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?

Answers: Machine learning is a form of artificial intelligence (AI) that teaches computers to think in a similar way to how humans do: Learning and improving upon past experiences. It works by exploring data and identifying patterns and involves minimal human intervention. the most common machine learning applications for business: customer lifetime value modeling, churn modeling, dynamic pricing, customer segmentation, image classification, and recommendation engines. When people can look at the source code for a program, they can see how it makes decisions. Another issue is the use of machine learning training data and possible biases. There have been several instances of racial and other biases making it into machine learning programs unintentionally.

2. Describe the process of human learning:

i. Under the supervision of experts

ii. With the assistance of experts in an indirect manner

iii. Self-education

Answers:

i. Subject matter experts accelerate the learning process by teaching the technology in real-time.

ii. Assistance from experts helps the learner to learn how to work on difficult tasks. Inherently motivating because it involves the transfer of responsibility, or control, for learning, from the teacher or more capable other to the learner. This transfer of control is motivating for the student as it acknowledges student mastery of the task, and hence the learner's developing efficacy.

iii. Assess readiness to learn. Students need various skills and attitudes towards learning for successful independent study. Set learning goals. Engage in the learning process. Evaluate learning.

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

Answers:

Machine Learning:

1. Supervised learning.
   1. Classification.
      1. Diagnostics
      2. Customer Retention
      3. Image Classification
      4. Identify Fraud detection
   2. Regression:
      1. Population growth prediction.
      2. Estimating life expectancy.
      3. Market forecasting
      4. Weather forecasting
      5. Advertising popularity prediction.
2. Unsupervised learning.
   1. Dimensionality reduction.
      1. Meaningful compression
      2. Big data Visualisation
      3. Structure Discovery
      4. Feature Elicitation
   2. Clustering
      1. Recommender systems
      2. Targeted Marketing
      3. Customer segmentation
3. Reinforcement Learning
   1. Game AI
   2. Skill Acquisition
   3. Learning tasks
   4. Robot navigation
   5. Real time decisions

4. Examine the various forms of machine learning.

Answers:

Different types of Machine Learning :-

* Supervised Learning.
* Unsupervised Learning.
* Reinforcement Learning.

Machine Learning Types

1. Supervised Learning :-

In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and output.

Two types of Supervised Learning :-

* Regression — Estimate continuous values (Real valued output)
* Classification — Identify a unique class (Discrete values, Boolean or Categories)

1.1 Regression :-

Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Regression can be used to estimate/ predict continuous values (Real valued output).

For example : Given a picture of a person, we have to predict the age on the basis of the given picture .

1.2 Classification :-

Classification means to group the output into a class. If the data set is discrete or categorical then it is a classification problem.

For example : Given data about the sizes of houses in the real estate market, making our output about whether the house “sells for more or less than the asking price” i.e. Classifying houses into two discrete categories.

2. Unsupervised Learning :-

It allows us to approach problems with little or no idea about what our results look like. We can derive structure from data where we don’t necessarily know the effect of the variables.

We can derive this structure by clustering the data based on relationships among the variables in the data.

2.1. Clustering :

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).

For example : Take a collection of 1,000,000 different genes, and find a way to automatically group these genes into groups that are somehow similar or related by different variables, such as lifespan, location, roles, and so on.

3. Reinforcement Learning :-

Reinforcement Learning is about taking suitable actions to maximize reward in a particular situation. It is employed by various software and machines to find the best possible behavior or path to take in a specific situation.

Reinforcement learning differs from the supervised learning in a way that in supervised learning the training data has the answer key with it, so the model is trained with the correct answer itself whereas in reinforcement learning, there is no answer and the reinforcement agent decides what to do in order to perform the given task. In the absence of training data set, it is bound to learn from its experience.

Applications of Machine Learning :-

* Virtual Personal Assistants.
* Predictions while commuting.
* Videos Surveillance.
* Social Media Services.
* Email Spam and Malware Filtering.
* Online Customer Support.
* Search Engine Result Refining.
* Product Recommendations.
* Online Fraud Detection.

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.

Answers:

Well Posed Learning Problem – A computer program is said to learn from experience E in context to some task T and some performance measure P, if its performance on T, as was measured by P, upgrades with experience E.

Any problem can be segregated as well-posed learning problem if it has three traits –

* Task
* Performance Measure
* Experience

Certain examples that efficiently defines the well-posed learning problem are –

1. To better filter emails as spam or not

* Task – Classifying emails as spam or not
* Performance Measure – The fraction of emails accurately classified as spam or not spam
* Experience – Observing you label emails as spam or not spam

2. A checkers learning problem

* Task – Playing checkers game
* Performance Measure – percent of games won against opposer
* Experience – playing implementation games against itself

3. Handwriting Recognition Problem

* Task – Acknowledging handwritten words within portrayal
* Performance Measure – percent of words accurately classified
* Experience – a directory of handwritten words with given classifications

4. A Robot Driving Problem

* Task – driving on public four-lane highways using sight scanners
* Performance Measure – average distance progressed before a fallacy
* Experience – order of images and steering instructions noted down while observing a human driver

5. Fruit Prediction Problem

* Task – forecasting different fruits for recognition
* Performance Measure – able to predict maximum variety of fruits
* Experience – training machine with the largest datasets of fruits images

6. Face Recognition Problem

* Task – predicting different types of faces
* Performance Measure – able to predict maximum types of faces
* Experience – training machine with maximum number of datasets of different face images

7. Automatic Translation of documents

* Task – translating one type of language used in a document to other language
* Performance Measure – able to convert one language to other efficiently
* Experience – training machine with a large dataset of different types of languages

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

Answers: ML Can’t Solve Everything. Here Are 5 Challenges That It Still Faces

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.
4. Regulatory Restriction For Data In ML:

ML usually need considerable amounts (in fact, massive) of data in stages such as training, cross-validation etc. Sometimes, data includes private as well as general information. This is where it gets complicated. Most tech companies have privatised data and these data are the ones which are actually useful for ML applications. But, there comes the risk of the wrong usage of data, especially in critical areas such as medical research, health insurance etc.,

1. Internal Working Of Deep Learning:

This sub-field of ML is actually responsible for today’s AI growth. What was once just a theory has appeared to be the most powerful aspect of ML. Deep Learning (DL) now powers applications such as voice recognition, image recognition and so on through artificial neural networks. But, the internal working of DL is still unknown and yet to be solved. Advanced DL algorithms still baffle researchers in terms of its working and efficiency. Millions of neurons that form the neural networks in DL increase abstraction at every level, which cannot be comprehended at all. This is why deep learning is dubbed a ‘black box’ since its internal agenda is unknown.

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

Answers:

Machine Learning Methods:

1. Regression.
2. Classification.
3. Clustering.
4. Dimensionality Reduction.
5. Ensemble Methods.
6. Neural Nets and Deep Learning.
7. Transfer Learning.
8. Reinforcement Learning.

Machine Learning technologies:

1. TensorFlow.
2. Keras
3. Scikit-learn
4. Microsoft Cognitive Toolkit
5. Theano
6. Caffe
7. Torch
8. Accord.NET

**Keras** is an open-source software library designed to simplify the creation of deep learning models. It is written in Python and can be deployed on top of other AI technologies such as TensorFlow, Microsoft Cognitive Toolkit (CNTK), and Theano. Keras is known for its user-friendliness, modularity, and ease of extensibility. It is suitable if you need a machine learning library that allows for easy and fast prototyping, supports both convolutional and recurrent networks, and runs optimally on both CPUs (central processing units) and GPUs (graphics processing units).

**Scikit-learn** is an open-source library developed for machine learning. This traditional framework is written in Python and features several machine learning models including classification, regression, clustering, and dimensionality reduction. Scikit-learn is designed on three other open-source projects—Matplotlib, NumPy, and SciPy—and it focuses on data mining and data analysis.

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

Answers:

Supervised Learning: -

In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and output.

Two types of Supervised Learning: -

* Regression — Estimate continuous values (Real valued output)
* Classification — Identify a unique class (Discrete values, Boolean or Categories)

1.1 Regression:-

Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Regression can be used to estimate/ predict continuous values (Real valued output).

For example: Given a picture of a person, we have to predict the age on the basis of the given picture .

1.2 Classification:-

Classification means to group the output into a class. If the data set is discrete or categorical then it is a classification problem.

For example: Given data about the sizes of houses in the real estate market, making our output about whether the house “sells for more or less than the asking price” i.e. Classifying houses into two discrete categories.

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

Answers:

The main difference between supervised and unsupervised learning: Labeled data:

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.

Other key differences between supervised and unsupervised learning:

Goals: In supervised learning, the goal is to predict outcomes for new data. You know up front the type of results to expect. With an unsupervised learning algorithm, the goal is to get insights from large volumes of new data. The machine learning itself determines what is different or interesting from the dataset.

Applications: Supervised learning models are ideal for spam detection, sentiment analysis, weather forecasting and pricing predictions, among other things. In contrast, unsupervised learning is a great fit for anomaly detection, recommendation engines, customer personas and medical imaging.

Complexity: Supervised learning is a simple method for machine learning, typically calculated through the use of programs like R or Python. In unsupervised learning, you need powerful tools for working with large amounts of unclassified data. Unsupervised learning models are computationally complex because they need a large training set to produce intended outcomes.

Drawbacks: Supervised learning models can be time-consuming to train, and the labels for input and output variables require expertise. Meanwhile, unsupervised learning methods can have wildly inaccurate results unless you have human intervention to validate the output variables.

10. Describe the machine learning process in depth.

Answers:

Machine learning is a subfield of computer science, but is often also referred to as predictive analytics, or predictive modeling. Its goal and usage is to build new and/or leverage existing algorithms to learn from data, in order to build generalizable models that give accurate predictions, or to find pat Imagine a dataset as a table, where the rows are each observation (aka measurement, data point, etc), and the columns for each observation represent the features of that observation and their values.

At the outset of a machine learning project, a dataset is usually split into two or three subsets. The minimum subsets are the training and test datasets, and often an optional third validation dataset is created as well.

Once these data subsets are created from the primary dataset, a predictive model or classifier is trained using the training data, and then the model’s predictive accuracy is determined using the test data.

As mentioned, machine learning leverages algorithms to automatically model and find patterns in data, usually with the goal of predicting some target output or response. These algorithms are heavily based on statistics and mathematical optimization.

Optimization is the process of finding the smallest or largest value (minima or maxima) of a function, often referred to as a loss, or cost function in the minimization case. One of the most popular optimization algorithms used in machine learning is called gradient descent, and another is known as the normal equation.

In a nutshell, machine learning is all about automatically learning a highly accurate predictive or classifier model, or finding unknown patterns in data, by leveraging learning algorithms and optimization techniques. terns, particularly with new and unseen similar data.

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

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

ii. Deep learning applications in healthcare

iii. Study of the market basket

iv. Linear regression (simple)

Answers:

MATLAB: Millions of engineers and scientists worldwide use MATLAB® to analyze and design the systems and products transforming our world. The matrix-based MATLAB language is the world’s most natural way to express computational mathematics. Built-in graphics make it easy to visualize and gain insights from data. The desktop environment invites experimentation, exploration, and discovery. These MATLAB tools and capabilities are all rigorously tested and designed to work together.

MATLAB helps you take your ideas beyond the desktop. You can run your analyses on larger data sets, and scale up to clusters and clouds. MATLAB code can be integrated with other languages, enabling you to deploy algorithms and applications within web, enterprise, and production systems

Deep learning provides the healthcare industry with the ability to analyze data at exceptional speeds without compromising on accuracy. It’s not machine learning, nor is it AI, it’s an elegant blend of both that uses a layered algorithmic architecture to sift through data at an astonishing rate. The benefits of deep learning in healthcare are plentiful – fast, efficient, accurate – but they don’t stop there. Even more benefits lie within the neural networks formed by multiple layers of AI and ML and their ability to learn. Yes, the secret to deep learning’s success is in the name – learning.

Deep learning uses mathematical models that are designed to operate a lot like the human brain. The multiple layers of network and technology allow for computing capability that’s unprecedented, and the ability to sift through vast quantities of data that would previously have been lost, forgotten or missed. These deep learning networks can solve complex problems and tease out strands of insight from reams of data that abound within the healthcare profession. It’s a skillset that hasn’t gone unnoticed by the healthcare profession.

11. Make a comparison between: -

1. Generalization and abstraction

2. Learning that is guided and unsupervised

3. Regression and classification

Answers:

1. Abstraction is the process of removing details of objects. We step back from concrete objects to consider several objects with identical properties. So, a concrete object can be looked at as a “superset” of a more abstract object. A generalization, then, is the formulation of general concepts from specific instances by abstracting common properties. A concrete object can be looked at as a “subset” of a more generalized object.
2. Learning that is guided consists of learning under the supervision of experts or with the assistance of experts in an indirect manner. Subject matter experts accelerate the learning process by teaching the technology in real-time. Assistance from experts helps the learner to learn how to work on difficult tasks. Inherently motivating because it involves the transfer of responsibility, or control, for learning, from the teacher or more capable other to the learner. Unsupervised learning is much similar as a human learns to think by their own experiences, which makes it closer to the real AI.
3. Regression — Estimate continuous values (Real valued output)

Classification — Identify a unique class (Discrete values, Boolean, or Categories)

Regression: - Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Regression can be used to estimate/ predict continuous values (Real valued output). For example: Given a picture of a person, we have to predict the age on the basis of the given picture.

Classification: - Classification means to group the output into a class. If the data set is discrete or categorical then it is a classification problem. For example: Given data about the sizes of houses in the real estate market, making our output about whether the house “sells for more or less than the asking price” i.e. Classifying houses into two discrete categories.