Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

AY: 2023-24

Class:	TE	**Semester:**	VI
Course Code:	CSL604	**Course Name:**	Machine Learning Lab

Name of Student:	Ojasi Prabhu
Roll No.:	43
Experiment No.:	3
Title of the Experiment:	Implementation of Logistic Regression
Date of Performance:	
Date of Submission:	

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Performance	5	
Understanding	5	
Journal work and timely submission	10	
Total	20	

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Performance	4-5	2-3	1
Understanding	4-5	2-3	1
Journal work and timely submission	8-10	5-8	1-4

Checked by

Name of Faculty : Mr Raunak Joshi

Signature :

Date :

Aim: Implementation of Logistic Regression Algorithm.

Objective: Able to perform various feature engineering tasks, apply logistic regression on the given dataset and maximize the accuracy.

Theory:

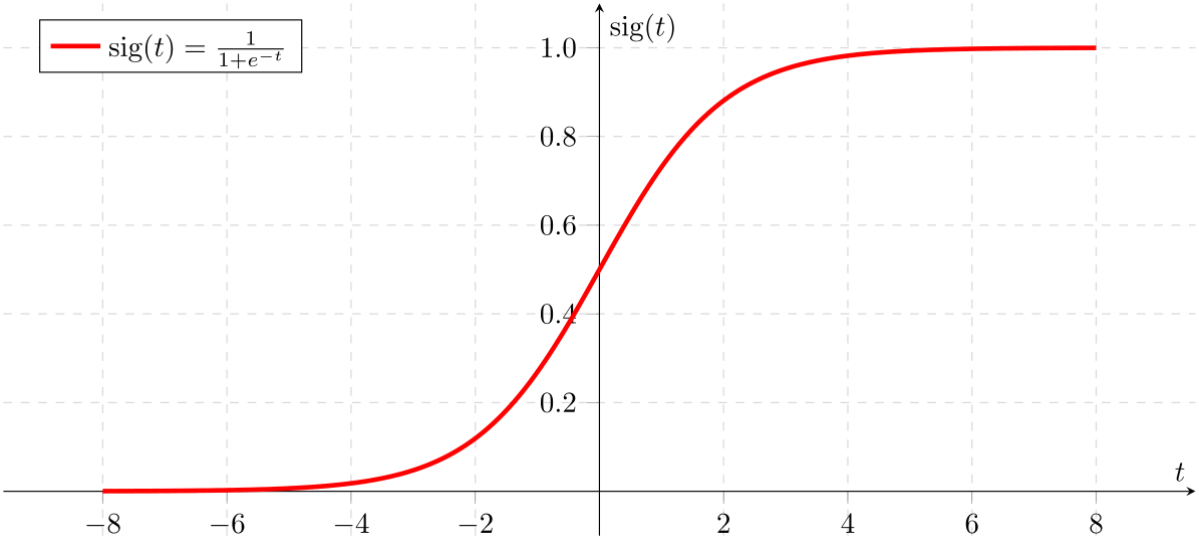
Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable(target) is categorical and is binary in nature. In order to perform binary classification, the logistic regression techniques makes use of Sigmoid function.

For example,

To predict whether an email is spam (1) or (0)

Whether the tumor is malignant (1) or not (0)

Consider a scenario where we need to classify whether an email is spam or not. If we use linear regression for this problem, there is a need for setting up a threshold based on which classification can be done. Say if the actual class is malignant, predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malignant which can lead to serious consequence in real time.



From this example, it can be inferred that linear regression is not suitable for classification problem. Linear regression is unbounded, and this brings logistic regression into picture. Their value strictly ranges from 0 to 1.

Implementation: =

```python
import numpy as np
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split

class LogisticRegression:
    def __init__ (self):
        self.params = np.zeros(int(np.random.random()), float)[:,np.newaxis]

    def fit (self, X, y):
        bias = np.ones (len (X))
        X_bias = np.c_[bias, X]
        inner_part = np.transpose (X_bias) @ X_bias
        inverse_part = np.linalg.inv (inner_part)
        outer_part = inverse_part @ np.transpose (X_bias)
        least_square_estimate = outer_part @ y
        self.params = least_square_estimate
        return self.params

    def predict (self, X):
        y_hat = list ()
        bias_testing = np.ones (len (X))
        X_test = np.c_[bias_testing, X]
        z = X_test @ self.params
        sigmoid = 1 / (1 + np.exp (-z))
        for _ in range (len (sigmoid)):
            if sigmoid[_] >= 0.5:
                y_hat.append (1)
            else:
                y_hat.append (0)
        return sigmoid, y_hat

if __name__ == '__main__':
    # X = np.array([.50, 1.50, 2.00, 4.25, 3.25, 5.50], ndmin=2).reshape((6,1))
    # y = np.array([0, 0, 0, 1, 1, 1])

    dataset = load_breast_cancer ()
    X = dataset.data
    y = dataset.target
    print (X.shape)

    X_train, X_test, y_train, y_test = train_test_split (X, y, test_size=0.1)

    model = LogisticRegression ()
    parameters = model.fit (X_train, y_train)
    # print (parameters)

    sig, y_pred = model.predict (X_test)
    print (f'The predicted outcome is {y_pred} and calculated sigmoid value is {sig}')

    print (f'First value of y_test : {y_test[14]} and first value of y_pred : {y_pred[14]}')
    print (f'The sigmoid probability for the tested value : {sig[14]}')
```

Output

```
(base) PS C:\Users\parth> python -u "c:\Users\parth\OneDrive\Desktop\ML LAB\logisticRegression.py"
(569, 30)
The predicted outcome is [0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1] and calculated sigmoid value is [0.49152494 0.71731219 0.74724397 0.67173795 0.62154607 0.706207
 0.72226664 0.71420152 0.73688669 0.65708673 0.65972199 0.62446522
 0.70558785 0.73901591 0.74335162 0.43961101 0.75874245 0.76282396
 0.70230753 0.69246187 0.65486738 0.63400112 0.62267494 0.50910988
 0.65025163 0.75450118 0.47047648 0.72010962 0.55366755 0.70787197
 0.5442737  0.70843621 0.70656497 0.66146519 0.71365817 0.64077053
 0.71588586 0.68163397 0.61453195 0.72674515 0.67348449 0.67172547
 0.64757279 0.74743731 0.76813744 0.7332035  0.73938532 0.68952185
 0.69701309 0.64759722 0.76973167 0.71826239 0.71506127 0.54787801
 0.60887205 0.49186898 0.76005579]
First value of y_test : 1 and first value of y_pred : 1
The sigmoid probability for the tested value : 0.7433516233164259
```

Conclusion:

Comment on the accuracy obtained.

While contemporary logistic regression implementations utilizing the least squares method may provide insights, their accuracy may be limited due to methodological limits. However, by using more appropriate optimization techniques such as maximum likelihood estimation or gradient descent, as well as thorough evaluation using metrics other than accuracy, such as precision, recall, and ROC AUC, we can improve the model's performance in classification tasks such as breast cancer detection.