1. What is the concept of human learning? Please give two examples.

Ans: Human learning can be defined as change in behaviour as a result of experience.Human learning is — observing something, identifying a pattern, building a theory (model) to explain this pattern and testing this theory to check if its fits in most or all observations.

Example: Learning to drive a motor car, learning how to solve a mathematical problem based on a concept, memorizing a poem

2. What different forms of human learning are there? Are there any machine learning equivalents?

Ans: Three major types of learning described by [behavioral psychology](https://www.verywellmind.com/behavioral-psychology-4157183) are classical conditioning, operant conditioning, and observational learning.

Operant conditioning is similar to reinforcement learning as it’s principles of reward and punishment are similar to reinforcement learning.

1. What is machine learning, and how does it work? What are the key responsibilities of machine learning?

Ans: Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

The Machine Learning process starts with inputting training data into the selected algorithm. Training data being known or unknown data to develop the final Machine Learning algorithm. The type of training data input does impact the algorithm, and that concept will be covered further momentarily.

New input data is fed into the machine learning algorithm to test whether the algorithm works correctly. The prediction and results are then checked against each other.

If the prediction and results don’t match, the algorithm is re-trained multiple times until the data scientist gets the desired outcome. This enables the machine learning algorithm to continually learn on its own and produce the optimal answer, gradually increasing in accuracy over time.

**Responsibility of ML**

* Selecting appropriate data sets.
* Picking appropriate data representation methods.
* Identifying differences in data distribution that affects model performance.
* Verifying [data quality](https://www.techtarget.com/searchdatamanagement/definition/data-quality).
* Transforming and converting data science prototypes.
* Performing [statistical analysis](https://www.techtarget.com/whatis/definition/statistical-analysis).

1. Define the terms "penalty" and "reward" in the context of reinforcement learning.

Ans: Reinforcement learning is an approach to machine learning that is inspired by behaviorist psychology. It is similar to how a child learns to perform a new task. Reinforcement learning contrasts with other machine learning approaches in that the algorithm is not explicitly told how to perform a task, but works through the problem on its own.

As an agent, which could be a self-driving car or a program playing chess, interacts with its environment, receives a reward state depending on how it performs, such as driving to destination safely or winning a game. Conversely, the agent receives a penalty for performing incorrectly, such as going off the road or being checkmated.

1. Explain the term "learning as a search"?

Ans: Searching through a large space of hypotheses implicitly defined by the hypothesis representation (same for more general learning).

The hypothesis representation defines the space of hypotheses the program can ever represent and therefore can ever learn.

For example, Sky has 3 possible values and Temp,Humidity,Wind,Water, and Forecast each have 2 possible values.

1. What are the various goals of machine learning? What is the relationship between these and human learning?

Ans: The goal of machine learning is to discover patterns in the user data and then make predictions based on these and intricate patterns for answering business questions and solving business problems. Machine learning helps in analysing the data as well as identifying trends. 7. Illustrate the various elements of machine learning using a real-life illustration.

Human learning also depends on the environmental parameters, principles to learn new things by establishing some patterns, so that in case of unknown situation it may be able to generalize well.

1. Illustrate the various elements of machine learning using a real-life illustration.

Ans: There are three main elements to every machine learning algorithm, and they include:

* Representation: what the model looks like; how knowledge is represented
* Evaluation: how good models are differentiated; how programs are evaluated
* Optimization: the process for finding good models; how programs are generated

For example in the case of machine learning model for risk estimation for Credit facilties offered by banks, the representative data should contain enough instances for both the high risk and low risk instances.

For evaluation purpose, the test data should contain approximately similar ratio positive and negative isntaces so that performance of the model could be ascertained.

Depending on the test result the model could be optimized by tuning its hyper parameters for better accuracy in predicting either good or bad loan portfolio.

1. Provide an example of the abstraction method.

Abstraction is defined as dealing with ideas instead of events. In the context of ML, that means worrying more about what the right algorithm is and less about how to implement it. Another way of looking at it, for those technically inclined, is as an API call (abstracted) vs. a self implemented function or series of functions.

Machine Learning is being abstracted into aproject that can take just a few days and a few hundred lines of code or less. There are a couple of trends that I think are pushing this envelope forward.

1. What is the concept of generalization? What function does it play in the machine learning process?

Ans: In machine learning, generalization demonstrate how well is a trained model to classify or forecast unseen data.

When we train a model to classify between dogs and cats. If the model is provided with dogs images dataset with only two breeds, it may obtain a good performance. But, it possibly gets a low classification score when it is tested by other breeds of dogs as well. This issue can result to classify an actual dog image as a cat from the unseen dataset. Therefore, data diversity is very important factor in order to make a good prediction. In the sample above, the model may obtain 85% performance score when it is tested by only two dog breeds and gains 70% if trained by all breeds. However, the first possibly gets a very low score (e.g. 45%) if it is evaluated by an unseen dataset with all breed dogs. This for the latter can be unchanged given than it has been trained by high data diversity including all possible breeds.

1. What is classification, exactly? What are the main distinctions between classification and regression?

Ans: The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, **Yes or No, 0 or 1, Spam or Not Spam, cat or dog,** etc. Classes can be called as targets/labels or categories.

In classification algorithm, a discrete output function(y) is mapped to input variable(x).

Unlike regression, the output variable of Classification is a category, not a value, such as "Green or Blue", "fruit or animal", etc. Since the Classification algorithm is a Supervised learning technique, hence it takes labeled input data, which means it contains input with the corresponding output.

1. What is regression, and how does it work? Give an example of a real-world problem that was solved using regression.

A regression problem is when the output variable is a real or continuous value, such as “salary” or “weight”.

It tries to fit data with the best hyper-plane which goes through the points.

Real world example:

Businesses often use linear regression to understand the relationship between advertising spending and revenue.

For example, they might fit a simple linear regression model using advertising spending as the predictor variable and revenue as the response variable. The regression model would take the following form:

**revenue = β0 + β1(ad spending)**

The coefficient **β0** would represent total expected revenue when ad spending is zero.

The coefficient **β1** would represent the average change in  total revenue when ad spending is increased by one unit (e.g. one dollar).

If β1 is negative, it would mean that more ad spending is associated with less revenue.

If β1 is close to zero, it would mean that ad spending has little effect on revenue.

And if β1 is positive, it would mean more ad spending is associated with more revenue.

Depending on the value of β1, a company may decide to either decrease or increase their ad spending.

1. Describe the clustering mechanism in detail.

**Clustering** is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

**For example**– The data points in the graph below clustered together can be classified into one single group. We can distinguish the clusters, and we can identify that there are 3 clusters in the below picture.

Clustering is very much important as it determines the intrinsic grouping among the unlabelled data present. There are no criteria for good clustering. It depends on the user, what is the criteria they may use which satisfy their need. For instance, we could be interested in finding representatives for homogeneous groups (data reduction), in finding “natural clusters” and describe their unknown properties (“natural” data types), in finding useful and suitable groupings (“useful” data classes) or in finding unusual data objects (outlier detection). This algorithm must make some assumptions that constitute the similarity of points and each assumption make different and equally valid clusters.

13. Make brief observations on two of the following topics:

i. Machine learning algorithms are used:

A. Image recognition is a well-known and widespread example of machine learning in the real world. It can identify an object as a digital image, based on the intensity of the pixels in black and white images or colour images.

B. Speech Recognition: Machine learning can translate speech into text. Certain software applications can convert live voice and recorded speech into a text file. The speech can be segmented by intensities on time-frequency bands as well.

C. Predictive Analysis- Machine learning can classify available data into groups, which are then defined by rules set by analysts. When the classification is complete, the analysts can calculate the probability of a fault.

ii. Studying under supervision

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.

In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.

Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).

Steps Involved in Supervised Learning:

* First Determine the type of training dataset
* Collect/Gather the labelled training data.
* Split the training dataset into training dataset, test dataset, and validation dataset.
* Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
* Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
* Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
* Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

iii. Studying without supervision

iv. Reinforcement learning is a form of learning based on positive reinforcement.