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?

**Ans:** Machine learning is basically means , that machine is learning something and it is able to perform an operation for which it is not being programmed explicitly.

Customer service automation, Credit card fraud detection are the business world example, where machine learning works.

Example of ethical concern is, if my self driving car kills someone, whose fault of this?

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

**Ans:** Under the supervision of expert is a proper guidance and way is being given by experts for human learning.

With assistance of experts in indirect manner is a watch is taken by experts however human mimic the actions provided/taught by experts.

Self education is we can say, where humas are trained by visualising, understanding and taking decisions by their selves, and provide appreciation and penalty to themselves as per their performance.

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

**Ans:** Spam filtering, house price prediction are example of supervised learning, anomaly detection is example of unsupervised learning, and self driving car is an example of reinforcement learning.

4. Examine the various forms of machine learning.

**Ans:** There are four types of machine learning algorithms: supervised, semi-supervised, unsupervised and reinforcement.

In supervised learning, the machine is taught by example. The operator provides the machine learning algorithm with a known dataset that includes desired inputs and outputs, and the algorithm must find a method to determine how to arrive at those inputs and outputs.

Semi-supervised learning is similar to supervised learning, but instead uses both labelled and unlabelled data. Labelled data is essentially information that has meaningful tags so that the algorithm can understand the data, whilst unlabelled data lacks that information. By using this  
combination, machine learning algorithms can learn to label unlabelled data.

Unsupervised machine learning algorithm studies data to identify patterns. There is no answer key or human operator to provide instruction. Instead, the machine determines the correlations and relationships by analysing available data. In an unsupervised learning process, the machine learning algorithm is left to interpret large data sets and address that data accordingly.

Reinforcement learning focuses on regimented learning processes, where a machine learning algorithm is provided with a set of actions, parameters and end values. By defining the rules, the machine learning algorithm then tries to explore different options and possibilities, monitoring and evaluating each result to determine which one is optimal.

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.

**Ans:** Finding a hidden pattern in a big number of observations - this may be an extremely difficult problem for, say, numeric methods of approximation, Classification of new observations, Finding unobvious groups in data.

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

**Ans:** No, machine learning comes up with certain concerns, e.g. ethical issues, deterministic problems, lack of good data, misapplications of implementations etc.

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

**Ans:** Regression, Classification, Clustering, Decision trees, Neural Networks, Support vector machine, Naïve Bayes, K means etc are various methods and techniques to solve different types of machine learning problems.

Regression: Regression methods are used for training supervised ML. The goal of regression techniques is typically to explain or predict a specific numerical value while using a previous data set. For example, regression methods can take historical pricing data, and then predict the price of a similar property to retail demand forecasting.

Linear regression is considered the simplest and most basic method. In this case, a dataset is modeled using the following equation:

( y = m \* x + b )

It is possible to train a regression model with multiple pairs of data, such as x, y. To do this, you need to define a position, as well as the slope of the line, with a minimal distance from all known data points. This is the line that best approximates the observations in the data, and can help make predictions for new unseen data.

Clustering - Clustering algorithms are unsupervised learning methods. A few common clustering algorithms are K-means, mean-shift, and expectation-maximization. They group data points according to similar or shared characteristics.

Grouping or clustering techniques are particularly useful in business applications when there is a need to segment or categorize large volumes of data. Examples include segmenting customers by different characteristics to better target marketing campaigns, and recommending news articles that certain readers will enjoy. Clustering is also effective in discovering patterns in complex data sets that may not be obvious to the human eye.

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

**Ans:** Below are various forms of supervised learning:

Regression - In regression, a single output value is produced using training data. This value is a probabilistic interpretation, which is ascertained after considering the strength of correlation among the input variables. For example, regression can help predict the price of a house based on its locality, size, etc.

Classification - It involves grouping the data into classes. If you are thinking of extending credit to a person, you can use classification to determine whether or not a person would be a loan defaulter. When the supervised learning algorithm labels input data into two distinct classes, it is called binary classification. Multiple classifications means categorizing data into more than two classes.

Naïve Bayesian Model - The Bayesian model of classification is used for large finite datasets. It is a method of assigning class labels using a direct acyclic graph. The graph comprises one parent node and multiple children nodes. And each child node is assumed to be independent and separate from the parent.

Decision Trees - A decision tree is a flowchart-like model that contains conditional control statements, comprising decisions and their probable consequences. The output relates to the labelling of unforeseen data.

In the tree representation, the leaf nodes correspond to class labels, and the internal nodes represent the attributes. A decision tree can be used to solve problems with discrete attributes as well as boolean functions. Some of the notable decision tree algorithms are ID3 and CART.

Random Forest Model - The random forest model is an ensemble method. It operates by constructing a multitude of decision trees and outputs a classification of the individual trees. Suppose you want to predict which undergraduate students will perform well in GMAT – a test taken for admission into graduate management programs. A random forest model would accomplish the task, given the demographic and educational factors of a set of students who have previously taken the test.

Neural Networks - This algorithm is designed to cluster raw input, recognize patterns, or interpret sensory data. Despite their multiple advantages, neural networks require significant computational resources. It can get complicated to fit a neural network when there are thousands of observations. It is also called the ‘black-box’ algorithm as interpreting the logic behind their predictions can be challenging.

Support Vector Machine -SVM separates hyperplanes, which makes it a discriminative classifier. The output is produced in the form of an optimal hyperplane that categorizes new examples. SVMs are closely connected to the kernel framework and used in diverse fields. Some examples include bioinformatics, pattern recognition, and multimedia information retrieval.

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

**Ans:** Supervised learning is the learning of the model where with input variable ( say, x) and an output variable (say, Y) and an algorithm to map the input to the output.

Example of Supervised Learning

Suppose there is a basket which is filled with some fresh fruits, the task is to arrange the same type of fruits at one place.

Also, suppose that the fruits are apple, banana, cherry, grape.

Unsupervised learning is where only the input data (say, X) is present and no corresponding output variable is there.

Again, Suppose there is a basket and it is filled with some fresh fruits. The task is to arrange the same type of fruits at one place. how to group similar fruits without any prior knowledge about those.  
First, any physical characteristic of a particular fruit is selected. Suppose *color*. Then the fruits are arranged on the basis of the color.

10. Describe the machine learning process in depth.

**Ans:** Data Gathering -> Data Preprocessing -> Split train and test data -> ML algorithm selection -> Model evaluation -> Validation -> Complexity & Improvement if any

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

MATLAB is one of the most widely used programming languages.

ii. Deep learning applications in healthcare – Drug discovery, Medical imaging and Diagnostics, Simplifying clinical trials, Improved health records and patient monitoring, Health insurance and fraud detection.

iii. Study of the market basket

iv. Linear regression (simple) - The goal of linear regression techniques is typically to explain or predict a specific numerical value while using a previous data set. For example, regression methods can take historical pricing data, and then predict the price of a similar property to retail demand forecasting.

Linear regression is considered the simplest and most basic method. In this case, a dataset is modeled using the following equation:

( y = m \* x + b )

It is possible to train a regression model with multiple pairs of data, such as x, y. To do this, you need to define a position, as well as the slope of the line, with a minimal distance from all known data points. This is the line that best approximates the observations in the data, and can help make predictions for new unseen data.

11. Make a comparison between:-

1. Generalization and abstraction - Abstraction reduces complexity by hiding irrelevant detail, generalization reduces complexity by replacing multiple entities which perform similar functions with a single construct.

2. Learning that is guided and unsupervised – It is a learning where guidance is provided but supervision does not takes place.

3. Regression and classification - In regression, a single output value is produced using training data, while classification involves grouping the data into classes.