Learning From Example

- Agents that can improve their behavior through diligent study of their own experiences
- An agent is learning if it improves its performance on future tasks after making observations about the world.

What is Machine Learning

- Machine Learning is said as a subset of artificial intelligence that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own.
- The term machine learning was first introduced by Arthur Samuel in 1959.
- Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed.
- Machine Learning is defined as a technology that is used to train machines to perform various actions such as predictions, recommendations, estimations, etc., based on historical data or past experience.
- Machine Learning enables computers to behave like human beings by training them with the help of past experience and predicted data.
- With the help of sample historical data, which is known as training data, machine learning algorithms build a mathematical model that helps in making predictions or decisions without being explicitly programmed.
- The more we will provide the information, the higher will be the performance. A machine has the ability to learn if it can improve its performance by gaining.
- A Machine Learning system learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it.
- The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.
- Suppose we have a complex problem, where we need to perform some predictions, so instead of writing a code for it, we just need to feed the data to generic algorithms, and with the help of these algorithms, machine builds the logic as per the data and predict the output. Machine learning has changed our way of thinking about the problem. The below block diagram explains the working of Machine Learning algorithm:

- There are three key aspects of Machine Learning, which are as follows:
    1. Task: A task is defined as the main problem in which we are interested. This task/problem can be related to the predictions and recommendations and estimations, etc.
    2. Experience: It is defined as learning from historical or past data and used to estimate and resolve future tasks.
    3. Performance: It is defined as the capacity of any machine to resolve any machine learning task or problem and provide the best outcome for the same. However, performance is dependent on the type of machine learning problems.

Forms of Learning

Techniques in Machine Learning

Machine Learning techniques are divided mainly into the following 4 categories:

1. Supervised Learning

- Supervised learning is a type of machine learning method in which we provide labeled data (input as well as output data) to the machine learning system in order to train it, and on the basis of that, it predicts the output.
- The labelled data means some input data is already tagged with the correct output.
- The labels in the data help the algorithm to correlate the features.
- The system creates a model using labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not.
- The goal of supervised learning is to map input data with the output data.
- The supervised learning is based on supervision, and it is the same as when a student learns things in the supervision of the teacher.
- Supervised learning technique helps us to predict future events with the help of past experience and labeled examples. Initially, it analyses the known training dataset, and later it introduces an inferred function that makes predictions about output values. Further, it also predicts errors during this entire learning process and also corrects those errors through algorithms.
- Example: Let's assume we have a set of images tagged as "dog". A machine learning algorithm is trained with these dog images so it can easily distinguish whether an image is a dog or not.
- The example of supervised learning is spam filtering.
- Supervised learning can be grouped further in two categories of algorithms:
    i. Classification
    ii. Regression

→ 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.
→ Types of supervised Machine learning Algorithms:
Supervised learning can be further divided into two types of problems:

1. Regression

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning:

- Linear Regression
- Regression Trees
- Non-Linear Regression
- Bayesian Linear Regression
- Polynomial Regression

2. Classification

→ Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc. Spam Filtering
→ 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.
→ The task of the classification algorithm is to find the mapping function to map the input(x) to the discrete output(y)
→ Types of ML Classification Algorithms:

- Random Forest
- Decision Trees
- Logistic Regression
- Support vector Machines
- Naïve Bayes

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.

Difference between Classification and Regression

Classification	Regression

Classification gives out discrete values.	Regression gives continuous values.
Given a group of data, this method helps group the data into different groups.	It uses the mapping function to map values to continuous output.
In classification, the nature of the predicted data is unordered.	Regression has ordered predicted data.
The mapping function is used to map values to pre–defined classes.	It attempts to find a best fit line. It tries to extrapolate the graph to find/predict the values.
Regression algorithms can be used to solve the regression problems such as Weather Prediction, House price prediction, etc.	Classification Algorithms can be used to solve classification problems such as Identification of spam emails, Speech Recognition, Identification of cancer cells, etc.
Example include Decision tree, logistic regression.	Examples include Regression tree (Random forest), Linear regression
Classification is done by measuring the accuracy.	Regression is done using the root mean square error method.

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.

When the output y is one of a finite set of values (such as sunny, cloudy or rainy), the learning problem is called classification, and is called Boolean or binary classification if there are only two values.

When y is a number (such as tomorrow's temperature), the REGRESSION learning problem is called regression

Linear Regression in Machine Learning

- Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.
- Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression.

- Linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.
- The linear regression model provides a sloped straight line representing the relationship between the variables.

  Mathematically, we can represent a linear regression as:

Linear Regression Line

A linear line showing the relationship between the dependent and independent variables is called a regression line.

A regression line can show two types of relationship:

i. **Positive Linear Relationship**: If the dependent variable increases on the Y-axis and independent variable increases on X-axis, then such a relationship is termed as a Positive linear relationship.
ii. **Negative Linear Relationship**: If the dependent variable decreases on the Y-axis and independent variable increases on the X-axis, then such a relationship is called a negative linear relationship.

Finding the best fit line:

- → When working with linear regression, our main goal is to find the best fit line that means the error between predicted values and actual values should be minimized.
- → The best fit line will have the least error.
- → The different values for weights or the coefficient of lines (a0, a1) gives a different line of regression, so we need to calculate the best values for a0 and a1 to find the best fit line, so to calculate this we use cost function.

Cost function

- → Cost function is used to estimate the values of the coefficient for the best fit line.
- → Cost function optimizes the regression coefficients or weights.
- → It measures how a linear regression model is performing.
- → We can use the cost function to find the accuracy of the mapping function, which maps the input variable to the output variable.
- → This mapping function is also known as Hypothesis function.
- → For Linear Regression, we use the Mean Squared Error (MSE) cost function, which is the average of squared error occurred between the predicted values and actual values. It can be written as:

Gradient Descent

- → Gradient descent is used to minimize the MSE by calculating the gradient of the cost function.
- → A regression model uses gradient descent to update the coefficients of the line by reducing the cost function.
- → It is done by a random selection of values of coefficient and then iteratively update the values to reach the minimum cost function.

Model Performance

- → The Goodness of fit determines how the line of regression fits the set of observations.
- → The process of finding the best model out of various models is called optimization.
- → It can be achieved by R-squared method.
- → R-squared is a statistical method that determines the goodness of fit. It measures the strength of the relationship between the dependent and independent variables on a scale of 0-100%.

→ The high value of R-square determines the less difference between the predicted values and actual values and hence represents a good model.
→ It is also called a coefficient of determination, or coefficient of multiple determination for multiple regression.
→ It can be calculated from the below formula:

Types of Linear Regression

Linear regression can be further divided into two types of the algorithm: •

i. **Simple Linear Regression**: If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

ii. **Multiple Linear regression**: If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression (pdf 16-22 {2 pdfs})

Advantages of supervised learning

1. The use of well-known and labelled input data makes supervised learning produce a far more accurate and reliable than unsupervised learning. With the access to labels, it can use to improve its performance on some task.
2. Supervised learning can be very helpful in classification problems.
3. Another typical task of supervised machine learning is to predict a numerical target value from some given data and labels.

Disadvantages of supervised learning

1. Supervised learning models are not suitable for handling the complex tasks.
2. Supervised learning cannot predict the correct output if the test data is different from the training dataset.
3. Training required lots of computation times. In supervised learning, we need enough knowledge about the classes of object.
4. Supervised learning cannot give you unknown information from the training data like unsupervised learning do.
5. It cannot cluster or classify data by discovering its features on its own, unlike unsupervised learning.

6. Similarly, let's say your training set does not include some examples that you want to have in a class. Then, when you use those examples after training, you might not get the correct class label as the output.

7. While you are training the classifier, you need to select a lot of good examples from each class. Otherwise, the accuracy of your model will be very less. This is difficult when you deal with a large amount of training data.

Advantages of unsupervised learning

1. There is lesser complexity compared to the supervised learning task. Here, no one is required to interpret the associated labels and hence it holds lesser complexities.
2. It is reasonably easier to obtain unlabeled data.

Disadvantages of unsupervised learning

1. It is not always certain that the obtained results will be useful since there is no label or output measure to confirm its usefulness.
2. Unsupervised learning is intrinsically more difficult than supervised learning as it does not have corresponding output.
3. The result of the unsupervised learning algorithm might be less accurate as input data is not labelled, and algorithms do not know the exact output in advance.

Advantages of reinforcement learning

1. Reinforcement learning can be used to solve very complex problems that cannot be solved by conventional techniques.
2. This technique is preferred to achieve long-term results, which are very difficult to achieve.
3. This learning model is very similar to the learning of human beings. Hence, it is close to achieving perfection.
4. The model can correct the errors that occurred during the training process.
5. Once an error is corrected by the model, the chances of occurring the same error are very less.
6. It can create the perfect model to solve a particular problem.
7. t can be useful when the only way to collect information about the environment is to interact with it.

Disadvantages of reinforcement learning

1. Too much reinforcement learning can lead to an overload of states, which can diminish the results.
2. Reinforcement learning is not preferable to use for solving simple problems.
3. Reinforcement learning needs a lot of data and a lot of computation. It is data-hungry.

2. Unsupervised Learning

- Unsupervised learning is a learning method in which a machine learns without any supervision. . Instead, models itself find the hidden patterns and insights from the given data.
- It can be compared to learning which takes place in the human brain while learning new things.
- The training is provided to the machine with the set of data that has not been labeled, classified, or categorized, and the algorithm needs to act on that data without any supervision. The training information is neither classified nor labeled; hence, a machine may not always provide correct output compared to supervised learning.
- In unsupervised learning, a machine is trained with some input samples or labels only, while output is not known.
- The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns.
- In unsupervised learning, we don't have a predetermined result. The machine tries to find useful insights from the huge amount of data.
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs
- It can be further classifieds into two categories of algorithms:
  - Clustering
  - Association

Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but no corresponding output data. The goal of unsupervised learning is to find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.

Example: Suppose the unsupervised learning algorithm is given an input dataset containing images of different types of cats and dogs. The algorithm is never trained upon the given dataset, which means it does not have any idea about the features of the dataset. The task of the unsupervised learning algorithm is to identify the image features on their own. Unsupervised learning algorithm will perform this task by clustering the image dataset into the groups according to similarities between images

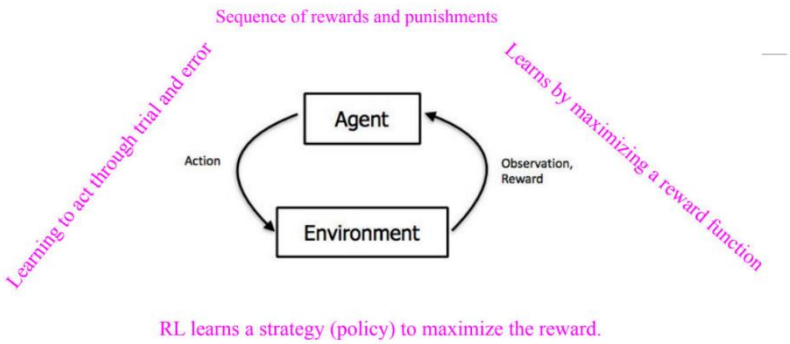
- Clustering:
  - Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group.
  - Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.
- Association:
  - An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database.
  - It determines the set of items that occurs together in the dataset.
  - Association rule makes marketing strategy more effective. Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item. A typical example of Association rule is Market Basket Analysis.

Below is the list of some popular unsupervised learning algorithms:

- K-means clustering
- KNN (k-nearest neighbors)
- Hierarchal clustering
- Anomaly detection
- Neural Networks
- Principle Component Analysis
- Independent Component Analysis
- Apriori algorithm
- Singular value decomposition

3. Reinforcement Learning

- Reinforcement learning is a feedback-based learning method.
- .In such type of learning, agents (computer programs) need to explore the environment, perform actions, and on the basis of their actions, they get rewards as feedback. Learning agent gets a reward for each right action and gets a penalty for each wrong action.



- The agent learns automatically with these feedbacks and improves its performance.
- In reinforcement learning, the agent interacts with the environment and explores it.
- The goal of an agent is to get the most reward points, and hence, it improves its performance.
- Since there is no labeled data, the agent is bound to learn by its experience only.
- 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.

4. Semi-supervised Learning

- Semi-supervised Learning is an intermediate technique of both supervised and unsupervised learning.
- It performs actions on datasets having few labels as well as unlabeled data. However, it generally contains unlabeled data. Hence, it also reduces the cost of the machine learning model as labels are costly, but for corporate purposes, it may have few labels.
- Further, it also increases the accuracy and performance of the machine learning model.
- Sem-supervised learning helps data scientists to overcome the drawback of supervised and unsupervised learning.
- Some important applications of Semi-supervised learning are:
  - Speech analysis,
  - web content classification

  - protein sequence classification
  - text documents classifiers

- Eg: Imagine that you are trying to build a system to guess a person's age from a photo.
  - You gather some labeled examples by snapping pictures of people and asking their age. That's supervised learning.
  - But in reality some of the people lied about their age. It's not just that there is random noise in the data; rather the inaccuracies are systematic, and to uncover them is an unsupervised learning problem involving images, self-reported ages, and true (unknown) ages.
  - Thus, both noise and lack of labels create a continuum between supervised and unsupervised learning

Applications of Machine Learning

Machine Learning is widely being used in approximately every sector, including healthcare, marketing, finance, infrastructure, automation, etc. There are some important real-world examples of machine learning, which are as follows:

1. Healthcare and Medical Diagnosis: Machine Learning is used in healthcare industries that help in generating neural networks. These self-learning neural networks help specialists for providing quality treatment by analyzing external data on a patient's condition, X-rays, CT scans, various tests, and screenings. Other than treatment, machine learning is also helpful for cases like automatic billing, clinical decision supports, and development of clinical care guidelines, etc.

2. Marketing: Machine learning helps marketers to create various hypotheses, testing, evaluation, and analyze datasets. It helps us to quickly make predictions based on the concept of big data. It is also helpful for stock marketing as most of the trading is done through bots and based on calculations from machine learning algorithms.

3. Self-driving cars: This is one of the most exciting applications of machine learning in today's world. Various automobile companies like Tesla, Tata, etc., are continuously working for the development of self-driving cars. It also becomes possible by the machine learning method (supervised learning), in which a machine is trained to detect people and objects while driving.

4. Speech Recognition: Nowadays, almost every mobile application comes with a voice search facility. This "Search By Voice" facility is also a part of speech recognition. In this method, voice instructions are converted into text, which is known as Speech to text" or "Computer speech recognition. Google assistant, SIRI, Alexa, etc., are some famous applications of speech recognition.

5. Traffic Prediction: Machine Learning also helps us to find the shortest route to reach our destination by using Google Maps. It also helps us in predicting traffic conditions, whether it is cleared or congested, through the real-time location of the Google Maps app and sensor.

6. Image Recognition: Image recognition is also an important application of machine learning for identifying objects, persons, places, etc. Face detection and auto friend tagging suggestion is the most famous application of image recognition used by Facebook, Instagram, etc. Whenever we upload photos with our Facebook friends, it automatically suggests their names through image recognition technology.

7. Product Recommendations: Machine Learning is widely used in business industries for the marketing of various products. Almost all big and small companies like Amazon, Alibaba, Walmart, Netflix, etc., are using machine learning techniques for products recommendation to their users. Whenever we search for any products on their websites, we automatically get started with lots of advertisements for similar products. This is also possible by Machine Learning algorithms that learn user's interests and, based on past data, suggest products to the user.

8. Automatic Translation: Automatic language translation is also one of the most significant applications of machine learning that is based on sequence algorithms by translating text of one language into other desirable languages. Google GNMT (Google Neural Machine Translation) provides this feature, which is Neural Machine Learning. Further, you can also translate the selected text on images as well as complete documents through Google Lens.

9. Virtual Assistant: A virtual personal assistant is also one of the most popular applications of machine learning. First, it records out voice and sends to cloud-based server then decode it with the help of machine learning algorithms. All big companies like Amazon, Google, etc., are using these features for playing music, calling someone, opening an app and searching data on the internet, etc.

10. Email Spam and Malware Filtering: Machine Learning also helps us to filter various Emails received on our mailbox according to their category, such as important, normal, and spam. It is possible by ML algorithms such as Multi-Layer Perceptron, Decision tree, and Naïve Bayes classifier.

Ockham's razor

- When there are multiple consistent hypotheses to solve a problem, the simpler one is to be preferred. This principle is called Ockham's razor.(pdf fig).
- This statement does not mean that simpler models are universally better than complex models, but rather that a model must be complex enough to learn the patterns in a dataset but simple enough to avoid overfitting.

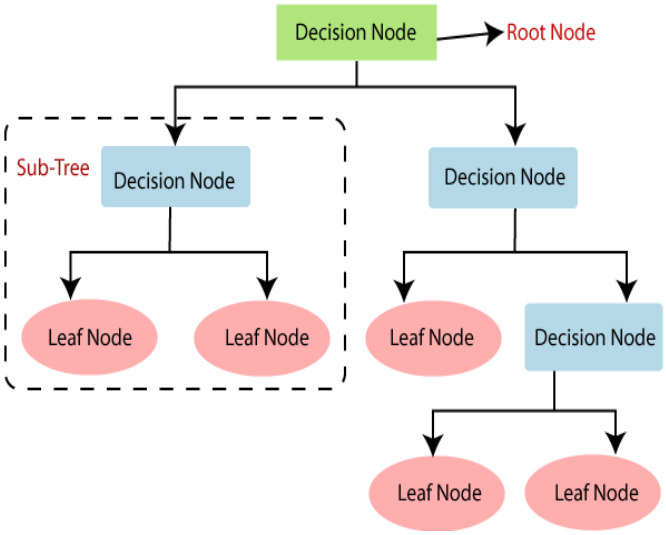
Overfitting in Machine Learning

Decision Tree Classification Algorithm

- Decision Tree is a **Supervised learning technique** that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where **internal nodes represent the features of a dataset, branches represent the decision rules** and **each leaf node represents the outcome.**
- In a Decision tree, there are two nodes, which are the **Decision Node** and **Leaf Node.** Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
- The decisions or the test are performed on the basis of features of the given dataset.
- ***It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.***
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- In order to build a tree, we use the **CART algorithm,** which stands for **Classification and Regression Tree algorithm.**
- A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

- Below diagram explains the general structure of a decision tree:

Note: A decision tree can contain categorical data (YES/NO) as well as numeric data.



Why use Decision Trees?

There are various algorithms in Machine learning, so choosing the best algorithm for the given dataset and problem is the main point to remember while creating a machine learning model. Below are the two reasons for using the Decision tree:

- Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.
- The logic behind the decision tree can be easily understood because it shows a tree-like structure.

Decision Tree Terminologies

- **Root Node:** Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.
- **Leaf Node:** Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.
- **Splitting:** Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.
- **Branch/Sub Tree:** A tree formed by splitting the tree.
- **Pruning:** Pruning is the process of removing the unwanted branches from the tree.
- **Parent/Child node:** The root node of the tree is called the parent node, and other nodes are called the child nodes.

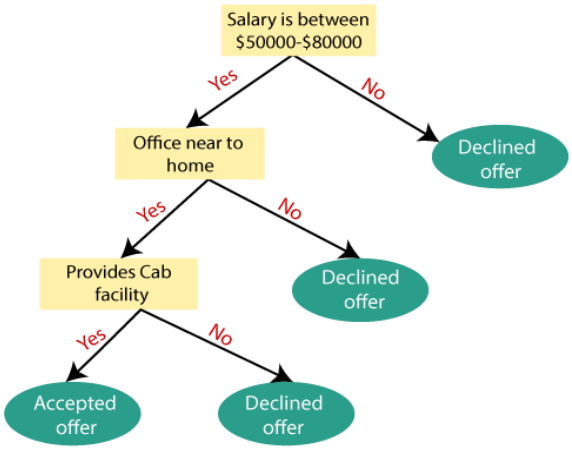
How does the Decision Tree algorithm Work?

In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node.

For the next node, the algorithm again compares the attribute value with the other sub-nodes and move further. It continues the process until it reaches the leaf node of the tree. The complete process can be better understood using the below algorithm:

- **Step-1:** Begin the tree with the root node, says S, which contains the complete dataset.
- **Step-2:** Find the best attribute in the dataset using **Attribute Selection Measure (ASM).**
- **Step-3:** Divide the S into subsets that contains possible values for the best attributes.
- **Step-4:** Generate the decision tree node, which contains the best attribute.
- **Step-5:** Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Example: Suppose there is a candidate who has a job offer and wants to decide whether he should accept the offer or Not. So, to solve this problem, the decision tree starts with the root node (Salary attribute by ASM). The root node splits further into the next decision node (distance from the office) and one leaf node based on the corresponding labels. The next decision node further gets split into one decision node (Cab facility) and one leaf node. Finally, the decision node splits into two leaf nodes (Accepted offers and Declined offer). Consider the below diagram:



Attribute Selection Measures

While implementing a Decision tree, the main issue arises that how to select the best attribute for the root node and for sub-nodes. So, to solve such problems there is a technique which is called as **Attribute selection measure or ASM.** By this measurement, we can easily select the best attribute for the nodes of the tree. There are two popular techniques for ASM, which are:

- **Information Gain**
- **Gini Index**

1. Information Gain:

- Information gain is the measurement of changes in entropy after the segmentation of a dataset based on an attribute.
- It calculates how much information a feature provides us about a class.
- According to the value of information gain, we split the node and build the decision tree.
- A decision tree algorithm always tries to maximize the value of information gain, and a node/attribute having the highest information gain is split first. It can be calculated using the below formula:

1. Information Gain= Entropy(S)- [(Weighted Avg) *Entropy(each feature)

Entropy: Entropy is a metric to measure the impurity in a given attribute. It specifies randomness in data. Entropy can be calculated as:

```
Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)
```

Where,

- **S= Total number of samples**
- **P(yes)= probability of yes**
- **P(no)= probability of no**

2. Gini Index:

- Gini index is a measure of impurity or purity used while creating a decision tree in the CART(Classification and Regression Tree) algorithm.
- An attribute with the low Gini index should be preferred as compared to the high Gini index.

- It only creates binary splits, and the CART algorithm uses the Gini index to create binary splits.
- Gini index can be calculated using the below formula:

$$\text{Gini Index} = 1 - \sum_j P_j^2$$

Pruning: Getting an Optimal Decision tree

Pruning is a process of deleting the unnecessary nodes from a tree in order to get the optimal decision tree.

A too-large tree increases the risk of overfitting, and a small tree may not capture all the important features of the dataset. Therefore, a technique that decreases the size of the learning tree without reducing accuracy is known as Pruning. There are mainly two types of tree **pruning** technology used:

- **Cost Complexity Pruning**
- **Reduced Error Pruning.**

Advantages of the Decision Tree

- It is simple to understand as it follows the same process which a human follow while making any decision in real-life.
- It can be very useful for solving decision-related problems.
- It helps to think about all the possible outcomes for a problem.
- There is less requirement of data cleaning compared to other algorithms.

Disadvantages of the Decision Tree

- The decision tree contains lots of layers, which makes it complex.
- It may have an overfitting issue, which can be resolved using the **Random Forest algorithm.**
- For more class labels, the computational complexity of the decision tree may increase.