**U18ISI6204 – Machine Learning Techniques**

**LAB- EXPERIMENT 4**

|  |  |
| --- | --- |
| NAME | Navina G A |
| ROLL NO | 20BIS028 |

# INTRODUCTION

In this experiment, we have to perform Logistic regression on the covid dataset.

Logistic regression is a [statistical model](https://en.wikipedia.org/wiki/Statistical_model) that in its basic form uses a [logistic function](https://en.wikipedia.org/wiki/Logistic_function) to model a [binar](https://en.wikipedia.org/wiki/Binary_variable)y [dependent variable](https://en.wikipedia.org/wiki/Dependent_variable), although many more complex [extensions](https://en.wikipedia.org/wiki/Logistic_regression#Extensions) exist. In [regression](https://en.wikipedia.org/wiki/Regression_analysis) [analysis](https://en.wikipedia.org/wiki/Regression_analysis), logistic regression (or logit regression) is [estimating](https://en.wikipedia.org/wiki/Estimation_theory) the parameters of a logistic model (a form of [binary regression](https://en.wikipedia.org/wiki/Binary_regression)).

# Linear Regression Equation:



Where, y is dependent variable and x1, x2 ... and Xn are explanatory variables.

# Sigmoid Function:

Apply Sigmoid function on linear regression:



**OBJECTIVE OF THE EXERCISE/EXPERIMENT**

To perform Logistic regression on the given dataset, using scikit library

ACQUISITION PROCEDURE:

STEP-1: Start the program.

STEP-2: import all the necessary libraries

1. Numpy – array manipulation
2. Pandas – dataframe manipulation
3. Matplotlib and seaborn – for data visualization
4. Sklearn.model\_selection – train test data split
5. Sklearn.metrics – f1 score.
6. Sklearn,linear\_model – for logistic regression

STEP-3: Loading the dataset using read\_csv method in pandas module.

STEP-4: Analyze the dataset using info method, which gives its data types and number of non- null values in each columns.

STEP-5: Perform basic statistic operation using describe() method.

STEP-6: Use heatmaps, correlation matrix, regression plots and pairplots in seaborn to find the relationship between features.

STEP-7: Implement Logistics regression(logreg) with all variable and calculate the f1 score.

STEP-8: Stop the program.

# PROGRAM:

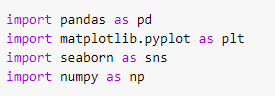
Importing libraries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

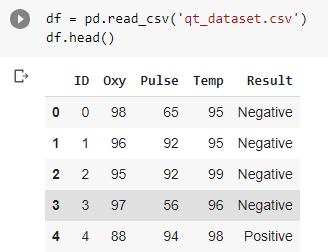
import numpy as np



# Loading dataset

# df = pd.read\_csv(‘qt\_dataset.csv’)

# df.head()



**Basic statistics operations**

Printf(df.info())

df.describe()

A screenshot of a computer

Description automatically generated with medium confidence

A picture containing text, screenshot, font, number

Description automatically generated

# Correlation between columns

# Sns.heatmap(df.corr(), annot = True)

A screenshot of a computer

Description automatically generated with medium confidence

Pairplots.

Sns.pairplot(df,hue = ‘Result’)

A screenshot of a graph

Description automatically generated with low confidence

# Train test split.

# from sklearn.model\_selection import train\_test\_split

# X\_train, X\_test, y\_train,y\_test = train\_test\_split(X,y,test\_size=0.2, random\_state =4)



Logistic regression :

From sklearn.linear\_model import LogisticRegression

logreg = LogisticRegression()

logreg.fit(X\_train, y\_train)

y\_pred = logreg.predict(X\_test)

from sklearn.metrics import f1\_score

print(‘f1 Score :’ ,f1\_score(y\_test,y\_pred,average =’micro’))

