**1.Explain the term machine learning, and how does it work? Explain two machine learning**

**applications in the business world. What are some of the ethical concerns that machine learning**

**applications could raise?**

Machine learning is a field of artificial intelligence that allows systems to learn and improve from experience without being explicitly programmed

Credit card default detection

Spam classification

**Ethical issues**

Accuracy, Bias, Privacy, Accountability, Decision making

**2. Describe the process of human learning:**

**i. Under the supervision of experts**

Human-guided machine learning is a process whereby subject matter experts accelerate the learning process by teaching the technology in real-time. For example, if the machine learning model comes across a piece of data it is uncertain about, a human can be asked to weigh in and give feedback

**ii. With the assistance of experts in an indirect manner**

Indirect guidance is provided through learners actively observing, listening, and engaging with social practices and norms, which serve to furnish models and goals for performance and individuals' learning.

**iii. Self-education**

Self-directed learning is a process in which individuals take the initiative, with or without the help of others, in diagnosing their learning needs, formulating learning goals, identifying human and material resources for learning, choosing and implementing appropriate learning strategies, and evaluating learning

**3. Provide a few examples of various types of machine learning.**

Supervised,- Regression, classification etc

Unsupervised – clustering, PCA etc

Reinforcement learning

**4. Examine the various forms of machine learning.**

Supervised, Unsupervised, and Reinforcement Learning

**5. Can you explain what a well-posed learning problem is? Explain the main characteristics that must be present to identify a learning problem properly.**

A problem is well-posed if a solution to it exists, if that solution is unique, and if that solution depends on the data / experience but it is not sensitive to (reasonably small) changes in the data / experience. A well-defined learning problem will have the features like class of tasks, the measure of performance to be improved, and the source of experience examples

**6. Is machine learning capable of solving all problems? Give a detailed explanation of your answer.**

Machine learning, a subset of artificial intelligence, has revolutionalized the world as we know it in the past decade. The information explosion has resulted in the collection of massive amounts of data, especially by large companies such as Facebook and Google. This amount of data, coupled with the rapid development of processor power and computer parallelization, has now made it possible to obtain and study huge amounts of data with relative ease.

1. Reasoning Power

One area where ML has not mastered successfully is reasoning power, a distinctly human trait. Algorithms available today are mainly oriented towards specific use-cases and are narrowed down when it comes to applicability. They cannot think as to why a particular method is happening that way or ‘introspect’ their own outcomes.

2. Contextual Limitation

If we consider the area of natural language processing (NLP), text and speech information are the means to understand languages by NLP algorithms. They may learn letters, words, sentences or even the syntax, but where they fall back is the context of the language. Algorithms do not understand the context of the language used

3. Scalability

Although we see ML implementations being deployed on a significant basis, it all depends on data as well as its scalability. Data is growing at an enormous rate and has many forms which largely affects the scalability of an ML project. Algorithms cannot do much about this unless they are updated constantly for new changes to handle data. This is where ML regularly requires human intervention in terms of scalability and remains unsolved mostly.

**7. What are the various methods and technologies for solving machine learning problems? Any two of them should be defined in detail.**

**Regression**

Regression methods fall within the category of supervised ML. They help to predict or explain a particular numerical value based on a set of prior data, for example predicting the price of a property based on previous pricing data for similar properties.

The simplest method is linear regression where we use the mathematical equation of the line (y = m \* x + b) to model a data set. We train a linear regression model with many data pairs (x, y) by calculating the position and slope of a line that minimizes the total distance between all of the data points and the line. In other words, we calculate the slope (m) and the y-intercept (b) for a line that best approximates the observations in the data.

**Classification**

Another class of supervised ML, classification methods predict or explain a class value. For example, they can help predict whether or not an online customer will buy a product. The output can be yes or no: buyer or not buyer. But classification methods aren’t limited to two classes. For example, a classification method could help to assess whether a given image contains a car or a truck. In this case, the output will be 3 different values: 1) the image contains a car, 2) the image contains a truck, or 3) the image contains neither a car nor a truck.

The simplest classification algorithm is logistic regression — which makes it sounds like a regression method, but it’s not. Logistic regression estimates the probability of an occurrence of an event based on one or more inputs.

**8. Can you explain the various forms of supervised learning? Explain each one with an example**

**application.**

**1. Regression**

Regression is used to understand the relationship between dependable and independent variables. Moreover, it is a type of supervised learning that learns from labelled data sets to predict continuous output for different data in an algorithm. It is believed to be widely used in scenarios where the output needs to be a finite value, for instance, height or weight, etc.

There two types of regression; they are as follows:

**Linear regression**

It is used to identify the relationship between two variables, typically used for making future predictions. Moreover, linear regression is sub-divided based on the number of independent and dependent variables.

For instance, if there is one independent and one dependent variable, it is known as simple linear regression. Meanwhile, if there are two or more independent and dependent variables, it is called multiple linear regression.

**Logistic regression**

Logistic regression is used when the dependent variable is categorical or has binary outputs like ‘yes’ or ‘no’. Moreover, logistic regression is used to solve binary classification problems; that’s why it predicts discreet values for variables.

**2. Naive Bayes**

A Naive Bayes algorithm is used for large datasets. The approach works on the fundamental that every programme in the algorithm works independently. This means that the presence of one feature will not impact the other. Generally, it is used in text classification, recommendation systems, and others.

There are different types of Naive Bayes models, and the decision tree remains the most popular among business organizations. A decision tree is a unique supervised learning algorithm structurally resembling a flowchart. However, they fundamentally perform different roles and responsibilities.

A decision tree consists of control statements containing decisions and their consequences. The output in a decision tree relates to the labelling of unforeseen data. ID3 and CART are some of the popular decision tree algorithms widely used across various industries.

**3. Classification**

It is a type of supervised learning algorithm that accurately assigns data into different categories or classes. It recognizes specific entities and analyses them to conclude where those entities must be categorized. Some of the classification algorithms are as follows:

K-nearest neighbor

Random forest

Support vector machines

Decision tree

Linear classifiers

**4. Neutral networks**

This type of supervised learning algorithm is used to group or categorize raw data. In addition, it is used for finding a pattern or interpreting sensory data. However, the algorithm requires numerous amounts of computation resources. As a result, its uses are constrained**.**

**5. Random forest**

A random forest algorithm is often called an ensemble method because it combines different supervised learning methods to conclude. Moreover, it uses many decision trees to output the classification of individual trees. As a result, it is widely used across industries.

**9. What is the difference between supervised and unsupervised learning? With a sample application in each region, explain the differences.**

|  |  |
| --- | --- |
| **Supervised Learning** | **Unsupervised Learning** |
| Supervised learning algorithms are trained using labeled data. | Unsupervised learning algorithms are trained using unlabeled data. |
| Supervised learning model takes direct feedback to check if it is predicting correct output or not. | Unsupervised learning model does not take any feedback. |
| Supervised learning model predicts the output. | Unsupervised learning model finds the hidden patterns in data. |
| In supervised learning, input data is provided to the model along with the output. | In unsupervised learning, only input data is provided to the model. |
| The goal of supervised learning is to train the model so that it can predict the output when it is given new data. | The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset. |
| Supervised learning needs supervision to train the model. | Unsupervised learning does not need any supervision to train the model. |
| Supervised learning can be categorized in Classification and Regression problems. | Unsupervised Learning can be classified in Clustering and Associations problems. |
| Supervised learning can be used for those cases where we know the input as well as corresponding outputs. | Unsupervised learning can be used for those cases where we have only input data and no corresponding output data. |
| Supervised learning model produces an accurate result. | Unsupervised learning model may give less accurate result as compared to supervised learning. |
| Supervised learning is not close to true Artificial intelligence as in this, we first train the model for each data, and then only it can predict the correct output. | Unsupervised learning is more close to the true Artificial Intelligence as it learns similarly as a child learns daily routine things by his experiences. |
| It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc. | It includes various algorithms such as Clustering, KNN, and Apriori algorithm. |

**10. Describe the machine learning process in depth.**

**a. Make brief notes on any two of the following:**

**MATLAB is one of the most widely used programming languages.**

MATLAB is a programming platform designed specifically for engineers and scientists to analyze and design systems and products that would transform our world. The MATLAB language is a matrix-based language allowing the most natural expression of computational mathematics. The programming language enables its users to take their ideas from research to production by deploying them to enterprise applications and embedded devices, as well as integrating model-based design. The language is designed in a way that will help the programmers think and work the way they feel like so that it is available and can be accessed by a novice and an expert, equally. Millions of engineers, scientists, and students around the world use MATLAB for a range of applications, both in industry and academia. They also use other technologies like deep learning, machine learning, signal processing, and communications, image and video processing, control systems, etc. Now in technical terms, they deal with MATLAB-enabled matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages.

**ii. Deep learning applications in healthcare**

**Drug Discovery**

The role of deep learning in identifying drug combinations is significant. During the pandemic, vaccine and drug development were funded by disruptive technologies like AI, machine learning, and deep learning. Since drug discovery is a complex task, deep learning can make it faster, cost-effective, and easier. Deep learning algorithms can predict the drug properties, drug-target interaction prediction, and in generating a compound with desired properties

**Medical Imaging and Diagnostics**

Deep learning models can interpret medical images like X-ray, MRI scan, CT scan, etc., to perform diagnosis. The algorithms can detect any risk and flag anomalies in the medical images. Deep learning is extensively used in detecting cancer. The recent innovation of computer vision was enabled by machine learning and deep learning. With a faster diagnosis through medical imaging, it becomes easier to treat diseases.

**Simplifying Clinical Trials**

Clinical trials are complicated and expensive. Machine learning and deep learning can be leveraged to perform predictive analytics to identify potential candidates for clinical trials and enable scientists to pool in people from different data points and sources. Deep learning will also enable continuous monitoring of these trials with minimum errors and human intervention.

**Personalized Treatment**

With deep learning models, it becomes easier to analyze patient’s health data, medical history, vital symptoms, medical test results, and others. Hence, this enables healthcare providers to understand each patient and provide personalized treatment for them. These disruptive technologies enable the detection of suitable and multiple treatment options for different patients. With real-time data collection through connected devices, machine learning models can use deep neural networks to predict upcoming health conditions or risks and provide specific medicines or treatments.

iii. Study of the market basket

A data mining technique that is used to uncover purchase patterns in any retail setting is known as Market Basket Analysis. In simple terms Basically, Market basket analysis in data mining is to analyze the combination of products which been bought together.

This is a technique that gives the careful study of purchases done by a customer in a supermarket. This concept identifies the pattern of frequent purchase items by customers. This analysis can help to promote deals, offers, sale by the companies, and data mining techniques helps to achieve this analysis task.

**iv. Linear regression (simple)**

Simple linear regression is used to estimate the relationship between two quantitative variables.

Simple Linear Regression is a type of Regression algorithms that models the relationship between a dependent variable and a single independent variable. The relationship shown by a Simple Linear Regression model is linear or a sloped straight line, hence it is called Simple Linear Regression.

The key point in Simple Linear Regression is that the dependent variable must be a continuous/real value. However, the independent variable can be measured on continuous or categorical values.

Simple Linear regression algorithm has mainly two objectives:

Model the relationship between the two variables. Such as the relationship between Income and expenditure, experience and Salary, etc.

Forecasting new observations. Such as Weather forecasting according to temperature, Revenue of a company according to the investments in a year, etc.

**11. Make a comparison between:-**

**1. Generalization and abstraction**

The term generalization describes the process of turning abstracted knowledge into a form that can be utilized for action. Generalization is a somewhat vague process that is a bit difficult to describe. Traditionally, it has been imagined as a search through the entire set of models (that is, theories) that could have been abstracted during training. Specifically, if you imagine a hypothetical set containing every possible theory that could be established from the data, generalization involves the reduction of this set into a manageable number of important findings.

Abstraction separates ideas from specific instances of those ideas. An abstract method is declared, but contains no implementation. Abstract classes cannot be instantiated, and require subclasses to provide implementations for the abstract methods.

**2. Learning that is guided and unsupervised**

The main distinction between the two approaches is the use of labeled datasets. To put it simply, supervised learning uses labeled input and output data, while an unsupervised learning algorithm does not.

In supervised learning, the algorithm “learns” from the training dataset by iteratively making predictions on the data and adjusting for the correct answer. While supervised learning models tend to be more accurate than unsupervised learning models, they require upfront human intervention to label the data appropriately. For example, a supervised learning model can predict how long your commute will be based on the time of day, weather conditions and so on. But first, you’ll have to train it to know that rainy weather extends the driving time.

Unsupervised learning models, in contrast, work on their own to discover the inherent structure of unlabeled data. Note that they still require some human intervention for validating output variables. For example, an unsupervised learning model can identify that online shoppers often purchase groups of products at the same time. However, a data analyst would need to validate that it makes sense for a recommendation engine to group baby clothes with an order of diapers, applesauce and sippy cups.

**3. Regression and classification**

Regression and Classification algorithms are Supervised Learning algorithms. Both the algorithms are used for prediction in Machine learning and work with the labeled datasets. But the difference between both is how they are used for different machine learning problems.

The main difference between Regression and Classification algorithms that Regression algorithms are used to predict the continuous values such as price, salary, age, etc. and Classification algorithms are used to predict/Classify the discrete values such as Male or Female, True or False, Spam or Not Spam, etc.

**Classification:**

Classification is a process of finding a function which helps in dividing the dataset into classes based on different parameters. In Classification, a computer program is trained on the training dataset and based on that training, it categorizes the data into different classes.

Example: The best example to understand the Classification problem is Email Spam Detection. The model is trained on the basis of millions of emails on different parameters, and whenever it receives a new email, it identifies whether the email is spam or not. If the email is spam, then it is moved to the Spam folder.

**Regression:**

Regression is a process of finding the correlations between dependent and independent variables. It helps in predicting the continuous variables such as prediction of Market Trends, prediction of House prices, etc.

The task of the Regression algorithm is to find the mapping function to map the input variable(x) to the continuous output variable(y).

Example: Suppose we want to do weather forecasting, so for this, we will use the Regression algorithm. In weather prediction, the model is trained on the past data, and once the training is completed, it can easily predict the weather for future days.