ROLL NO.: A-25	**NAME:SHRADDHA SAWANT**
CLASS:T.E.	**BATCH:A-2**
DATE OF PERFORMANCE:	**DATE OF SUBMISSION:27/3/2025**
GRADE:	

Experiment No.04

Aim: Implementation of Logistic Regression Algorithm.

Outcome:
After completing this practical, the **Logistic Regression algorithm** was successfully implemented, demonstrating its working principle for **binary classification problems**. Hands-on experience was gained in **data preprocessing, model training, evaluating performance using accuracy and confusion matrix, and visualizing decision boundaries**. This knowledge enhances the ability to apply logistic regression for real-world classification tasks.

Theory:
Logistic regression is a **supervised machine learning algorithm** used for **classification tasks** where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is a statistical algorithm which analyze the relationship between two data factors. The article explores the fundamentals of logistic regression, it's types and implementations.

Logistic regression is used for binary classification where we use sigmoid function, that takes input as independent variables and produces a probability value between 0 and 1.

For example, we have two classes Class 0 and Class 1 if the value of the logistic function for an input is greater than 0.5 (threshold value) then it belongs to Class 1 otherwise it belongs to Class 0. It's referred to as regression because it is the extension of linear regression but is mainly used for classification problems.

Key Points:

- Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value.

- It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).

Types of Logistic Regression

On the basis of the categories, Logistic Regression can be classified into three types:

1. **Binomial:** In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.

2. **Multinomial:** In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep".

3. **Ordinal:** In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

Understanding Sigmoid Function

So far, we've covered the basics of logistic regression, but now let's focus on the most important function that forms the core of logistic regression**.**

- The **sigmoid function** is a mathematical function used to map the predicted values to probabilities.
- It maps any real value into another value within a range of **0 and 1**. The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "**S**" form.
- The S-form curve is called the **Sigmoid function or the logistic function.**
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

How does Logistic Regression work?

The logistic regression model transforms the linear regression function continuous value output into categorical value output using a sigmoid function, which maps any real-valued set of independent variables input into a value between 0 and 1. This function is known as the logistic function.

Let the independent input features be:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ x_{21} & \cdots & x_{2m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

and the dependent variable is Y having only binary value i.e. 0 or 1.

$$Y = \begin{cases} 0 & \text{if } Class\ 1 \\ 1 & \text{if } Class\ 2 \end{cases}$$

then, apply the multi-linear function to the input variables X.

$$z = \left(\sum_{i=1}^{n} w_i x_i\right) + b$$

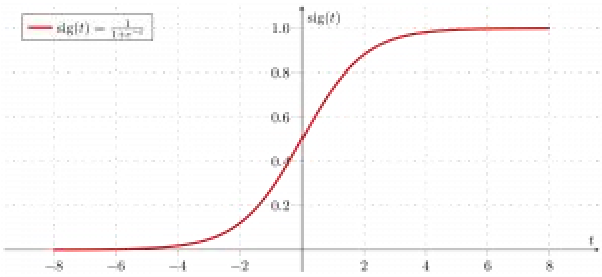
Here xi x_i is the ith observation of X, wi=[w1,w2,w3,⋯,wm] $w_i = [w_1, w_2, w_3, \cdots, w_m]$ is the weights or Coefficient, and b is the bias term also known as intercept. simply this can be represented as the dot product of weight and bias.

$$z = w \cdot X + b$$

whatever we discussed above is the linear regression.

Sigmoid Function

Now we use the sigmoid function where the input will be z and we find the probability between 0 and 1. i.e. predicted y.



As shown above, the figure sigmoid function converts the continuous variable data into the probability i.e. between 0 and 1.

- σ(z) $\sigma(z)$ tends towards 1 as z→∞ $z \to \infty$
- σ(z) $\sigma(z)$ tends towards 0 as z→−∞ $z \to -\infty$
- σ(z) $\sigma(z)$ is always bounded between 0 and 1

where the probability of being a class can be measured as:

$$P(y=1) = \sigma(z) \quad P(y=0) = 1 - \sigma(z)$$

Terminologies involved in Logistic Regression

Here are some common terms involved in logistic regression:

Independent variables: The input characteristics or predictor factors applied to the dependent variable's predictions

Dependent variable: The target variable in a logistic regression model, which we are trying to predict

Logistic function: The formula used to represent how the independent and dependent variables relate to one another. The logistic function transforms the input variables into a probability value between 0 and 1, which represents the likelihood of the dependent variable being 1 or 0.

Odds: It is the ratio of something occurring to something not occurring. it is different from probability as the probability is the ratio of something occurring to everything that could possibly occur.

Log-odds: The log-odds, also known as the logit function, is the natural logarithm of the odds. In logistic regression, the log odds of the dependent variable are modeled as a linear combination of the independent variables and the intercept.

Coefficient: The logistic regression model's estimated parameters, show how the independent and dependent variables relate to one another.

Intercept: A constant term in the logistic regression model, which represents the log odds when all independent variables are equal to zero.

CODE:

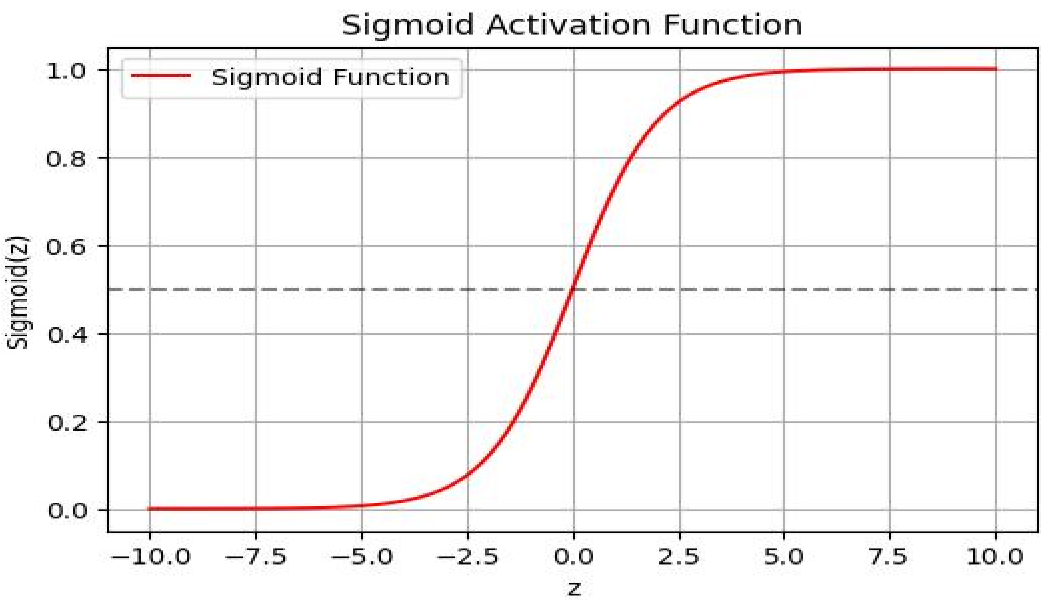
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.datasets import load_breast_cancer
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.decomposition import PCA

# Load dataset
X, y = load_breast_cancer(return_X_y=True)
```

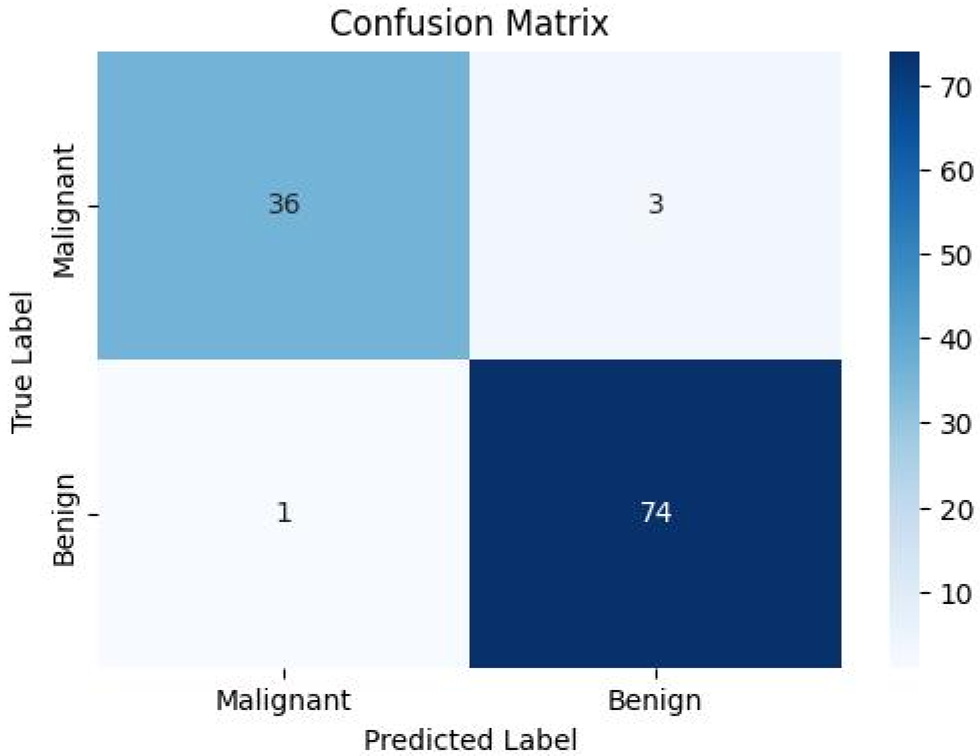
```
# Splitting data into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=23)
```

```
# Initialize and train the Logistic Regression model
clf = LogisticRegression(max_iter=10000, random_state=0)
clf.fit(X_train, y_train)
```

```
# Predictions
y_pred = clf.predict(X_test)
acc = accuracy_score(y_test, y_pred) * 100
```

```python
# Print accuracy
print(f"Logistic Regression model accuracy (in %): {acc:.2f}%\n")

# ------------------- Visualization 1: Sigmoid Function -------------------
def sigmoid(z):
    return 1 / (1 + np.exp(-z))

z = np.linspace(-10, 10, 100)
sigmoid_values = sigmoid(z)

plt.figure(figsize=(6, 4))
plt.plot(z, sigmoid_values, label="Sigmoid Function", color="red")
plt.axhline(0.5, linestyle="dashed", color="black", alpha=0.5)
plt.xlabel("z")
plt.ylabel("Sigmoid(z)")
plt.title("Sigmoid Activation Function")
plt.legend()
plt.grid()
plt.show()

# ------------------- Visualization 2: Confusion Matrix -------------------
conf_matrix = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(6, 4))
sns.heatmap(conf_matrix, annot=True, fmt="d", cmap="Blues", xticklabels=["Malignant",
"Benign"], yticklabels=["Malignant", "Benign"])
plt.xlabel("Predicted Label")
plt.ylabel("True Label")
plt.title("Confusion Matrix")
plt.show()
```





Conclusion:

Successfully implemented the **Logistic Regression Algorithm**, analyzed model accuracy, visualized the sigmoid function, confusion matrix, and decision boundary, gaining practical insights into classification tasks.