**Name:- Medhansh K**

**PRN:-122A8035**

**Batch:-B2**

**Experiment No. 3**

**Aim:-** Implementation of logistic regression for a given dataset.

**Theory:-**

**Logistic Regression**

A statistical model for binary classification is called [logistic regression](https://www.geeksforgeeks.org/ml-linear-regression-vs-logistic-regression/). Using the sigmoid function, it forecasts the likelihood that an instance will belong to a particular class, guaranteeing results between 0 and 1. To minimize the log loss, the model computes a linear combination of input characteristics, transforms it using the sigmoid, and then optimizes its coefficients using methods like gradient descent. These coefficients establish the decision boundary that divides the classes. Because of its ease of use, interpretability, and versatility across multiple domains, Logistic Regression is widely used in machine learning for problems that involve binary outcomes. Overfitting can be avoided by implementing regularization.

**How the Logistic Regression Algorithm Works?**

[Logistic Regression](https://www.geeksforgeeks.org/understanding-logistic-regression/) models the likelihood that an instance will belong to a particular class. It uses a linear equation to combine the input information and the sigmoid function to restrict predictions between 0 and 1. Gradient descent and other techniques are used to optimize the model’s coefficients to minimize the [log loss](https://www.geeksforgeeks.org/ml-log-loss-and-mean-squared-error/). These coefficients produce the resulting decision boundary, which divides instances into two classes. When it comes to binary classification, logistic regression is the best choice because it is easy to understand, straightforward, and useful in a variety of settings. Generalization can be improved by using regularization.

Key Concepts of Logistic Regression

Important key concepts in logistic regression include:

**Sigmoid Function:** The main function that ensures outputs are between 0 and 1 by converting a linear combination of input data into probabilities.  
The [sigmoid function](https://www.geeksforgeeks.org/derivative-of-the-sigmoid-function/) is denoted as σ(z)*σ*(*z*), and is defined as:  
σ(z)=11+ez*σ*(*z*)=1+*ez*1​  
Where, z is linear combination of input features and coefficients.

**Hypothesis Function:** uses the sigmoid function and weights (coefficients) to combine input features to estimate the likelihood of falling into a particular class.  
In logistic regression, the [hypothesis function](https://www.geeksforgeeks.org/ml-understanding-hypothesis/) is provided by:  
hθ(x)=σ(θTx)*hθ*​(*x*)=*σ*(*θTx*)  
Where, hθ(x)*hθ*​(*x*) is the predicted probability that y = 1, θ*θ* is the vector of coefficients, and x is the vector of input features.

**Log Loss:** The optimization [cost function HYPERLINK "https://www.geeksforgeeks.org/what-is-cost-function/" HYPERLINK "https://www.geeksforgeeks.org/what-is-cost-function/" HYPERLINK "https://www.geeksforgeeks.org/what-is-cost-function/"](https://www.geeksforgeeks.org/what-is-cost-function/) is a measure of the discrepancy between actual class labels and projected probability.  
The definition of the log loss for a single instance is:  
J(θ)=−(ylog⁡hθ(x)+(1–y)log⁡(1−hθ(x)))*J*(*θ*)=−(*y*log*hθ*​(*x*)+(1–*y*)log(1−*hθ*​(*x*)))

**Decision Boundary:** The surface or line used to divide instances into several classes according to the determined probability.

**Probability Threshold:** a number (usually 0.5) that is used to calculate the class assignment using the probabilities that are anticipated.

**Odds Ratio:** The likelihood that an event will occur as opposed to not, which sheds light on how characteristics and the target variable are related.

**Code:**







**Conclusion:**-

As we can clearly see from the plot, we get a straight line for linear models. We can use the model to test on similar datasets with more number of independent variables.