Decision tree knowledge discovery through neural Networks

Presented by: Jaskaran Kaur

Decision tree

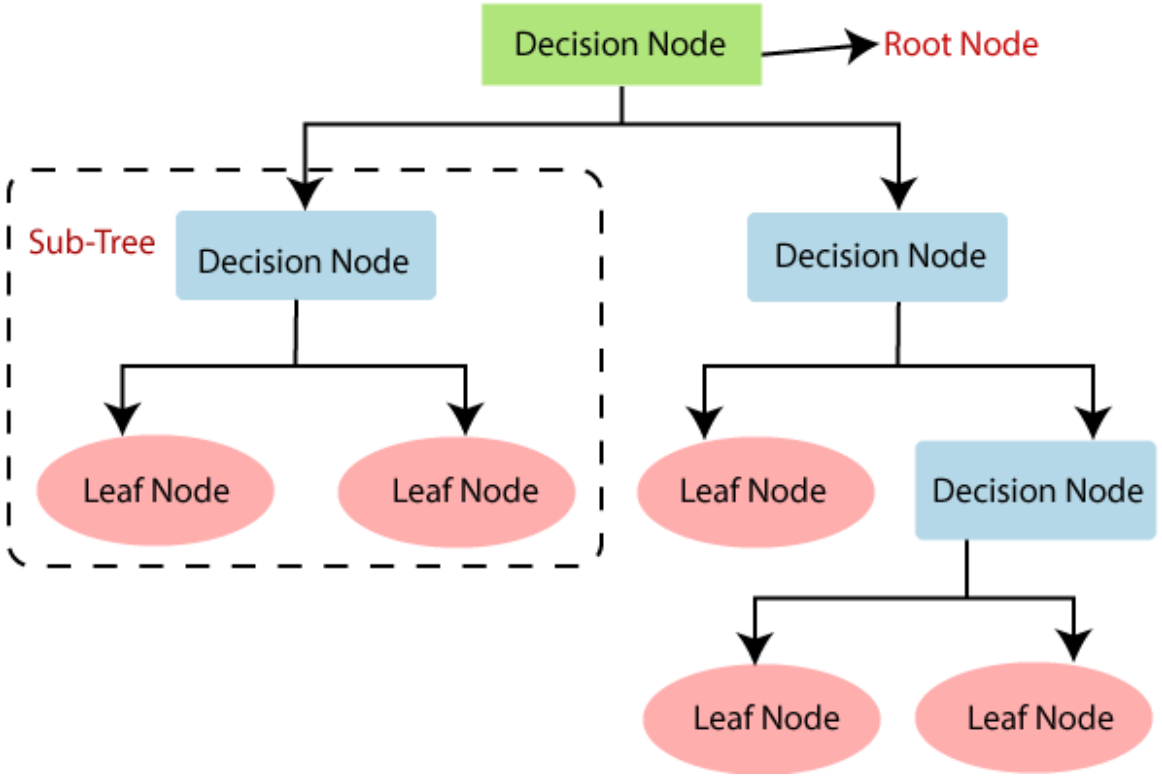
- Decision Tree Mining is a type of data mining technique that is used to build Classification Models.
- It builds classification models in the form of a tree-like structure, just like its name. This type of mining belongs to supervised class learning.
- In supervised learning, the target result is already known.
- Decision trees can be used for both categorical and numerical data.
- The categorical data represent gender, marital status, etc. while the numerical data represent age, temperature, etc.

Continued…….

- 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.
- 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.

General structure of a decision tree:



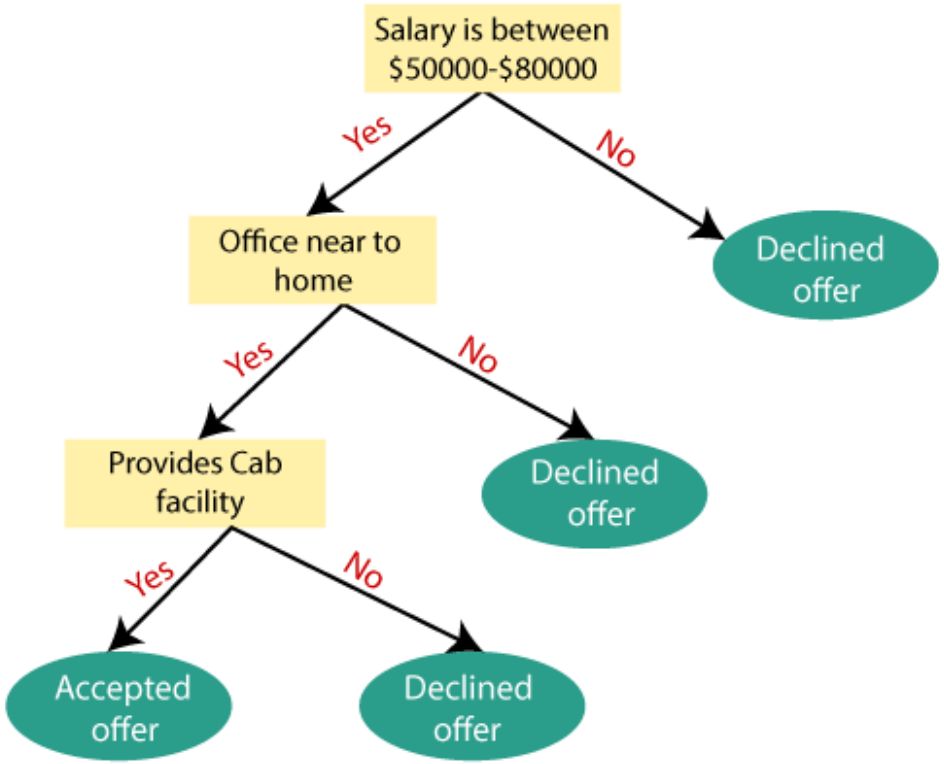
Decision Tree Terminologies

- **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.

Example: Suppose there is a candidate who has a job offer and wants to decide whether he should accept the offer or Not.



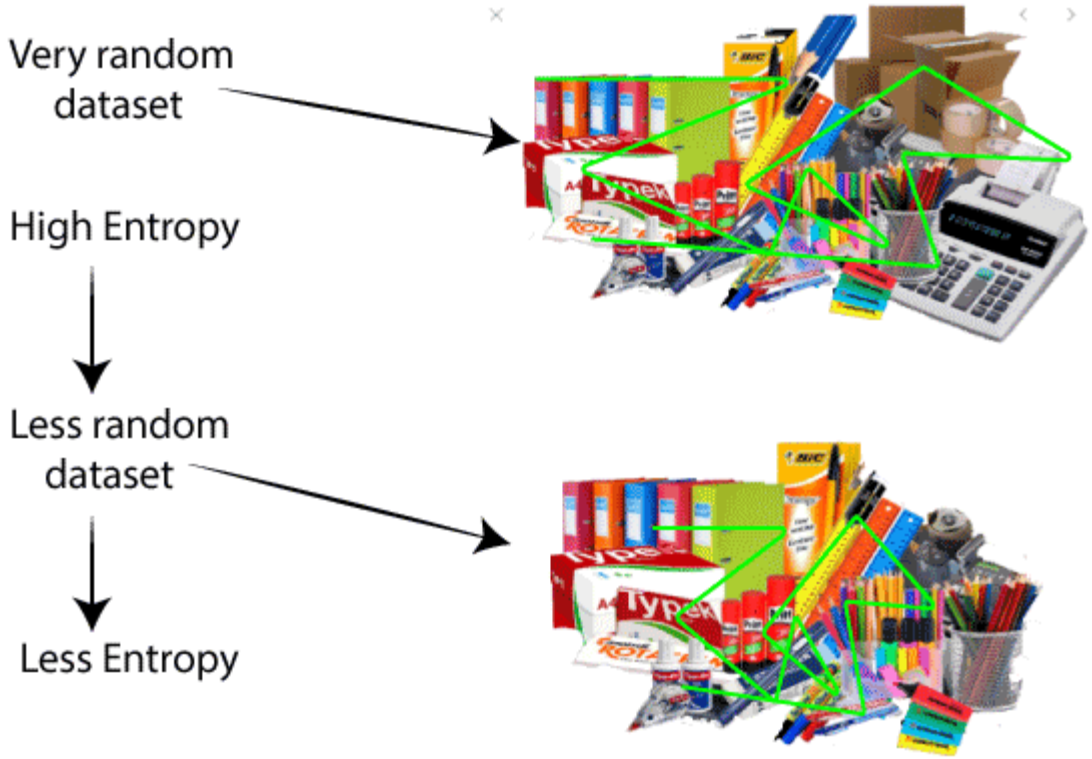
Attribute Selection Measures

- 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.

Key factors:

1. Entropy:

- Entropy refers to a common way to measure impurity. In the decision tree, it measures the randomness or impurity in data sets.



Formula:

- 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

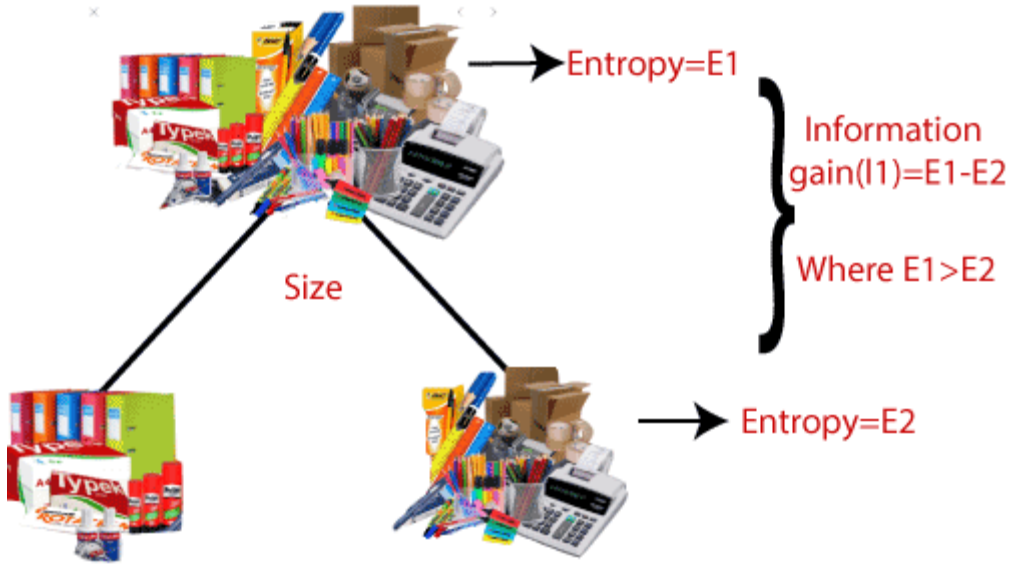
2. Information Gain:

- 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.
- It is also called **Entropy Reduction**.
- 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.

Formula:

- It can be calculated using the below formula:

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



Relation between Entropy and Information Gain:

- Inverse
- If information gain increases entropy decreases and vice versa.

Use:

- It is used to represent rules or generate rules
- Easy to understand.
- Visually define the rules which are simple to interpretation and understand.

Neural network

- An artificial neural network (ANN), usually called "neural network" (NN), is a mathematical model or computational model that tries to simulate the structure and/or functional aspects of biological neural networks.

Neural Network

- The term "**Artificial Neural Network**" is derived from Biological neural networks that develop the structure of a human brain.
- Similar to the human brain that has neurons interconnected to one another, artificial neural networks also have neurons that are interconnected to one another in various layers of the networks.
- These neurons are known as nodes.
- An **Artificial Neural Network** in the field of **Artificial intelligence** where it attempts to mimic the network of neurons makes up a human brain So that computers will have an option to understand things and make decisions in a human-like manner.

Continued….

- The artificial neural network is designed by programming computers to behave simply like interconnected brain cells.
- There are around 1000 billion neurons in the human brain. Each neuron has an association point somewhere in the range of 1,000 and 100,000.
- In the human brain, data is stored in such a manner as to be distributed, and we can extract more than one piece of this data when necessary from our memory parallelly.
- We can say that the human brain is made up of incredibly amazing parallel processors.

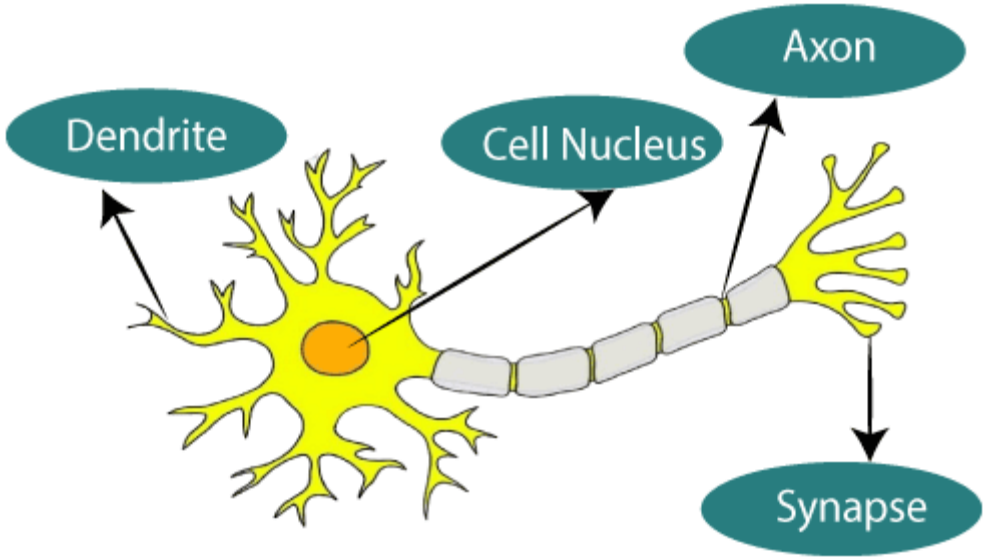
Example:

We can understand the artificial neural network with an example,

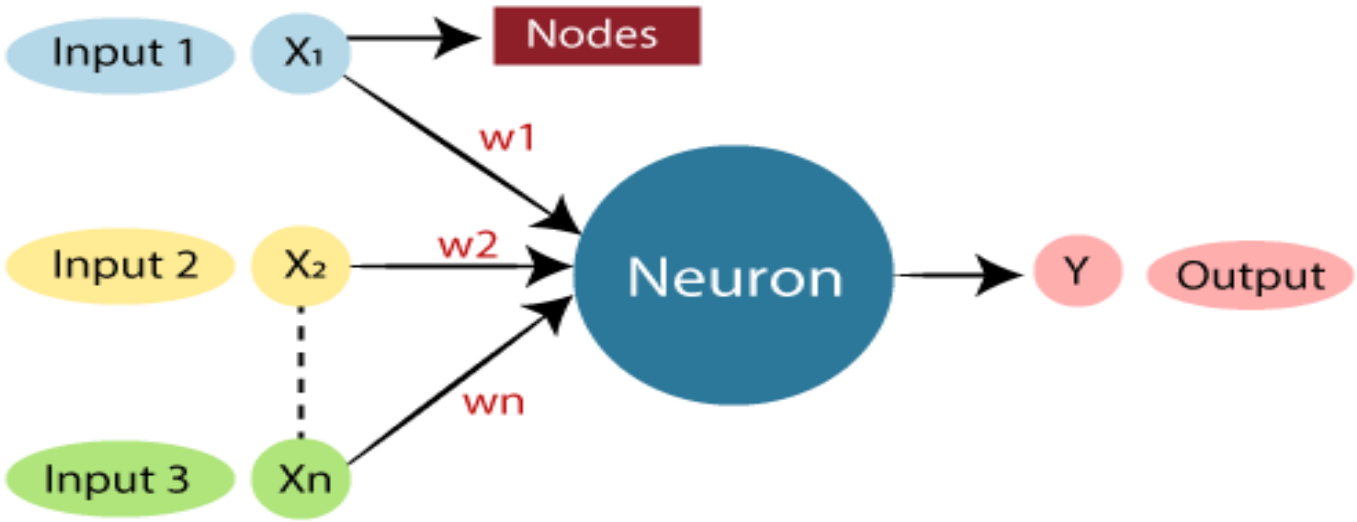
- Consider an example of a digital logic gate that takes an input and gives an output.
- "OR" gate, which takes two inputs. If one or both the inputs are "On," then we get "On" in output.
- If both the inputs are "Off," then we get "Off" in output.
- Here the output depends upon input. Our brain does not perform the same task.
- The outputs to inputs relationship keep changing because of the neurons in our brain, which are "learning.“

*Artificial neural networks are trained using a training set.

Diagram of Biological Neural Network



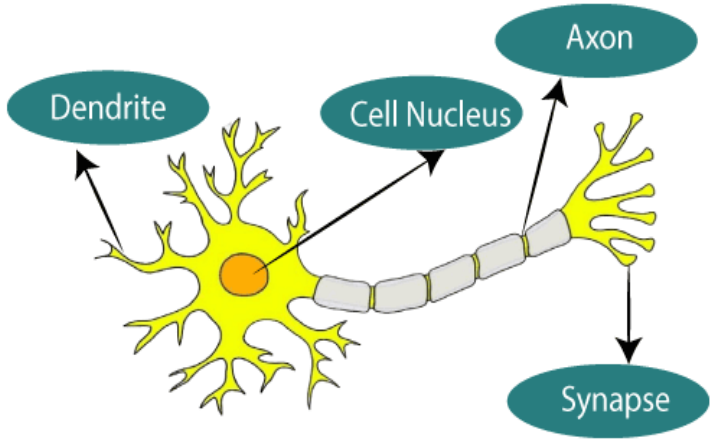
The typical Artificial Neural Network looks something like the given figure.

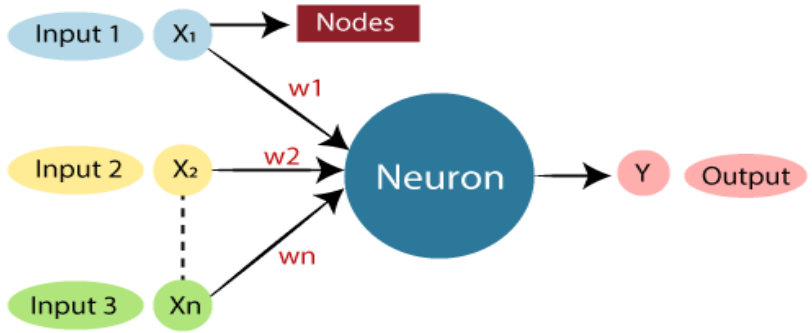


Dendrites from Biological Neural Network represent inputs in Artificial Neural Networks, cell nucleus represents Nodes, synapse represents Weights, and Axon represents Output.

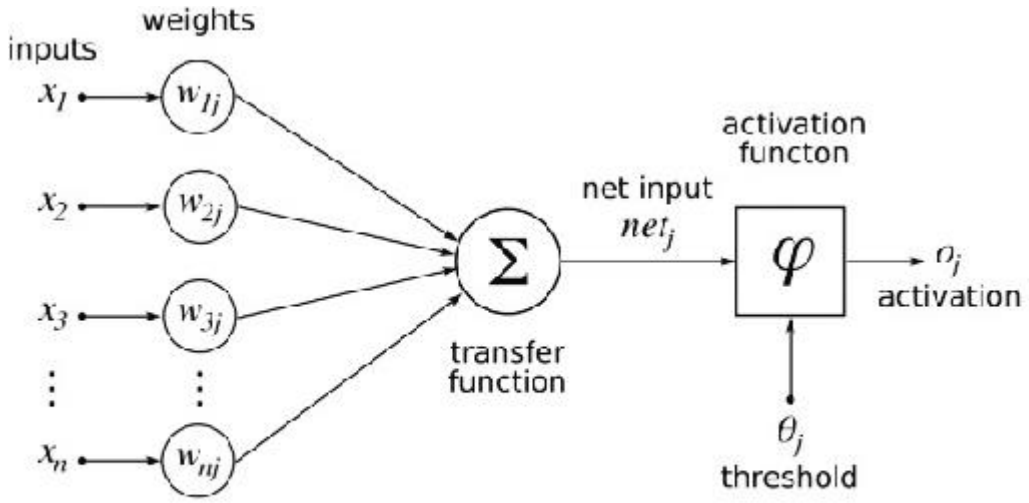
Relationship between Biological neural network and artificial neural network:

Biological Neural Network	Artificial Neural Network
Dendrites	Inputs
Cell nucleus	Nodes
Synapse	Weights
Axon	Output

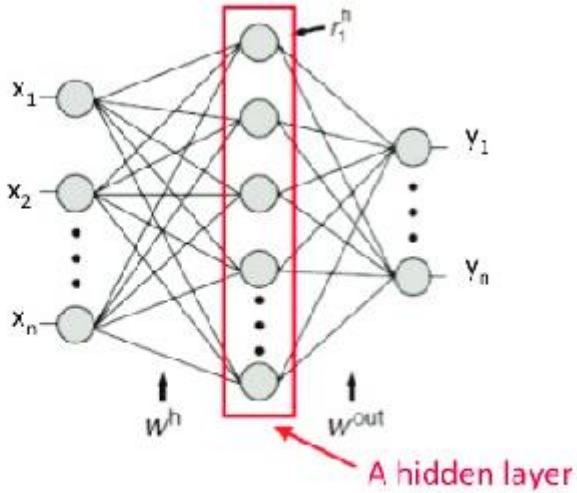




Neaural Network Components:

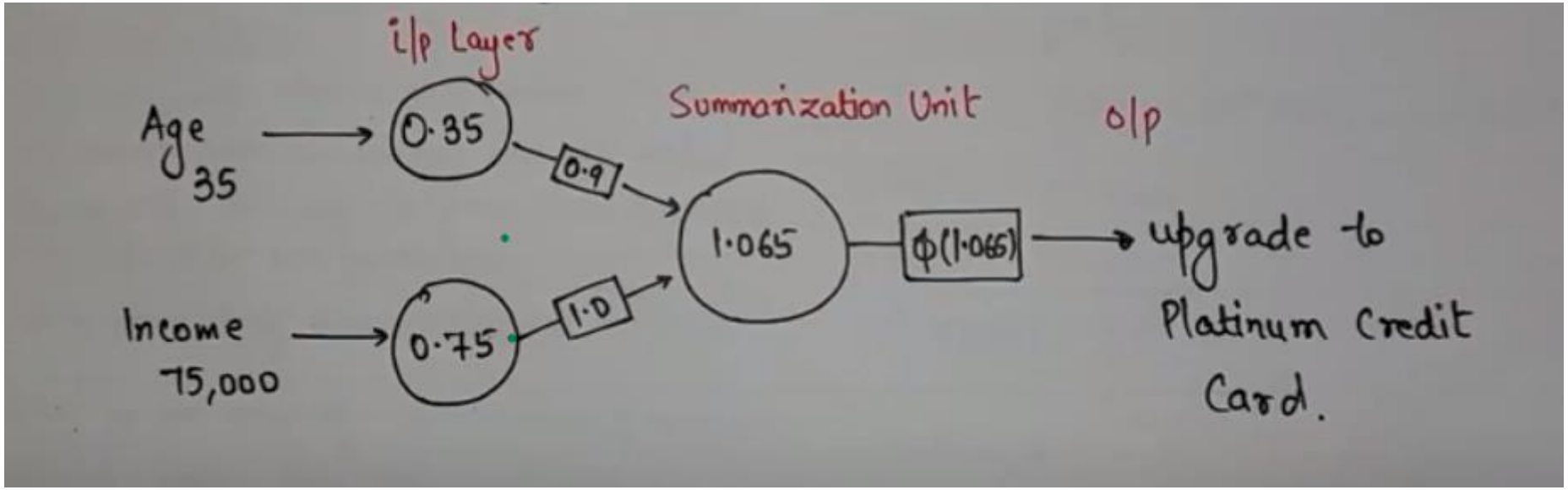


(a)



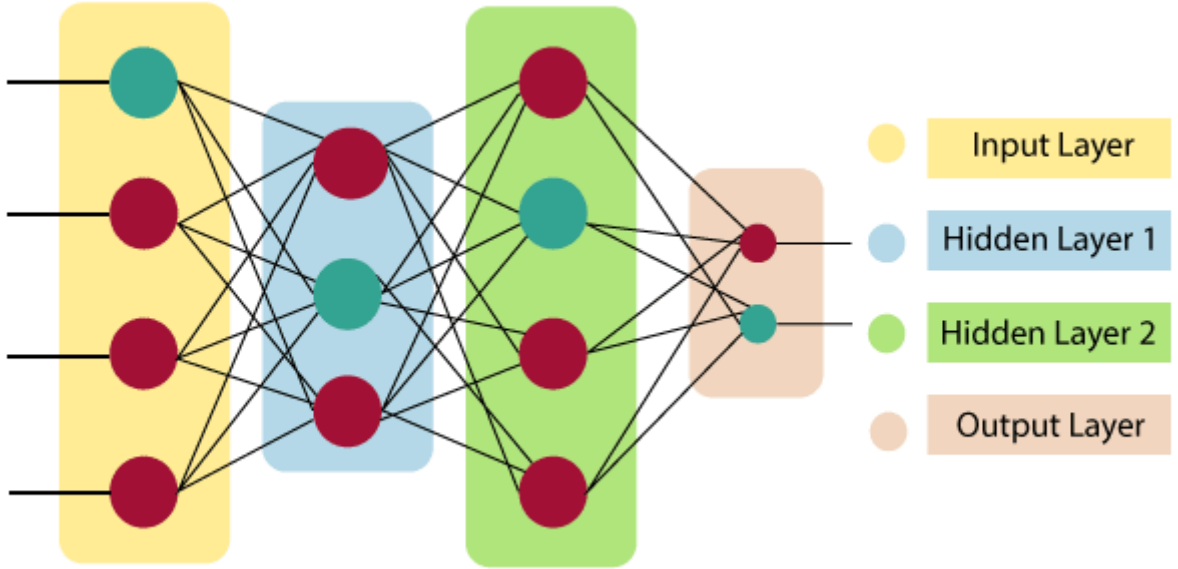
(b)

Example


i/p Layer
Summarization Unit
o/p
Age
35
0·35
0·9
Income
75,000
0·75
1·0
1·065
ϕ(1·065)
upgrade to
Platinum Credit
Card.

The architecture of an artificial neural network:

- Artificial Neural Network primarily consists of three layers:



- **Input Layer:**

As the name suggests, it accepts inputs in several different formats provided by the programmer.

- **Hidden Layer:**

The hidden layer presents in-between input and output layers. It performs all the calculations to find hidden features and patterns.

- **Output Layer:**

The input goes through a series of transformations using the hidden layer, which finally results in output that is conveyed using this layer.

In other words,

- The input layer receives data from the outside world which the neural network needs to analyze or learn about.
- Then this data passes through one or multiple hidden layers that transform the input into data that is valuable for the output layer.
- Finally, the output layer provides an output in the form of a response of the Artificial Neural Networks to input data provided.

Advantages of Artificial Neural Network (ANN)

- **Parallel processing capability:** Artificial neural networks have a numerical value that can perform more than one task simultaneously.
- **Storing data on the entire network:** Data that is used in traditional programming is stored on the whole network, not on a database. The disappearance of a couple of pieces of data in one place doesn't prevent the network from working.
- **Capability to work with incomplete knowledge:** After ANN training, the information may produce output even with inadequate data. The loss of performance here relies upon the significance of missing data.

Applications of Artificial Neural Networks

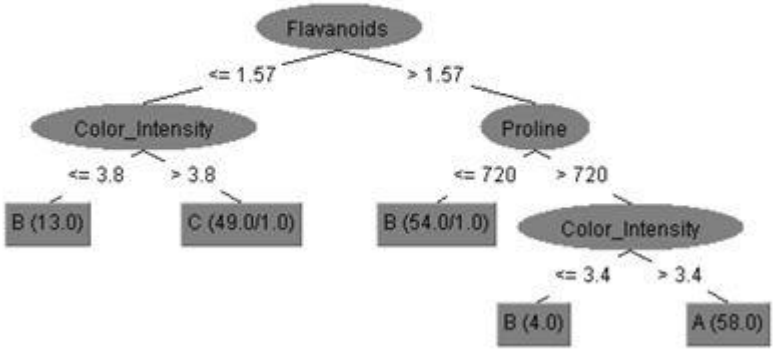
- Social Media
- Marketing and Sales
- Healthcare
- Personal Assistants

Decision Trees vs. Neural Networks

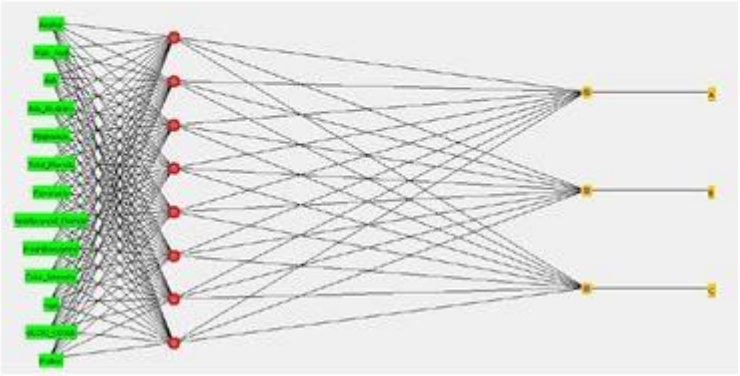
- Two popular data modeling techniques are Decision Trees, also called classification trees and Neural Networks.
- These two data modeling techniques are very different from the way they look to the way they find relationship within variables.
- The neural network is an assembly of nodes, looks somewhat like the human brain.
- While the decision tree is an easy to follow top down approach of looking at the data.
- Neural networks are often compared to decision trees because both methods can model data that has nonlinear relationships between variables, and both can handle interactions between variables. However, neural networks have a number of drawbacks compared to decision trees.

Diagram

Decision Tree



Neural Network



Decision Trees

- Decision trees have an easy to follow natural flow. They are also easy to program for computer systems with IF, THEN, ELSE statements.
- We can see that the top node in the tree is the most influential piece of data that affects the response variable in the model. Because these trees are so easy to understand, they are very useful as modeling techniques and provide visual representations of the data.

Neural Networks

- The neural network is not so easy to understand from the visual representation. It is very difficult to create computer systems from them, and almost impossible to create an explanation from the model. Neural networks can handle binary data better than decision trees but cannot handle categorical values.
- Neural Network achieve 99% accuracy on a data set while the decision tree model only achieved 86% accuracy on the same data set. The best fitted model is the one that most accurately fits your data.

Thank You