PART A

(PART A: TO BE REFFERED BY STUDENTS)

Experiment No. 1

A.1 Aim:

To implement Logistic Regression.

A.2 Prerequisite:

Python Basic Concepts

A.3 Outcome:

Students will be able to implement Logistic Regression.

A.4 Theory:

Machine Learning, being a subset of Artificial Intelligence (AI), has been playing a dominant role in our daily lives. Data science engineers and developers working in various domains are widely using machine learning algorithms to make their tasks simpler and life easier.

Types of supervised learning

As supervised learning is used to classify something or predict a value, naturally there are two types of algorithms for supervised learning - *classification models and regression models.*

1. **Classification model** - In simple terms, a classification model predicts possible outcomes. *Example: Predicting if a transaction is fraud or not.*
2. **Regression model** - Are used to predict a numerical value. *Example: Predicting the sale price of a house.*

Logistic regression

Logistic regression is an example of supervised learning. It is used to calculate or predict the probability of a binary (yes/no) event occurring. An example of logistic regression could be applying machine learning to determine if a person is likely to be infected with COVID-19 or not. Since we have two possible outcomes to this question - yes they are infected, or no they are not infected - this is called *binary classification.*

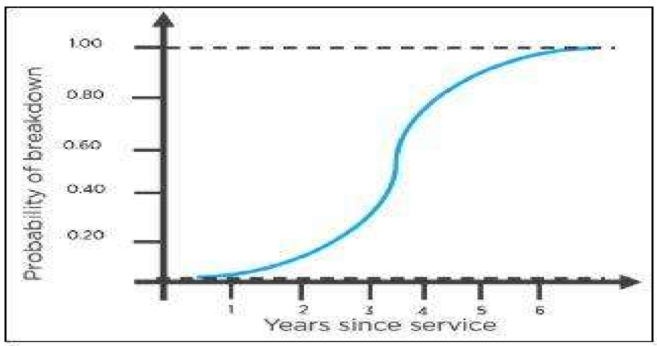
Sigmoid function, which produces an S-shaped curve. It always returns a probability value between 0 and 1. The Sigmoid function is used to convert expected values to probabilities. The function converts any real number into a number between 0 and 1. We utilize sigmoid to translate predictions to probabilities in machine learning.The mathematically sigmoid function can be

$$f(x) = \frac{1}{1 + e^{-(x)}}$$

,

Where to use logistic regression

- In health care, logistic regression can be used to predict if a tumor is likely to be benign or malignant.
- In the financial industry, logistic regression can be used to predict if a transaction is fraudulent or not.
- In marketing, logistic regression can be used to predict if a targeted audience will respond or not.



Advantages of the Logistic Regression Algorithm

- Logistic regression performs better when the data is linearly separable
- It does not require too many computational resources as it's highly interpretable
- There is no problem scaling the input features—It does not require tuning
- It is easy to implement and train a model using logistic regression
- It gives a measure of how relevant a predictor (coefficient size) is, and its direction of association (positive or negative)

PART B

(PART B : TO BE COMPLETED BY STUDENTS)

Roll. No. BE-A15	Name: Khan Mohammad TAQI
Class: BE-Comps	Batch: A1
Date of Experiment:	Date of Submission:
Grade:	

B.1 Software Code written by student:

```
# To implement Logistic Regression.
import numpy as np
import pandas as pd
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, accuracy_score, confusion_matrix

data = load_breast_cancer()
X = pd.DataFrame(data.data, columns=data.feature_names)
y = pd.Series(data.target)

print(X.head())
print(y.head())

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = LogisticRegression(max_iter=10000)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nAccuracy Score:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

B.2 Input and Output:

```
taqi@Taqi:~/experiments/ml$ python3 exp1.py
   mean radius  mean texture  ...  worst symmetry  worst fractal dimension
0        17.99         10.38  ...          0.4601                  0.11890
1        20.57         17.77  ...          0.2750                  0.08902
2        19.69         21.25  ...          0.3613                  0.08758
3        11.42         20.38  ...          0.6638                  0.17300
4        20.29         14.34  ...          0.2364                  0.07678

[5 rows x 30 columns]
0    0
1    0
2    0
3    0
4    0
dtype: int64
Confusion Matrix:
 [[39  4]
 [ 1 70]]

Accuracy Score: 0.956140350877193

Classification Report:
               precision    recall  f1-score   support

           0       0.97      0.91      0.94        43
           1       0.95      0.99      0.97        71

    accuracy                           0.96       114
   macro avg       0.96      0.95      0.95       114
weighted avg       0.96      0.96      0.96       114
```

B.3 Observations and learning:

During this experiment, I successfully implemented the Logistic Regression algorithm, a key supervised learning technique for binary classification. I observed that the core of this model is the Sigmoid function, which effectively transforms numerical outputs into a probability score between 0 and 1. This mechanism allows the model to predict the likelihood of a binary outcome, such as 'yes' or 'no', or 'true' or 'false'. The theoretical study highlighted the algorithm's efficiency and interpretability, especially with linearly separable data. I noted its wide-ranging applications in various fields, including healthcare and finance, for tasks like tumor classification and fraud detection. The implementation process was straightforward, confirming its advantage of not requiring extensive computational resources.

B.4 Conclusion:

In conclusion, this experiment successfully achieved its aim of implementing a functional Logistic Regression model. I have demonstrated a practical understanding of how to apply this algorithm to solve binary classification problems. The process reinforced the understanding that Logistic Regression is a powerful yet simple tool for predictive modeling, providing clear, probabilistic outcomes that are easy to interpret. The successful implementation confirms its value as a foundational machine learning algorithm, essential for any data scientist's toolkit due to its efficiency, simplicity, and broad applicability.