Experiment 3

Implement logistic regression using sklearn python library.

Aim: Implement logistic regression using sklearn python library.

Theory:

What is Logistic Regression?

- **Logistic regression** is a supervised machine learning algorithm used for classification problems, where the dependent variable is categorical (often binary).

- Instead of predicting exact values (as in linear regression), logistic regression predicts the probability of a given sample belonging to a particular class.

Mathematical Background

- The logistic regression model uses the **sigmoid function** to map input values (linear combination of features) to a probability between 0 and 1.

- The **sigmoid function** is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

where $z = w0 + w1x1 + w2x2 + ... + wnxn$

- The model predicts class 1 if the probability is greater than a threshold (usually 0.5); otherwise, it predicts class 0.

Working with sklearn

- sklearn.linear_model.LogisticRegression is the main class used.

- The steps involved include:

    o Data preprocessing (splitting, scaling if required)

    o Creating a model object

    o Training (fitting) the model with training data

    o Predicting outcomes for test data

o   Evaluating model performance (accuracy, confusion matrix, etc.)

Conclusion:

- Logistic regression provides an efficient and interpretable approach for binary and multiclass classification tasks.

- Using sklearn, the implementation is simple due to high-level APIs for data preparation, training, and evaluation.

- The model's accuracy and effectiveness depend on the quality of data and feature selection.

Viva Questions:

- What is the difference between logistic regression and linear regression?

- Explain the sigmoid function and its role in logistic regression.

- Why is logistic regression used for classification and not regression tasks?

- How do you interpret the coefficients in logistic regression?

- What does the fit() method do in sklearn's logistic regression?

- How can you evaluate the performance of a logistic regression model?

- Explain the concept of regularization in logistic regression.

- What happens if the data is not linearly separable?

Program:

python

```python
from sklearn.datasets import load_iris

from sklearn.model_selection import train_test_split

from sklearn.linear_model import LogisticRegression

from sklearn.metrics import accuracy_score
```

```python
# Load dataset
iris = load_iris()
X = iris.data
y = (iris.target == 0).astype(int)  # Binary classification: Setosa vs Others

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Create logistic regression model
model = LogisticRegression()

# Train the model
model.fit(X_train, y_train)

# Predict
y_pred = model.predict(X_test)

# Evaluate accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy:.2f}")
```