**Question Bank**

**2 marks**

1. **What is what if Analysis?**

What-If Analysis is **the process of changing the values in cells to see how those changes will affect the outcome of formulas on the worksheet**. Three kinds of What-If Analysis tools come with Excel: Scenarios, Goal Seek, and Data Tables. Scenarios and Data tables take sets of input values and determine possible results.

1. **List different types of What if Analysis.**

Three kinds of What-If Analysis tools come with Excel: Scenarios, Goal Seek, and Data Tables. Scenarios and Data tables take sets of input values and determine possible results.

1. **Do we always need a formulae cell to do what-if analysis?**

**Yes**

**4. What is the advantage of sorting?**

The ability to sort is very important for analysis and manipulation of data. You can sort data by text (A to Z or Z to A), numbers (smallest to largest or largest to smallest), and dates and times (oldest to newest and newest to oldest) in one or more columns. You can also sort by a custom list or by format, including cell color, font color, or icon set. Most sort operations are column sorts, but you can also sort by rows.

Excel will attempt to sort the entire range of data selected.   This is very important from the standpoint of data integrity.

**5. What is the advantage of regression analysis?**

Regression analysis is a reliable method of identifying which variables have impact on a topic of interest. The process of performing a regression allows you to confidently determine which factors matter most, which factors can be ignored, and how these factors influence each other.

**6. What is sorting?**

* Sorting in Excel is arranging data according to our requirements.
* It can be done alphabetically or numerically.
* Basic Sorting works when sorting is to do on only one column.
* Advanced Sorting is used in multi-level sorting, viz sorting required in 2 or more than 2 columns.
* Two methods can do advanced Sorting:

1. Sorting Dialogue Method
2. Sorting Icon Method

**7. What is the filter?**

The filter in excel **helps display relevant data by eliminating the irrelevant entries temporarily from the view**. The data is filtered as per the given criteria. The purpose of filtering is to focus on the crucial areas of a dataset.

**8. What is machine learning?**

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning focuses on the development of computer programs** that can access data and use it to learn for themselves.

**9. Define statistics**

**1:**a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data

**2:**a collection of quantitative data

**10. Where is unsupervised learning used?**

Unsupervised learning is commonly used for **finding meaningful patterns and groupings inherent in data**, extracting generative features, and exploratory purposes.

**5 marks**

* 1. **Explain correlation with the help of an example.**
* Correlation means association - more precisely it is a measure of the extent to which two variables are related. There are three possible results of a correlational study: a positive correlation, a negative correlation, and no correlation.
*  A **positive correlation** is a relationship between two variables in which both variables move in the same direction. Therefore, when one variable increases as the other variable increases, or one variable decreases while the other decreases. An example of positive correlation would be height and weight. Taller people tend to be heavier.
*  A **negative correlation** is a relationship between two variables in which an increase in one variable is associated with a decrease in the other. An example of negative correlation would be height above sea level and temperature. As you climb the mountain (increase in height) it gets colder (decrease in temperature).
* A **zero correlation** exists when there is no relationship between two variables. For example there is no relationship between the amount of tea drunk and level of intelligence.

1. **Explain regression analysis.**

### *Regression Analysis*

In the domain of statistics, the most commonly used statistical technique is Regression Analysis which is used to estimate particular relationships among variables. Under this technique, the main focus is upon the relationship between dependent variable and any one or more independent variables. There are several techniques within this analysis that are used for modeling and analyzing several variables. This technique helps you see how the particular value of a dependent variable changes when any one of the independent variable varies with all others fixed. In simple terms, through this approach you get to estimate the conditional expectation or the average value of the dependent variable. Thoroughly in all the cases, the target for estimation is a function of any one or more independent variables, which is termed as regression function. The main goal of regression analysis is to ascertain the values of all the parameters to derive a function that will fit the data observations in the best way possible.

### *Common uses of Regression Analysis*

Most widely, this technique is used for predicting and forecasting. In both these fields, the uses overlap with that of the domain of machine learning. The technique also helps you in figuring out the form and type of relationships that forms between a dependent variable and the independent variable. It also interprets the casual relationships between the same.

# Discuss the role of data filter.

Data filtering is the process of choosing a smaller part of your data set and using that subset for viewing or analysis. Filtering is generally (but not always) temporary – the complete data set is kept, but only part of it is used for the calculation.

Filtering may be used to:

* Look at results for a particular period of time.
* Calculate results for particular groups of interest.
* Exclude erroneous or "bad" observations from an analysis.
* Train and validate statistical models.

Filtering requires you to specify a rule or logic to identify the cases you want to included in your analysis. Filtering can also be referred to as “subsetting” data, or a data “drill-down”.

# What is the difference between filter and sorting?

Data filtering is the process of choosing a smaller part of your data set and using that subset for viewing or analysis. Filtering is generally (but not always) temporary – the complete data set is kept, but only part of it is used for the calculation.

Filtering may be used to:

* Look at results for a particular period of time.
* Calculate results for particular groups of interest.
* Exclude erroneous or "bad" observations from an analysis.
* Train and validate statistical models.

Filtering requires you to specify a rule or logic to identify the cases you want to included in your analysis. Filtering can also be referred to as “subsetting” data, or a data “drill-down”.

**5. Name 5 programming languages most commonly used for AI.**

Top 5 AI programming languages:

* LISP
* Python
* C++
* Java
* Prolog

## LISP

LISP is the oldest AI programming language. It is the second oldest programming language after Fortran. The term Artificial intelligence was made up by John McCarthy who invented LISP. Another pioneer was Marvin Minsky, who founded the AI lab at MIT.

There would be no progress in AI at that time if it weren’t for LISP. It had fresh ideas ( if-then-else, construct, recursion), which were very useful to express the ideas programmers had. Because of the huge adoption of LISP, it became a standard AI language. LISP is a very flexible AI programming language and is often called “the most intelligent way to misuse a computer.”

## Python

[Python](https://ncube.com/technologies/python-remote-teams) is an AI programming language that has gained huge popularity. The main reasons are the simple syntax, less coding and a large number of available libraries ready for use. Simple syntax means you can focus on the core value of programming, thinking, or problem-solving.

The earlier mentioned libraries include NumPy, SciPy, matplotlib, nltk, SimpleAI. Python is an open-source AI programming language. That’s why it has a huge fan base among programmers. Because it can be used broadly, to make small scripts and up to enterprise applications, it’s suitable for AI.

Where other AI programming languages use punctuation, Python uses English keywords. It’s designed to be readable. It has only a few keywords and has a clearly defined syntax. If you are a student, you will pick up the language quickly.

The libraries are portable across platforms such as UNIX, Windows, and Macintosh.

## C++

Bjarne Stroustrup developed C++ in 1983, and it holds the title for the “fastest programming language.” Time is important for AI projects and C++ is the usual choice when it comes to that.

Search engines use C++ to have less response time, and the development of [computer games](https://ncube.com/blog/looking-for-needle-in-the-hay-how-to-hire-the-right-ai-talent) takes advantage of the fast execution.

Because C++ has a complex syntax, it might not be your first choice as an AI programming language, but if you are working in an embedded environment and can’t afford Java Virtual Machine, this is the perfect option for you.

## Java

Java is the most widely used programming language in the world and is one of the best choices of AI programming. Because of its Virtual Machine Technology, it’s easy to implement on different platforms. That means once it’s written and compiled on one platform, you don’t have to compile it again. It’s called the WORA (Once Written Read/Run Anywhere) principle.

Most of the open-source big data stack is written in Java Virtual Machine.

## Prolog

[Prolog](https://www.swi-prolog.org/) is another AI programming languages and is one of the oldest ones. It stands for “Programming in logic.” The language is based on a few basic mechanisms like pattern matching and automatic backtracking. It was made in the 1970s by a French computer scientist named Alain Colmerauer.

In Prolog AI programming, the programmer has to specify the rules and the facts and the end goal. After the programmer has done that, Prolog will try to find the connection between those three. This is an unusual way of AI programming, and it has its advantages and disadvantages.

## Discuss descriptive statistics.

Descriptive statistics are brief descriptive coefficients that summarize a given data set, which can be either a representation of the entire population or a sample of a population. Descriptive statistics are broken down into measures of central tendency and measures of variability (spread). Measures of central tendency include the mean, median, and mode, while measures of variability include standard deviation, variance, minimum and maximum variables, [kurtosis](https://www.investopedia.com/terms/k/kurtosis.asp), and [skewness](https://www.investopedia.com/terms/s/skewness.asp).

Descriptive statistics, in short, help describe and understand the features of a specific data set by giving short summaries about the sample and measures of the data. The most recognized types of descriptive statistics are measures of center: the [mean](https://www.investopedia.com/terms/m/mean.asp), [median](https://www.investopedia.com/terms/m/median.asp), and [mode](https://www.investopedia.com/terms/m/mode.asp), which are used at almost all levels of math and statistics. The mean, or the average, is calculated by adding all the figures within the data set and then dividing by the number of figures within the set.

## Discuss inferential statistics.

nferential statistics is one of the two main branches of statistics.

Inferential statistics use a random sample of data taken from a population to describe and make inferences about the population. Inferential statistics are valuable when examination of each member of an entire population is not convenient or possible. For example, to measure the diameter of each nail that is manufactured in a mill is impractical. You can measure the diameters of a representative random sample of nails. You can use the information from the sample to make generalizations about the diameters of all of the nails.

## What are Pivot Tables? How it helps in data interpretation and processing.

A PivotTable is an interactive way to quickly summarize large amounts of data. You can use a PivotTable to analyze numerical data in detail, and answer unanticipated questions about your data. A PivotTable is especially designed for:

* Querying large amounts of data in many user-friendly ways.
* Subtotaling and aggregating numeric data, summarizing data by categories and subcategories, and creating custom calculations and formulas.
* Expanding and collapsing levels of data to focus your results, and drilling down to details from the summary data for areas of interest to you.
* Moving rows to columns or columns to rows (or "pivoting") to see different summaries of the source data.
* Filtering, sorting, grouping, and conditionally formatting the most useful and interesting subset of data enabling you to focus on just the information you want.
* Presenting concise, attractive, and annotated online or printed reports.

**9. How Machine learning helps in recommending your favourite movie/web series at**

**Online Media or suggesting your products online.**

**5 Use Cases of AI/Data/Machine Learning at Netflix**

1. **Personalization of Movie Recommendations**— Users who watch A are likely to watch B. This is perhaps the most well known feature of a Netflix. Netflix uses the watching history of other users with similar tastes to recommend what you may be most interested in watching next so that you stay engaged and continue your monthly subscription for more.
2. **Auto-Generation and Personalization of Thumbnails / Artwork**— Using thousands of video frames from an existing movie or show as a starting point for thumbnail generation, Netflix annotates these images then ranks each image in an effort to identify which thumbnails have the highest likelihood of resulting in your click. These calculations are based on what others who are similar to you have clicked on. One finding could be that users who like certain actors / movie genres are more likely to click thumbnails with certain actors/image attributes.
3. **Location Scouting for Movie Production (Pre-Production)** — Using data to help decide on where and when best to shoot a movie set — given constraints of scheduling (actor/crew availability), budget(venue, flight/hotel costs), and production scene requirements (day vs night shoot, likelihood of weather event risks in a location). Notice this is more of a data science optimization problem rather than a machine learning model that makes predictions based on past data.
4. **Movie Editing (Post-Production)** —Using historical data of when quality control checks have failed in the past (when syncing of subtitles to sound/movements were off in the past) — to predict when a manual check is most beneficial in what could otherwise be a very time-intensive and laborious process.
5. **Streaming Quality** — Using past viewing data to predict bandwidth usage to help Netflix decide when to cache regional servers for faster load times during peak (expected) demand.

These 5 use cases / applications of data science or machine learning just in Netflix alone have had such scalable impact that they have forever changed the technology landscape and user experience for millions and more to come. Adoption of these AI-related solutions is only going to get stronger over time.

**10. Discuss supervised learning with the help of an example.**

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)**.

In the real-world, supervised learning can be used for **Risk Assessment, Image classification, Fraud Detection, spam filtering**, etc.

**8 marks**

### What is the role of charts in data analysis? Explain with examples.

A chart can represent tabular numeric data, functions or some kinds of qualitative structure and provides different info. ... A chart can create a clearer picture of a set of data values than a table with rows of numbers in it, allowing managers to incorporate this understanding into analysis and future planning.

# 1. Column Chart

Column charts use vertical columns to show numerical comparisons between categories, and the number of columns should not be too large (the labels of the axis may appear incomplete if there are too many columns). The column chart takes advantage of the height of the column to reflect the difference in the data, and the human eye is sensitive to height differences. The limitation is that it is only suitable for small and medium-sized data sets.

# 2. Bar Chart

Bar charts are similar to column charts, but the number of bars can be relatively large. Compared with the column chart, the positions of its two axes are changed.

# 3. Line Chart

A line chart is used to show the change of data over a continuous time interval or time span. It is characterized by a tendency to reflect things as they change over time or ordered categories.

It should be noted that the number of data records of the line graph should be greater than 2, which can be used for trend comparison of large data volume. And it is better not to exceed 5 polylines on the same graph.

# Area Chart

The area chart is formed on the basis of the line chart. It fills the area between the polyline and the axis in the line chart with color. The filling of the color can better highlight the trend information.

The fill color of the area chart should have a certain transparency. The transparency can help the user to observe the overlapping relationship between different series. The area without transparency will cause the different series to cover each other.

# 5. Pie Chart

Pie charts are widely used in various fields to represent the proportion of different classifications, and to compare various classifications by the arc.

The pie chart is not suitable for multiple series of data, because as the series increase, each slice becomes smaller, and finally the size distinction is not obvious.

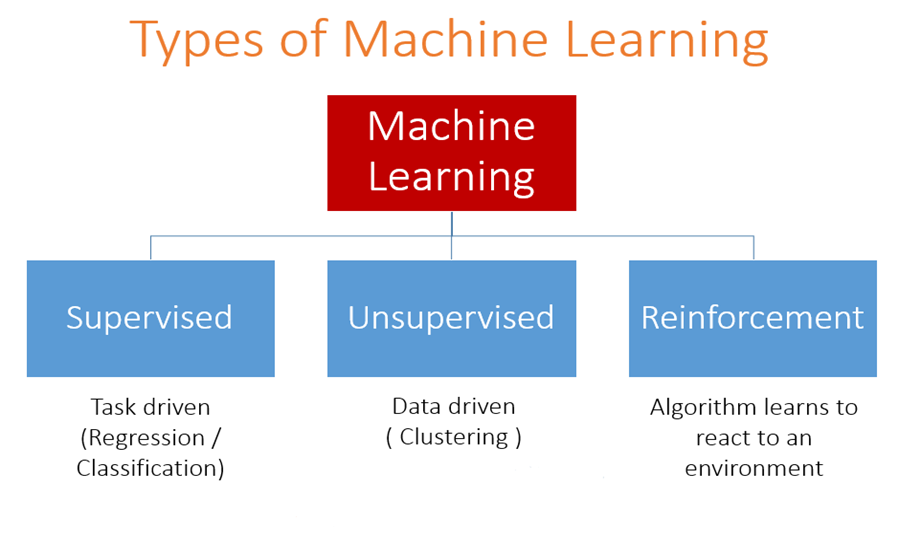
A pie chart can also be made into a multi-layer pie chart, showing the proportion of different categorical data, while also reflecting the hierarchical relationship.

1. **Name and discuss types of Machine Learning.**

There are three kinds of Machine Learning Algorithms.

a. Supervised Learning

b. Unsupervised Learning

c. Reinforcement Learning

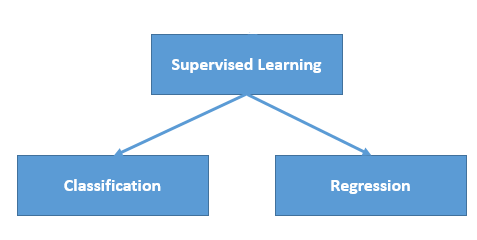
a. **Supervised Learning**

A majority of practical machine learning uses supervised learning.

In supervised learning, the system tries to learn from the previous examples that are given. (On the other hand, in unsupervised learning, the system attempts to find the patterns directly from the example given.)

Speaking mathematically, supervised learning is where you have both input variables (x) and output variables(Y) and can use an algorithm to derive the mapping function from the input to the output. The mapping function is expressed as Y = f(X).

## ****Example****



Supervised learning problems can be further divided into two parts, namely classification, and regression.

**b. Unsupervised Learning**

In unsupervised learning, the algorithms are left to themselves to discover interesting structures in the data. Mathematically, unsupervised learning is when you only have input data (X) and no corresponding output variables.

This is called unsupervised learning because unlike supervised learning above, there are no given correct answers and the machine itself finds the answers.

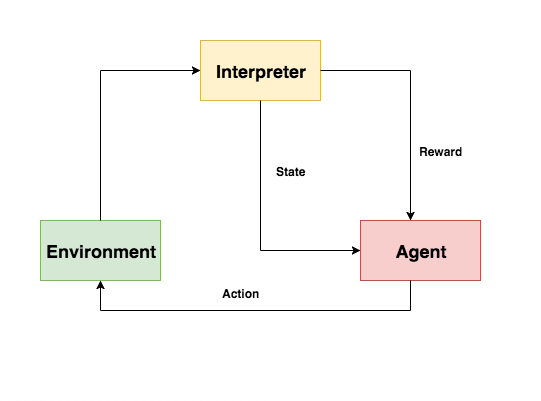
Unsupervised learning problems can be further divided into association and clustering problems.

* **Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as “people that buy X also tend to buy Y”.
* **Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.

**c. Reinforcement Learning**

A computer program will interact with a dynamic environment in which it must perform a particular goal (such as playing a game with an opponent or driving a car). The program is provided feedback in terms of rewards and punishments as it navigates its problem space. Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it continuously trains itself using trial and error method.

**Example:**



**3. Discuss the various trending applications that are immensely influenced by**

**Machine Learning.**

## Social Media (Facebook)

One of the most common applications of Machine Learning is **Automatic Friend Tagging Suggestions**in Facebook or any other social media platform. Facebook uses **face detection** and **Image recognition** to automatically find the face of the person which matches it’s Database and hence suggests us to tag that person based on DeepFace.

Facebook’s Deep Learning project **DeepFace**is responsible for the recognition of faces and identifying which person is in the picture. It also provides Alt Tags (Alternative Tags) to images already uploaded on facebook.  For eg., if we inspect the following image on Facebook, the alt-tag has a description.

## Self Driving Cars

Well, here is one of the coolest application of Machine Learning. It’s here and people are already using it. Machine Learning plays a very important role in Self Driving Cars and I’m sure you guys might have heard about **Tesla**. The leader in this business and their current ***Artificial Intelligence*** is driven by hardware manufacturer **NVIDIA**, which is based on Unsupervised Learning Algorithm.

NVIDIA stated that they didn’t train their model to detect people or any object as such. The model works on ***Deep Learning*** and it crowdsources data from all of its vehicles and its drivers. It uses internal and external sensors which are a part of **IOT**.  According to the data gathered by McKinsey, the automotive data will hold a tremendous value of **$750 Billion.**

## Google Translate

Remember the time when you traveled to a new place and you find it difficult to communicate with the locals or finding local spots where everything is written in a different language. Well, those days are gone now. Google’s **GNMT**(Google Neural Machine Translation) is a Neural Machine Learning that works on thousands of languages and dictionaries, uses [***Natural Language Processing***](https://www.youtube.com/watch?v=5ctbvkAMQO4) to provide the most accurate translation of any sentence or words. Since the tone of the words also matters, it uses other techniques like POS Tagging, NER (Named Entity Recognition) and Chunking. It is one of the best and most used Applications of Machine Learning.

## Online Video Streaming (Netflix)

With over 100 million subscribers, there is no doubt that Netflix is the daddy of the online streaming world. Netflix’s speedy rise has all movie industrialists taken aback – forcing them to ask, “**How on earth could one single website take on Hollywood?”**. The answer is Machine Learning.

The Netflix algorithm constantly gathers massive amounts of data about users’ activities like:

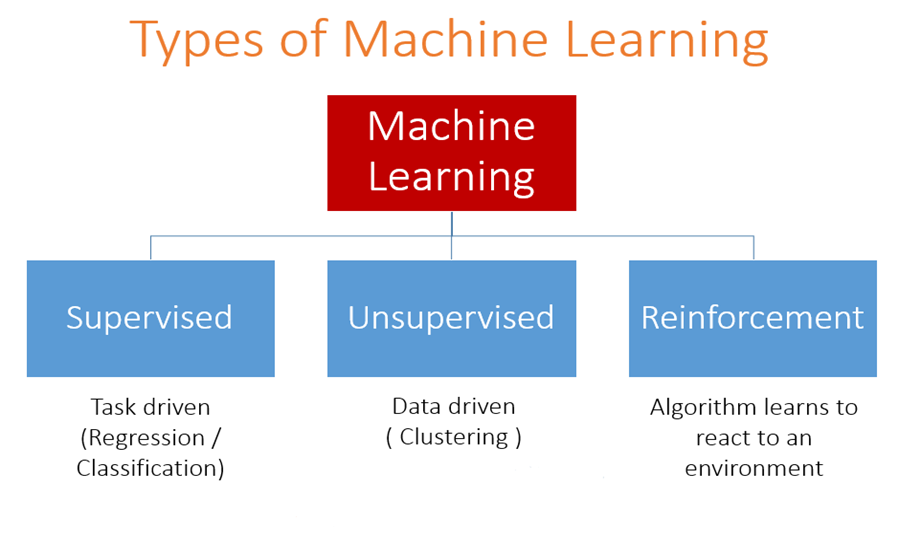
* When you pause, rewind, or fast forward
* What day you watch content (TV Shows on Weekdays and Movies on Weekends)
* The Date and Time you watch
* When you pause and leave content (and if you ever come back)
* The ratings Given (about 4 million per day), Searches (about 3 million per day)
* Browsing and Scrolling Behavior

## Discuss different types of machine learning with examples.

There are three kinds of Machine Learning Algorithms.

a. Supervised Learning

b. Unsupervised Learning

c. Reinforcement Learning

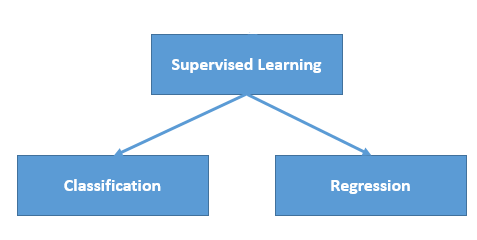
a. **Supervised Learning**

A majority of practical machine learning uses supervised learning.

In supervised learning, the system tries to learn from the previous examples that are given. (On the other hand, in unsupervised learning, the system attempts to find the patterns directly from the example given.)

Speaking mathematically, supervised learning is where you have both input variables (x) and output variables(Y) and can use an algorithm to derive the mapping function from the input to the output. The mapping function is expressed as Y = f(X).

## ****Example****



Supervised learning problems can be further divided into two parts, namely classification, and regression.

**b. Unsupervised Learning**

In unsupervised learning, the algorithms are left to themselves to discover interesting structures in the data. Mathematically, unsupervised learning is when you only have input data (X) and no corresponding output variables.

This is called unsupervised learning because unlike supervised learning above, there are no given correct answers and the machine itself finds the answers.

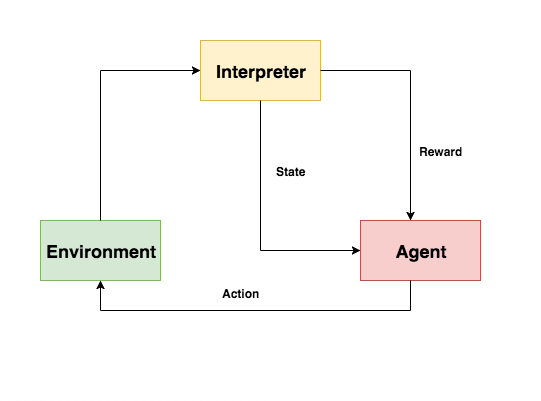
Unsupervised learning problems can be further divided into association and clustering problems.

* **Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as “people that buy X also tend to buy Y”.
* **Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.

**c. Reinforcement Learning**

A computer program will interact with a dynamic environment in which it must perform a particular goal (such as playing a game with an opponent or driving a car). The program is provided feedback in terms of rewards and punishments as it navigates its problem space. Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it continuously trains itself using trial and error method.

**Example:**



## Describe what is the filter? Steps to use filters in excel. What is its purpose?

In addition to sorting, you may find that adding a filter allows you to better analyze your data. When data is filtered, only rows that meet the filter criteria will display and other rows will be hidden. With filtered data, you can then copy, format, print, etc., your data, without having to sort or move it first. To use a filter,

* Go to the Home ribbon, click the arrow below the Sort & Filtering icon in the Editing group and choose Filter.

OR

* Go to the Data ribbon, and then click Filter in the Sort & Filter group.

You will notice that all of your column headings now have an arrow next to the heading name. Click on the arrow next to the heading with which you want to filter, and you will see a list of all the unique values in that column. Check the box next to the criteria you wish to match and click OK. Click on the arrow next to another heading to further filter the data.  
  
To clear the filter, choose one of these options:

* Click on the Filter icon next to the heading and choose Clear Filter from “Name of Heading”.
* Go to the Data ribbon and click the Clear icon in the Sort & Filter group.
* Go to the Home ribbon, click the arrow below the Sort & Filter icon in the Editing Sorting Data

For a quick sort, click the arrow below the Sort & Filtering icon in the Editing group of the Home ribbon and choose the Sort A to Z / Z to A icons in the Sort & Filter group of the Data ribbon. In Excel 2013, these are labeled Sort Smallest to Largest and vice versa.

For a more complex sort, go to the Home ribbon, click the arrow below the Sort & Filter icon in the Editing group and choose Custom Sort. This takes you to the same Sort dialog box you get with the Sort icon in the Sort & Filter group of the Data ribbon.

1. Under Column, choose the first column that you would like to sort. If you want to sort multiple columns, click the Add Level button.

2. Under Sort On, choose how you would like to sort. Note that Excel can sort by cell or font color in addition to values.

3. Under Order, choose A to Z (ascending), Z to A (descending), or Custom List.

4. Click OK to perform the sort.

**6. What is reinforcement machine learning? Explain with the help of an example. List**

**2 areas of applications.**

Reinforcement learning is an area of Machine Learning. It is about taking suitable action to maximize reward in a particular situation. It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation. Reinforcement learning differs from 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 but the reinforcement agent decides what to do to perform the given task. In the absence of a training dataset, it is bound to learn from its experience.

**Main points in Reinforcement learning –** 

* Input: The input should be an initial state from which the model will start
* Output: There are many possible outputs as there are a variety of solutions to a particular problem
* Training: The training is based upon the input, The model will return a state and the user will decide to reward or punish the model based on its output.
* The model keeps continues to learn.
* The best solution is decided based on the maximum reward.

**Types of Reinforcement:** There are two types of Reinforcement: 

1. **Positive –**   
   Positive Reinforcement is defined as when an event, occurs due to a particular behavior, increases the strength and the frequency of the behavior. In other words, it has a positive effect on behavior.

Advantages of reinforcement learning are:

* + Maximizes Performance
  + Sustain Change for a long period of time
  + Too much Reinforcement can lead to an overload of states which can diminish the results

1. **Negative –**   
   Negative Reinforcement is defined as strengthening of behavior because a negative condition is stopped or avoided.

Advantages of reinforcement learning:

* + Increases Behavior
  + Provide defiance to a minimum standard of performance
  + It Only provides enough to meet up the minimum behavior

**Various Practical applications of Reinforcement Learning –** 

* RL can be used in robotics for industrial automation.
* RL can be used in machine learning and data processing
* RL can be used to create training systems that provide custom instruction and materials according to the requirement of students.

RL can be used in large environments in the following situations: 

1. A model of the environment is known, but an analytic solution is not available;
2. Only a simulation model of the environment is given (the subject of simulation-based optimization)
3. The only way to collect information about the environment is to interact with it.

**7. What are soring and filters? How sorting and filter are useful for AI?**