to induce a general rule.

12. What is the activation function in Machine Learning?

A function (for example, ReLU or sigmoid) that takes in the weighted sum of all of the inputs from the previous layer and then generates and passes an output value (typically nonlinear) to the next layer.

13. How does deductive and inductive machine learning differ?

Deductive machine learning starts with a conclusion, then learns by deducing what is right or wrong about that conclusion. Inductive machine learning starts with examples from which to draw conclusions.

14. How do you choose an algorithm for a classification problem?

The answer depends on the degree of accuracy needed and the size of the training set. If you have a small training set, you can use a low variance/high bias classifier. If your training set is large, you will want to choose a high variance/low bias classifier.

15. How do bias and variance play out in machine learning?

Both bias and variance are errors. Bias is an error due to flawed assumptions in the learning algorithm. Variance is an error resulting from too much complexity in the learning algorithm.

16. What is a class in machine learning?

One of a set of enumerated target values for a label. For example, in a binary classification model that detects spam, the two classes are spam and not spam. In a multi-class classification model that identifies dog breeds, the classes would be a poodle, beagle, pug, and so on.

17. What is the baseline for machine learning?

A simple model or heuristic used as reference point for comparing how well a model

is performing. A baseline helps model developers quantify the minimal, expected performance on a particular problem.

18. 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.

19. What is a checkpoint in machine learning?

Data that captures the state of the variables of a model at a particular time. Checkpoints enable exporting model weights, as well as performing training across multiple sessions. Checkpoints also enable training to continue past errors (for example, job preemption). Note that the graph itself is not included in a checkpoint.

20. 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().

21. What is bucketing in machine learning?

Converting a (usually continuous) feature into multiple binary features called buckets or bins, typically based on value range. For example, instead of representing temperature as a single continuous floating-point feature, you could chop ranges of temperatures into discrete bins. Given temperature data
sensitive to a tenth of a degree, all temperatures between 0.0 and 15.0 degrees could be put into one bin, 15.1 to 30.0 degrees could be the second bin, and 30.1 to 50.0 degrees could be a third bin.

22. What are some methods of reducing dimensionality?

You can reduce dimensionality by combining features with feature engineering, removing collinear features, or using algorithmic dimensionality reduction.

23. How do classification and regression differ?

Classification predicts group or class membership. Regression involves predicting a response. Classification is the better technique when you need a more definite answer.

24. What is supervised versus unsupervised learning?

Supervised learning is a process of machine learning in which outputs are fed back into a computer for the software to learn from for more accurate results the next time. With supervised learning, the "machine" receives initial training to start. In contrast, unsupervised learning means a computer will learn without initial training.

25. Define A HashTable?

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

26. What is the bias in machine learning?

An interceptor offset from an origin. Bias (also known as the bias term) is referred to as b or w0 in machine learning models.

27. What is the use of gradient descent?

The use of gradient descent plainly lies with the fact that it is easy to implement and is compatible with most of the ML algorithms when it comes to optimization. This technique works on the principle of a cost function.

28. What is backpropagation in machine learning?

The primary algorithm for performing gradient descent on neural networks. First, the output values of each node are calculated (and cached) in a forward pass. Then, the partial derivative of the error with respect to each parameter is calculated in a backward pass through the graph.

The Area Under the ROC curve is the probability that a classifier will be more confident that a randomly chosen positive example is actually positive than that a randomly chosen negative example is positive.

29. What is a sigmoid function in Machine learning?

    A function that maps logistic or multinomial regression output (log odds) to probabilities, returning a value between 0 and 1.

30. Explain The Concept Of Machine Learning And Assume That You Are Explaining This To A 5-year-old Baby?

    Yes, 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.

31.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.

32.What is the calibration layer in machine learning?

A post-prediction adjustment, typically to account for prediction bias. The adjusted predictions and probabilities should match the distribution of an observed set of labels.

33.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.

34.What is an AdaGrad algorithm in machine learning?

A sophisticated gradient descent algorithm that rescales the gradients of each parameter, effectively giving each parameter an independent learning rate.

35.Please, State Few Popular Machine Learning Algorithms?

Nearest Neighbour

Neural Networks

Decision Trees etc

Support vector machines

36.What Is The Difference Between Bias And Variance?

Bias: Bias can be defined as a situation where an error has occurred due to the use of assumptions in the learning algorithm.

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

37.What is a binary classification in machine learning?

A type of classification task that outputs one of two mutually exclusive classes. For example, a machine learning model that evaluates email messages and outputs either "spam" or "not spam" is a binary classifier.

38.What is the difference between supervised and unsupervised machine learning?

Supervised learning requires training labeled data. For example, in order to do classification (a supervised learning task), you'll need to first label the data you'll use to train the model to classify data into your labeled groups. Unsupervised learning, in contrast, does not require labeling data explicitly.

39.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.

40.What Is Deep Learning?

Deep learning is a process where it is considered to be a subset of the machine learning process.

1. What is Rectified Linear Unit (ReLU) in Machine learning?

    An activation function with the following rules:

    (a). If the input is negative or zero, the output is 0.

    (b). If the input is positive, the output is equal to input.

2. What is batch in machine learning?

    The set of examples used in one iteration (that is, one gradient update) of model training.

3. What is batch size machine learning?

    The number of examples in a batch. For example, the batch size of SGD is 1, while the batch size of a mini-batch is usually between 10 and 1000. Batch size is usually fixed during training and inference.

4. What is a class-imbalanced dataset in machine learning?

    A binary classification problem in which the labels for the two classes have significantly different frequencies. For example, a disease data set in which 0.0001 of examples have positive labels and 0.9999 have negative labels is a class-imbalanced problem, but a football game predictor in which 0.51 of examples label one team winning and 0.49 label the other team winning is not a class- imbalanced problem.

5. What is the classification model in machine learning?

    A type of machine learning model for distinguishing between two or more discrete classes. For example, a natural language processing classification model could determine whether an input sentence was in French, Spanish, or Italian. Compare with the regression model.

6. What is the classification threshold in machine learning?

   A scalar-value criterion that is applied to a model's predicted score in order to separate the positive class from the negative class. Used when mapping logistic regression results to binary classification.

7. What is collaborative filtering in machine learning?

   Making predictions about the interests of one user based on the interests of many other users. Collaborative filtering is often used in recommendation systems.

8. What is the confusion matrix in machine learning?

   An NxN table that summarizes how successful a classification model's predictions were; that is, the correlation between the label and the model's classification. One axis of a confusion matrix is the label that the model predicted, and the other axis is the actual label. N represents the number of classes.

9. What is the convex function in machine learning?

   A function in which the region above the graph of the function is a convex set. The prototypical convex function is shaped something like the letter U.
   A strictly convex function has exactly one local minimum point, which is also the global minimum point. The classic U-shaped functions are strictly convex functions. However, some convex functions (for example, straight lines) are not.

10. Which is better for image classification? Supervised or unsupervised classification. Justify.1

    In a supervised classification, the images are interpreted manually by the ML expert to create feature classes whereas this is not the case in unsupervised classification wherein the ML software creates feature classes based on image pixel values. Therefore, it is better to opt for supervised classification for image classification in terms of accuracy.

11. What Are The Three Stages To Build The Model In Machine Learning?

    (a). Model building
    (b). Model testing
    (c). Applying the model

12. What is convergence in machine learning?

Informally, often refers to a state reached during training in which training loss and validation loss change very little or not at all with each iteration after a certain number of iterations.

In other words, a model reaches convergence when additional training on the current data will not improve the model. In deep learning, loss values sometimes stay constant or nearly so for many iterations before finally descending, temporarily producing a false sense of convergence.

13. Explain the difference between L1 and L2 regularization.

L2 regularization tends to spread error among all the terms, while L1 is more binary/sparse, with many variables either being assigned a 1 or 0 in weighting. L1 corresponds to setting a Laplace a prior on the terms, while L2 corresponds to a Gaussian prior.

14. What's your favorite algorithm, and can you explain it to me in less than a minute?

This type of question tests your understanding of how to communicate complex and technical nuances with poise and the ability to summarize quickly and efficiently. Make sure you have a choice and make sure you can explain different algorithms so simply and effectively that a five-year-old could grasp the basics!

15. How is ML different from artificial intelligence?

AI involves machines that execute tasks which are programmed and based on human intelligence, whereas ML is a subset application of AI where machines are made to learn information. They gradually perform tasks and can automatically build models from the learnings.

16. Differentiate between statistics and ML.

In statistics, the relationships between relevant data (variables) are established; but in ML, the algorithms rely on data regardless of their statistical influence. In other words, statistics are concerned about inferences in the data whereas ML looks at optimization.

17. What are neural networks and where do they find their application in ML? Elaborate.

Neural networks are information processing models that derive their functions based on biological neurons found in the human brain. The reason they are the choice of technique in ML is that they help discover patterns in data that are sometimes too complex to comprehend by humans.

18. Differentiate between a parameter and a hyperparameter?

Parameters are attributes in training data that can be estimated during ML. Hyperparameters are attributes that cannot be determined beforehand in the training data. Example: Learning rate in neural networks.

19. What is 'tuning' in ML?

Generally, the goal of ML is to automatically provide accurate output from the vast amounts of input data without human intervention. Tuning is a process which makes this possible and it involves optimizing hyperparameters for an algorithm or an ML model to make them perform correctly.

20. What is optimization in ML?

Optimisation, in general, refers to minimizing or maximizing an objective function (in linear programming). In the context of ML, optimization refers to the tuning of hyperparameters which result in minimizing the error function (or loss function).

21. What is dimensionality reduction? Explain in detail.

The process of reducing variables in an ML classification scenario is called Dimensionality reduction. The process is segregated into sub-processes called feature extraction and feature selection. Dimensionality reduction is done to enhance visualization of training data. It finds the appropriate set of variables known as principal variables.

22. Explain Principal Component Analysis (PCA).

PCA is a dimensionality-reduction technique which mathematically transforms a set of correlated variables into a smaller set of uncorrelated variables called principal components.

23. What value do you optimize when using a support vector machine (SVM)?

For a linear function, SVM optimizes the product of input vectors as well as the coefficients. In other words, the algorithm with the linear function can be restructured into a dot-product.

24. On what basis do you choose a classifier?

Classifiers must be chosen based on the accuracy it provides on the trained data. Also, the size of the dataset sometimes affects accuracy. For example, Naive Bayes classifiers suit smaller datasets in terms of accuracy due to higher asymptotic errors.

25. Mention key business metrics that help ML?

Identify the key services/products/functions that hold good for ML. For example, if you consider a commercial bank, metrics such as a number of new accounts, type of accounts, leads generated and so on, can be evaluated through ML methods.

26. What is kernel SVM?

Kernel SVM is the abbreviated version of kernel support vector machine. Kernel methods are a class of algorithms for pattern analysis and the most common one is the kernel SVM.

27. What is the decision tree classification?

A decision tree builds classification (or regression) models as a tree structure, with datasets broken up into ever smaller subsets while developing the decision tree, literally in a tree-like way with branches and nodes. Decision trees can handle both categorical and numerical data.

28. What is a recommendation system?

Anyone who has used Spotify or shopped at Amazon will recognize a recommendation system: It's an information filtering system that predicts what a user might want to hear or see based on choice patterns provided by the user.

29. What are the five popular algorithms of Machine Learning?

Decision Trees

Neural Networks (back propagation)

Probabilistic networks

Nearest Neighbor

Support vector machines

30. What are the different Algorithm techniques in Machine Learning?

The different types of techniques in Machine Learning are

Supervised Learning

Unsupervised Learning

Semi-supervised Learning

Reinforcement Learning

Transduction

Learning to Learn

31. What are the three stages to build the hypotheses or model in machine learning?

Model building

Model testing

Applying the model

32. What is the standard approach to supervised learning?

The standard approach to supervised learning is to split the set of example into the training set and the test.

33. What is 'Training set' and 'Test set'?

In various areas of information science like machine learning, a set of data is used to discover the potentially predictive relationship known as 'Training Set'. The training set is an example given to the learner, while Test set is used to test the accuracy of the hypotheses generated by the learner, and it is the set of example held back from the learner. The training set is distinct from the Test set.

34. List down various approaches to machine learning?

The different approaches in Machine Learning are

i) Concept Vs Classification Learning

ii) Symbolic Vs Statistical Learning

iii) Inductive Vs Analytical Learning

1. What is not Machine Learning?

   i) Artificial Intelligence

   ii) Rule-based inference

2. Explain what is the function of 'Unsupervised Learning'?

   Find clusters of the data

   Find low-dimensional representations of the data

   Find interesting directions in data

   Interesting coordinates and correlations

   Find novel observations/ database cleaning

3. Explain what is the function of 'Supervised Learning'?

   Classifications

   Speech recognition

   Regression

   Predict time series

   Annotate strings

4. What is algorithm independent machine learning?

   Machine learning in where mathematical foundations are independent of any particular classifier or learning algorithm is referred to as algorithm independent machine learning?

5. What is the difference between artificial learning and machine learning?

   Designing and developing algorithms according to the behaviors based on empirical data are known as Machine Learning. While artificial intelligence in addition to machine learning, it also covers other aspects like knowledge representation, natural language processing, planning, robotics etc.

6. What is the classifier in machine learning?

   A classifier in a Machine Learning is a system that inputs a vector of discrete or continuous feature values and outputs a single discrete value, the class.

7. What are the advantages of Naive Bayes?

   In a Naïve Bayes classifier will converge quicker than discriminative models like

logistic regression, so you need less training data. The main advantage is that it can't learn interactions between features.

8. In what areas Pattern Recognition is used?

   Pattern Recognition can be used in

   Computer Vision

   Speech Recognition

   Data Mining

   Statistics

   Informal Retrieval

   Bio-Informatics

9. How would you approach the "Netflix Prize" competition?

   The Netflix Prize was a famed competition where Netflix offered $1,000,000 for a better collaborative filtering algorithm. The team that won called BellKor had a 10% improvement and used an ensemble of different methods to win. Some familiarity with the case and its solution will help demonstrate you've paid attention to machine learning for a while.

10. What are the different methods of Sequential Supervised Learning?

    The different methods to solve Sequential Supervised Learning problems are

    a)Sliding-window methods

    b)Recurrent sliding windows

    c)Hidden Markow models

    d)Maximum entropy Markow models

    e)Conditional random fields

    f)Graph transformer networks

85.What are the areas in robotics and information processing where the sequential prediction problem arises?

The areas in robotics and information processing where the sequential prediction problem arises are

a)Imitation Learning

b)Structured prediction

c)Model-based reinforcement learning

1. What is batch statistical learning?

   Statistical learning techniques allow learning a function or predictor from a set of observed data that can make predictions about unseen or future data. These techniques provide guarantees on the performance of the learned predictor on the future unseen data based on a statistical assumption on the data generating process.

2. What is PAC Learning?

   PAC (Probably Approximately Correct) learning is a learning framework that has been introduced to analyze learning algorithms and their statistical efficiency.

3. What are the different categories you can categorize the sequence learning process?

   a)Sequence prediction

   b)Sequence generation

   c)Sequence recognition

   d)Sequential decision

89.What is sequence learning?

Sequence learning is a method of teaching and learning in a logical manner.

1. What are two techniques of Machine Learning?

   The two techniques of Machine Learning are

   a)Genetic Programming

   b)Inductive Learning

91.What is the bias-variance decomposition of classification error in the ensemble method?

The expected error of a learning algorithm can be decomposed into bias and variance. A bias term measures how closely the average classifier produced by the learning algorithm

matches the target function. The variance term measures how much the learning algorithm's prediction fluctuates for different training sets.

1. What is an Incremental Learning algorithm in the ensemble?

   The incremental learning method is the ability of an algorithm to learn from new data that may be available after the classifier has already been generated from the already available dataset.

2. What are PCA, KPCA, and ICA used for?

   PCA (Principal Components Analysis), KPCA (Kernel-based Principal Component Analysis) and ICA (Independent Component Analysis) are important feature extraction techniques used for dimensionality reduction.

3. What is dimension reduction in Machine Learning?

   In Machine Learning and statistics, dimension reduction is the process of reducing the number of random variables under considerations and can be divided into feature selection and feature extraction

4. What are support vector machines?

   Support vector machines are supervised learning algorithms used for classification and regression analysis.

5. What are the components of relational evaluation techniques?

   The important components of relational evaluation techniques are

   a)Data Acquisition

   b)Ground Truth Acquisition

   c)Cross-Validation Technique

   d)Query Type

   e)Scoring Metric

   f)Significance Test

6. What is ensemble learning?

   To solve a particular computational program, multiple models such as classifiers or experts are strategically generated and combined. This process is known as ensemble learning.

7. Why ensemble learning is used?

   Ensemble learning is used to improve the classification, prediction, function approximation etc of a model.

8. When to use ensemble learning?

   Ensemble learning is used when you build component classifiers that are more accurate and independent from each other.

9. What are the two paradigms of ensemble methods?

   The two paradigms of ensemble methods are

   a)Sequential ensemble methods

   b)Parallel ensemble methods