Machine Learning

Practical File

BACHELOR OF TECHNOLOGY

Information Technology

SUBMITTED BY

PRABHJOT SINGH

University Roll no. 2104546

Class Roll no.2121088

GURU NANAK DEV ENGINEERING COLLEGE

LUDHIANA-141006, INDIA

- **INTRODUCTION TO PYTHON LIBRARIES**

1 NUMPY – Numpy stands for numerical Python. It is a library which deals with n-dimensional array. It is used for numeric and scientific operation. It serves as a building Numpy stands for numerical Python. It is a block for many libraries available in Python.

PROGRAM :

```
import numpy as np
c = np.array ([(1,2,3),(4,5,6)])
print (c.shape)
```

Output –

```
(2, 3)
```

2 PANDAS– Pandas is an open source Python package that is most widely used for data science, data analysis and machine learning tasks. It is built on top of another package named Numpy. Using pandas data can be cleanes, transformed, manipulated and analysed. It is suitable for tabular form of data.

PROGRAM

```
 import pandas as pd
S = pd.Series([11,28,72,3,5,8])
print (S.index)
print(S.values)
```

Output

```
RangeIndex(start=0, stop=6, step=1)
[11 28 72  3  5  8]
```

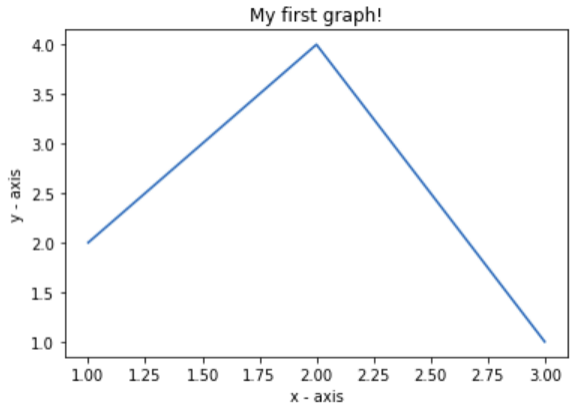
3 MATPLOT LIB – Matplotlib is a cross-platform , data visualization and graphical plotting for Python and its numerical extension Numpy. It can be used in python scripts, shells, web application servers and other graphical user interface toolkits.

PROGRAM

```
import matplotlib.pyplot as plt


x = [1,2,3]

y = [2,4,1]


plt.plot(x, y)


plt.xlabel('x - axis')

plt.ylabel('y - axis')
```

```
plt.title('My first graph!'
```

OUTPUT



4 SCIKIT-LEARN – Scikit-learn (Sklearn) is the most useful and robust library for machine learning in python. It provides a selection of efficient modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library ,which is largely written in Python , is built upon Numpy ,Scipy and Mataplotlib.

- **DEFINITION OF MACHINE LEARNING (ML)**

Machine Learning is a branch of artificial intelligence that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so .

➢ TYPES OF MACHINE LEARNING

- **SUPERVISED LEARNING**
- **UNSUPERVISED LEARNING**
- **SEMISUPERVISED LEARNING**

• Supervised learning

➢ Input data is labeled
➢ Has a feedback mechanism
➢ Data is classified based on the training dataset
➢ Divide into regression and classification
➢ Used for prediction
➢ Algorithms include – decision trees, logistic regressions , support vector machine
➢ A known number of classes
➢ Supervised learning model produces an accurate result
➢ It is not truly close to artificial intelligence

- **Unsupervised learning**

- Input data is unlabeled
- Has no feedback
- Assigns properties of given data to classify it
- Divided into clustering and association
- Used for analysis
- Algorithms include- k-means clustering, hierarchical clustering, apriori algorithm
- A unknown number of classes
- Give less accurate result as compare to supervised learning
- It is more close to the true artificial intelligence

- **MLvs traditional learning**

- Machine Learning
- ML algorithms do not depend on rules defined by humans experts. Instead ,they process data in raw form-for eg: text, emails, documents ,social media content, etc.
- An ML system is truly a learning system if it is not programmed to perform a task ,but is programmed to learn to perform the task.

- **Traditinal programmimg**

- Traditional programming is manual process , meaning a person creates a program.
- But without anyone programming the logic , one has to manually formulate or code rules.

• Advantages of Machine Learning

- Effiecient handling of data
- Automation for everything
- Continuous improvement
- Handling multidimensional and multi-variety data
- Wide applications

• Advantages of AI

- It handles the information better than humans
- Helpful for the conversion of information into knowledge
- Improve the work efficiency to reduce time duration
- Introduces a new technique to solve new problems

• Applications of machine learning

- Product recommendations

- Self driving cars
- Google translate
- Fraud detection
- Dynamic pricing
- Social media
- Virtual personal assistants

- **Applications of AI**

- Voice detection
- Autonomous vehicles
- Google maps
- Chatbots
- Health care
- Smart speaker
- Music and media streamning services

PRACTICAL-1

To print some message in python program

Implementation/Code

```
print (" Hello World")
```

Output:

Hello World

PRACTICAL-2

To declare and print different variable and type

Implementation/Code

```
a = 10
b = "hello world"
c =7.7
print (a,b,c)
print (type (a) ,type (b) ,type (c) )
```

OUTPUT:

```
10 hello world 7.7
<class 'int'>  <class 'str'>  <class 'float'>
```

REGRESSION

Regression is a technique for investigating the relationship between independent variables or features and a dependent variable or outcome. It's used as a method for predictive modelling in machine learning, in which an algorithm is used to predict continuous outcomes.

TYPES OF REGRESSION:

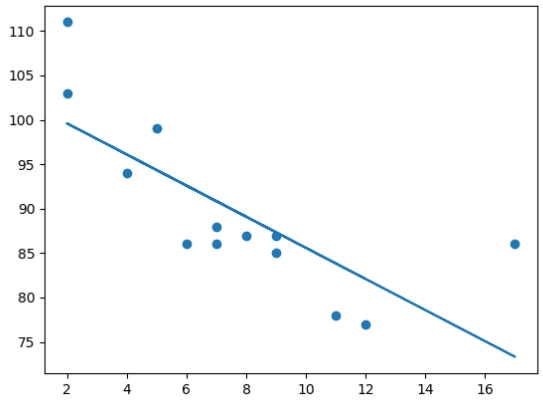
Some of the most common regression techniques in machine learning can be grouped into the following types of regression analysis:

- Simple Linear Regression
- Multiple linear regression
- Logistic regression

LINEAR REGRESSION

Linear regression uses the relationship between the data-points to draw a straight line through all them.

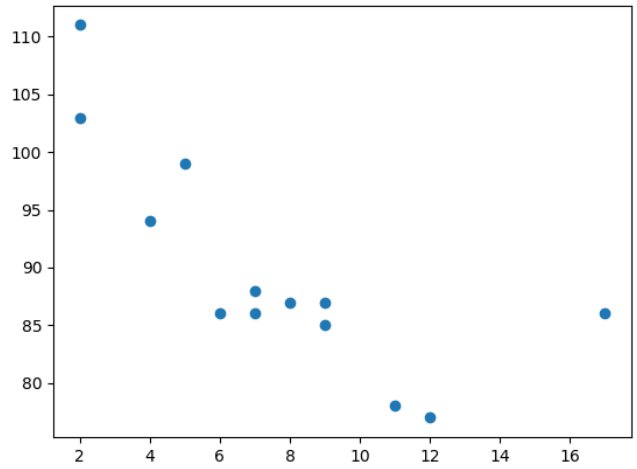
This line can be used to predict future values.



PROGRAM : 6

```
import matplotlib.pyplot as plt
x = [5,7,8,7,2,17,2,9,4,11,12,9,6]
y = [99,86,87,88,111,86,103,87,94,78,77,85,86]

plt.scatter(x, y)
plt.show()
```

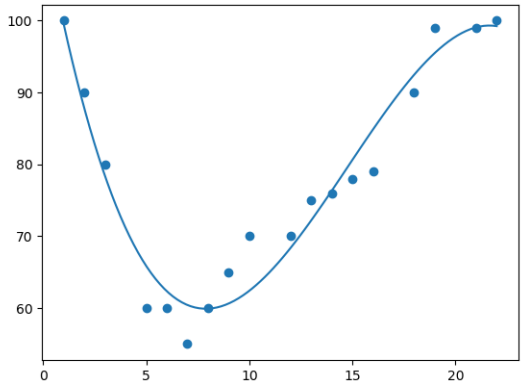
OUTPUT



Polynomial Regression

If your data points clearly will not fit a linear regression (a straight line through all data points), it might be ideal for polynomial regression.

Polynomial regression, like linear regression, uses the relationship between the variables x and y to find the best way to draw a line through the data points.



Multiple Regression

Multiple regression is like linear regression, but with more than one independent value, meaning that we try to predict a value based on **two or more** variables.

PROGRAM:8

```
import pandas
from sklearn import linear_model

df = pandas.read_csv("cars.csv")

X = df[['Weight', 'Volume']]
y = df['CO2']

regr = linear_model.LinearRegression()
regr.fit(X, y)

predictedCO2 = regr.predict([[2300, 1300]])

print(predictedCO2)
```

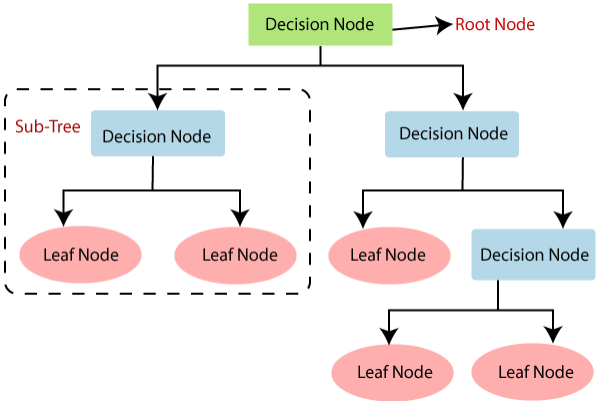
OUTPUT

```
[107.2087328]
```

.

DECISION TREE

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.**



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

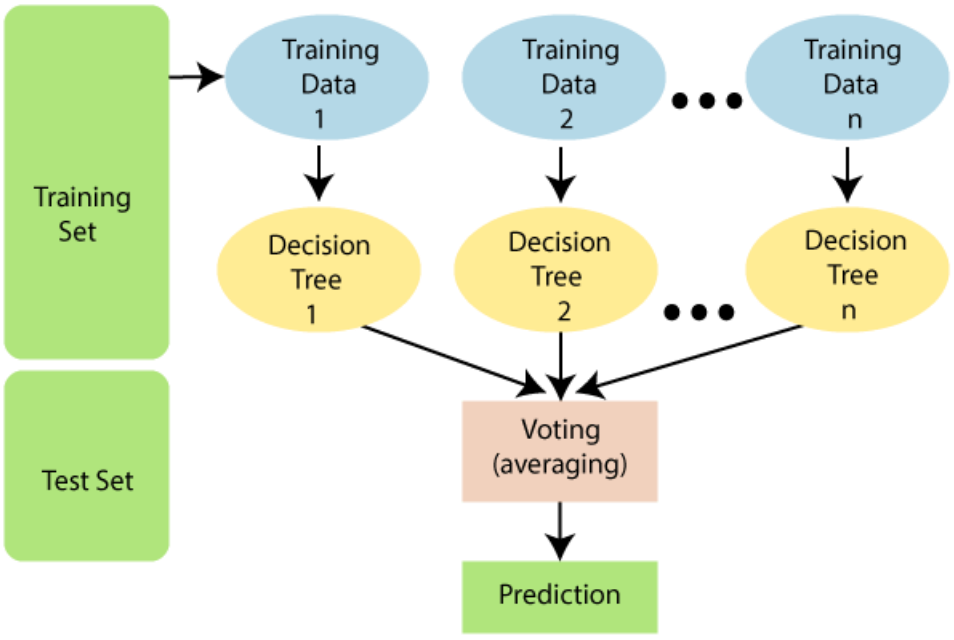
- The decision tree contains lots of layers, which makes it complex.
- It may have an overfitting issue, which can be resolved using the **Random Forest algorithm.**
- For more class labels, the computational complexity of the decision tree may increase.

PROGRAM 9

```
import pandas
from sklearn import tree
import pydotplus
from sklearn.tree import DecisionTreeClassifier
import matplotlib.pyplot as plt
import matplotlib.image as pltimg

df = pandas.read_csv("shows.csv")

print(df)
```

Random Forest : Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model.*

As the name suggests, **"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.



Advantages of Random Forest

- It is capable of handling large datasets with high dimensionality.
- It enhances the accuracy of the model and prevents the overfitting issue.

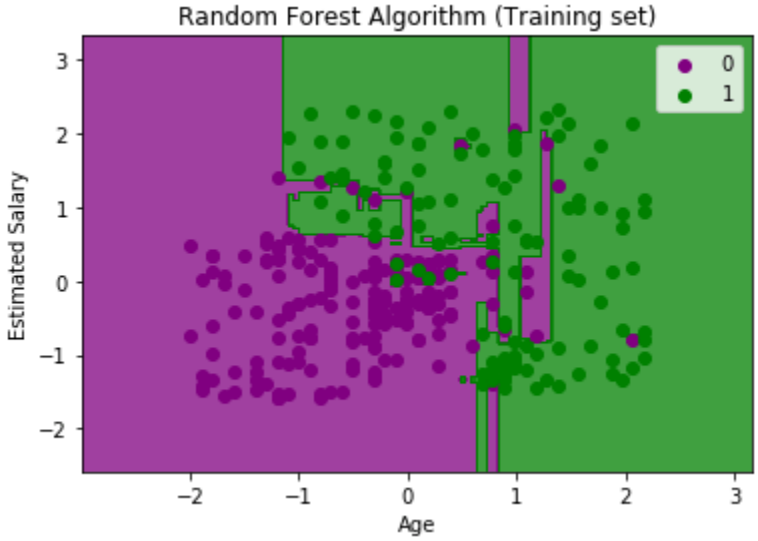
Disadvantages of Random Forest

- Although random forest can be used for both classification and regression tasks, it is not more suitable for Regression tasks.

PROGRAM 10

```
 import ListedColormap
x_set, y_set = x_train, y_train
x1, x2 = nm.meshgrid(nm.arange(start = x_set[:, 0].min() -
 1, stop = x_set[:, 0].max() + 1, step  =0.01),
nm.arange(start = x_set[:, 1].min() -
 1, stop = x_set[:, 1].max() + 1, step = 0.01))
mtp.contourf(x1, x2, classifier.predict(nm.array([x1.ravel(), x2.ravel()]).T).resha
pe(x1.shape),
alpha = 0.75, cmap = ListedColormap(('purple','green' )))
mtp.xlim(x1.min(), x1.max())
mtp.ylim(x2.min(), x2.max())
for i, j in enumerate(nm.unique(y_set)):
    mtp.scatter(x_set[y_set == j, 0], x_set[y_set == j, 1],
        c = ListedColormap(('purple', 'green'))(i), label = j)
mtp.title('Random Forest Algorithm (Training set)')
mtp.xlabel('Age')
mtp.ylabel('Estimated Salary')
mtp.legend()
mtp.show()
```

OUTPUT

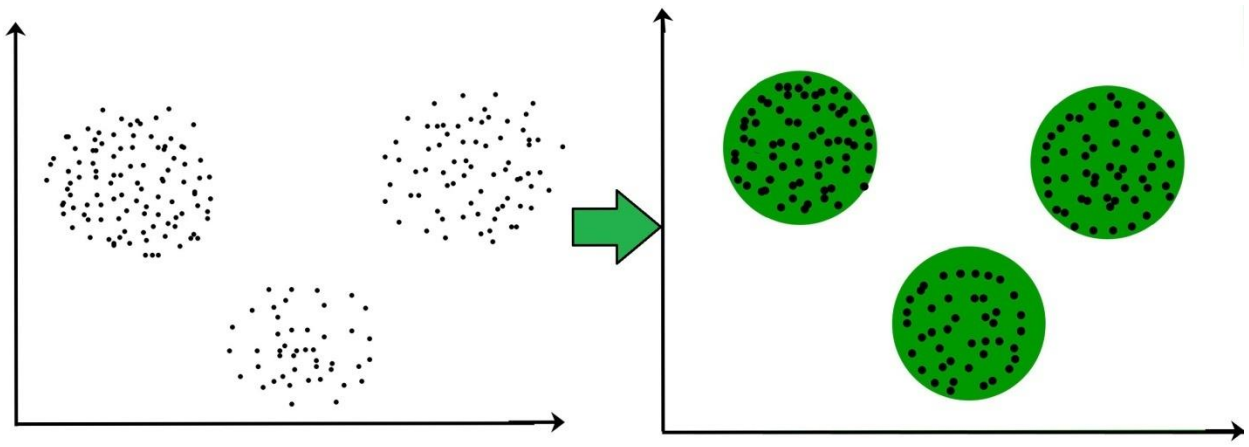


Clustering

It is basically a type of *unsupervised learning method*. An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labeled responses. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent in a set of examples.

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 ex– 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.



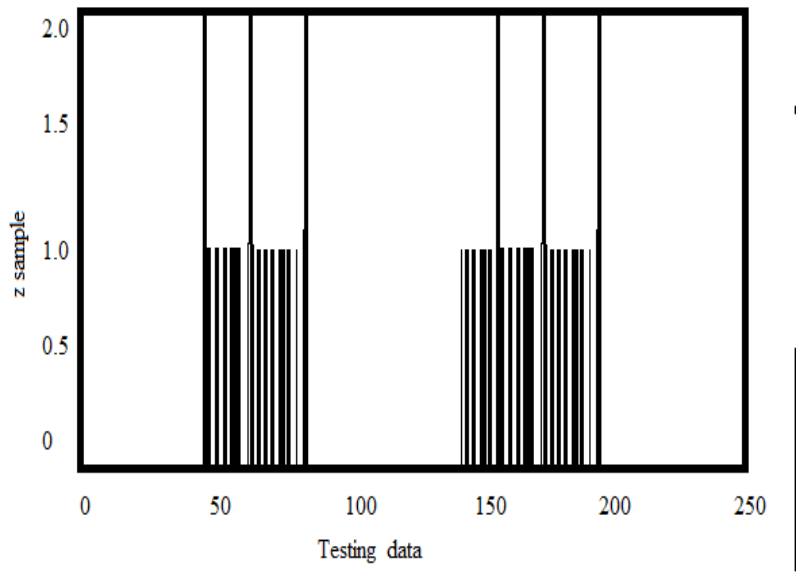
Applications of Clustering in different fields

- **Marketing:** It can be used to characterize & discover customer segments for marketing purposes.
- **Biology:** It can be used for classification among different species of plants and animals.
- **Libraries:** It is used in clustering different books on the basis of topics and information.
- **Insurance:** It is used to acknowledge the customers, their policies and identifying the frauds.

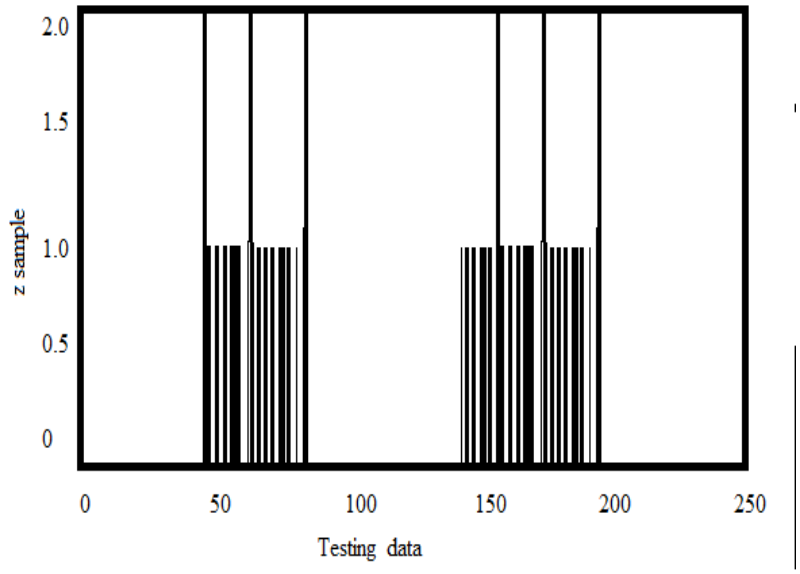
PROGRAM 11

```
import numpy as np

import pyplot as plt


X = np.random.randint(10,35,(25,2))
Y = np.random.randint(55,70,(25,2))
Z = np.vstack((X,Y))
Z = Z.reshape((50,2))

Z = np.float32(Z)

plt.xlabel('Test Data')
plt.ylabel('Z samples')



plt.hist(Z,256,[0,256])


plt.show()
```

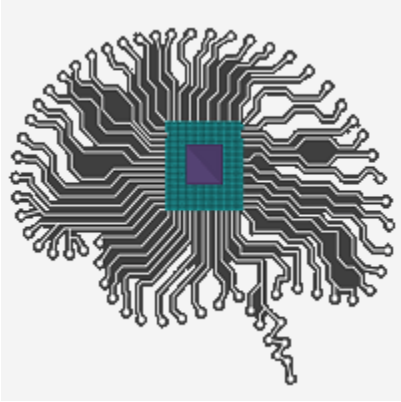
OUTPUT



Output:



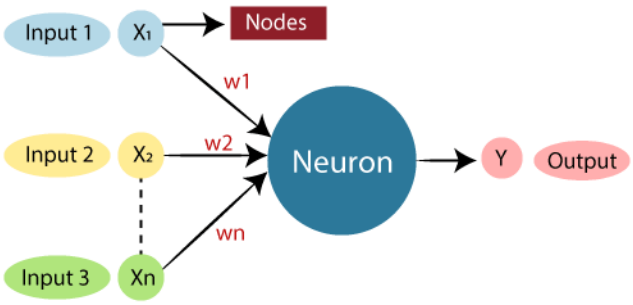
Artificial Neural Network Tutorial



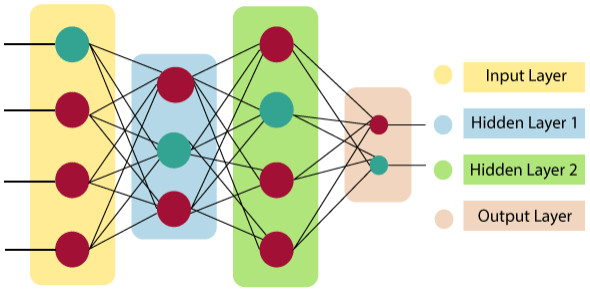
Artificial Neural Network Tutorial provides basic and advanced concepts of ANNs. Our Artificial Neural Network tutorial is developed for beginners as well as professions.

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 ANN.

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



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.

The artificial neural network takes input and computes the weighted sum of the inputs and includes a bias. This computation is represented in the form of a transfer function.