



KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY (KIIT)

Deemed to be University U/S 3 of UGC Act, 1956

Name : Abhishek Mallick

Roll : 21051706

ML_CS-3

Machine Learning

Assignment 2 – Logistic Regression

[Github](#)

AIM: The aim of this assignment is to classify the given dataset using logistic regression.

- Implement logistic regression for dataset classification using batch gradient descent.
- Analyze convergence by monitoring cost function and learning parameter values.
- Visualize decision boundaries and dataset, comparing learning rates and evaluating model performance.

Dataset used:

Independent/Predictor Variable -

<https://drive.google.com/file/d/1BtrY1Ke0JjM8bGt4rcLq87sfqqiiTw7/view?usp=sharing>

Dependent/Response Variable -

<https://drive.google.com/file/d/1oGSum358X0V31pAOQH0hOokD75fa4JjI/view?usp=sharing>

1. Use logistic regression to find decision boundary For the given database. Set your learning rate to 0.1. What is the cost function value and learning parameter value after convergence?.

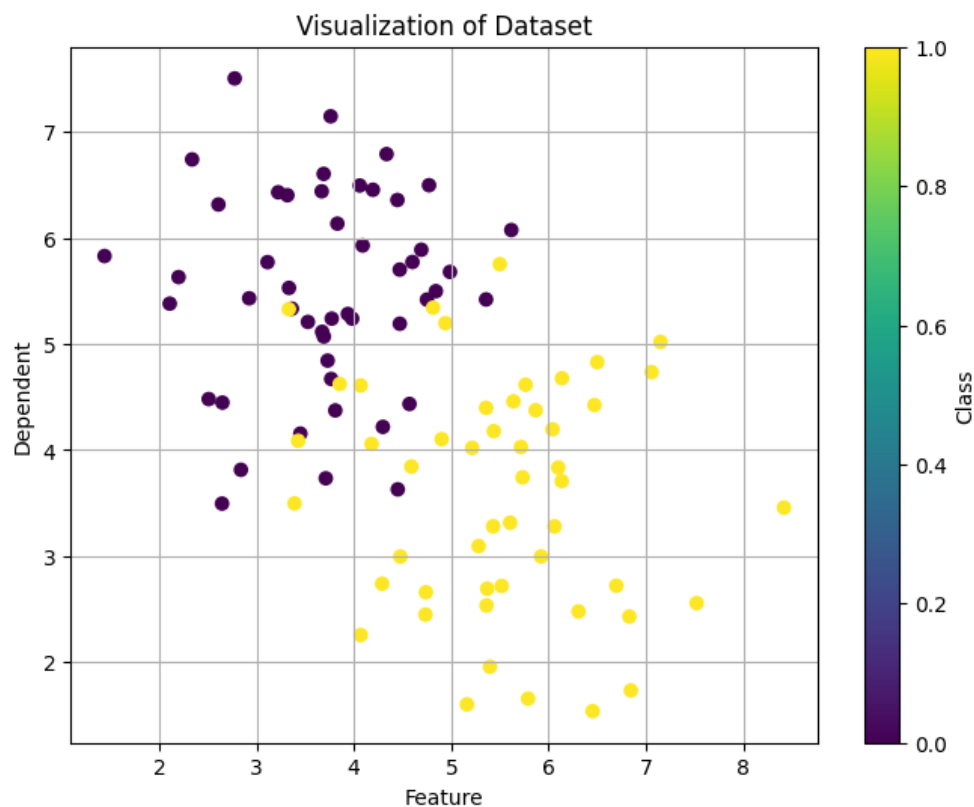
Merged dataframe representation:

```
5 print(merged_df)
```

	x	Dependent
0	3.6898 6.6051	0
1	2.7763 7.5057	0
2	3.1137 5.7724	0
3	2.9251 5.4315	0
4	3.6699 6.4406	0
..
94	5.7629 4.6161	1
95	4.1817 4.0577	1
96	6.3065 2.4776	1
97	5.6043 3.3146	1
98	7.0567 4.7346	1

[99 rows x 3 columns]

Data Vizualization:



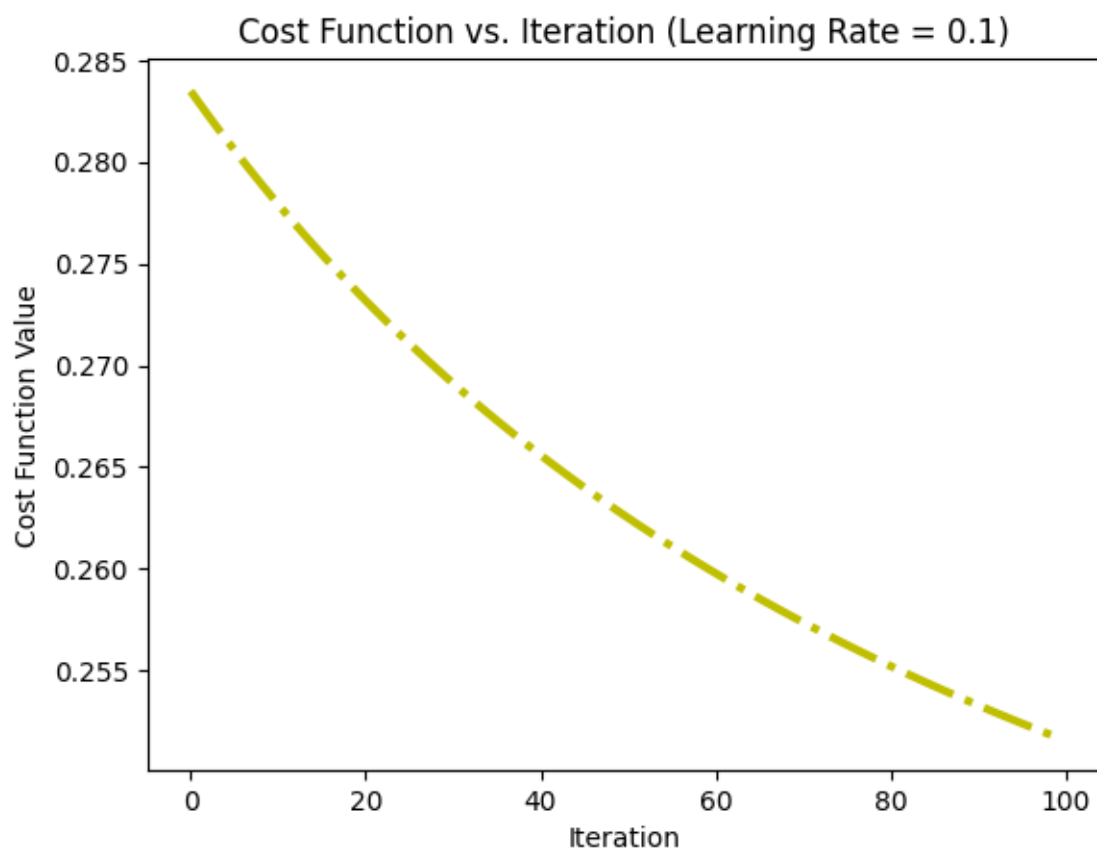
Using Cost Function:

```
def cost_function(X, y, theta):  
    m = len(y)  
    h = sigmoid(X.dot(theta))  
    cost = -1 / m * (y.T @ np.log(h) + (1 - y).T @ np.log(1 - h))  
    return cost
```

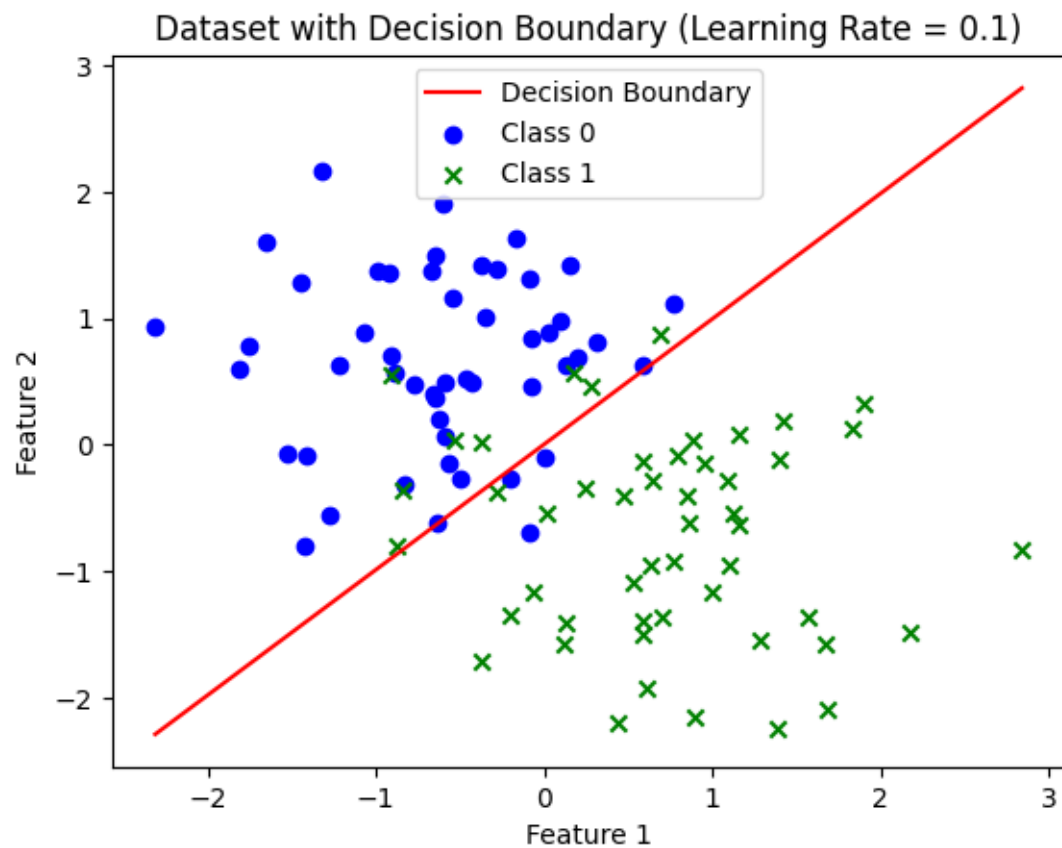
2. Plot cost function v/s iteration graph for the model trained in question 1. Plot the line as shown here .Do not use scatter plots for this.

For,
learning_rate = 0.1
iterations = 100

Cost Function vs Iteration (learning rate – 0.1)

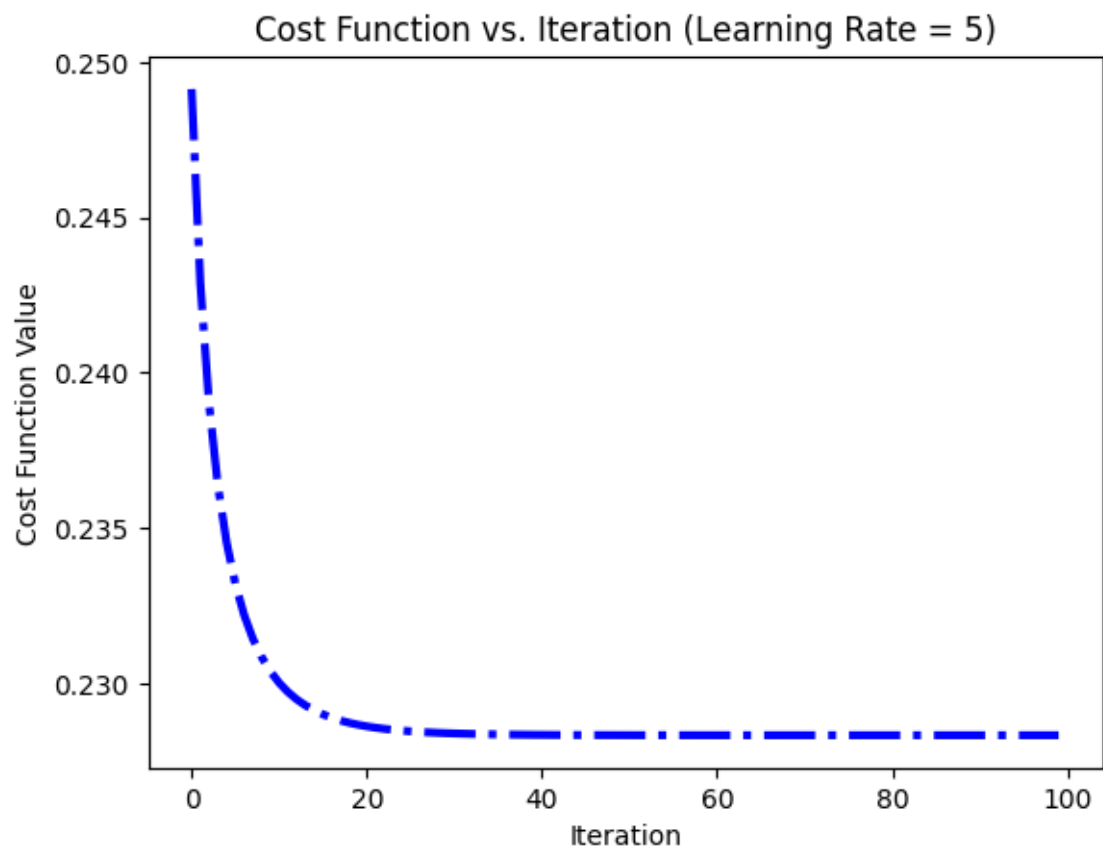


3. Plot the given dataset on a graph, use different colours for different classes and also show the decision boundary you obtained in question 1. Do not use scatter plot.

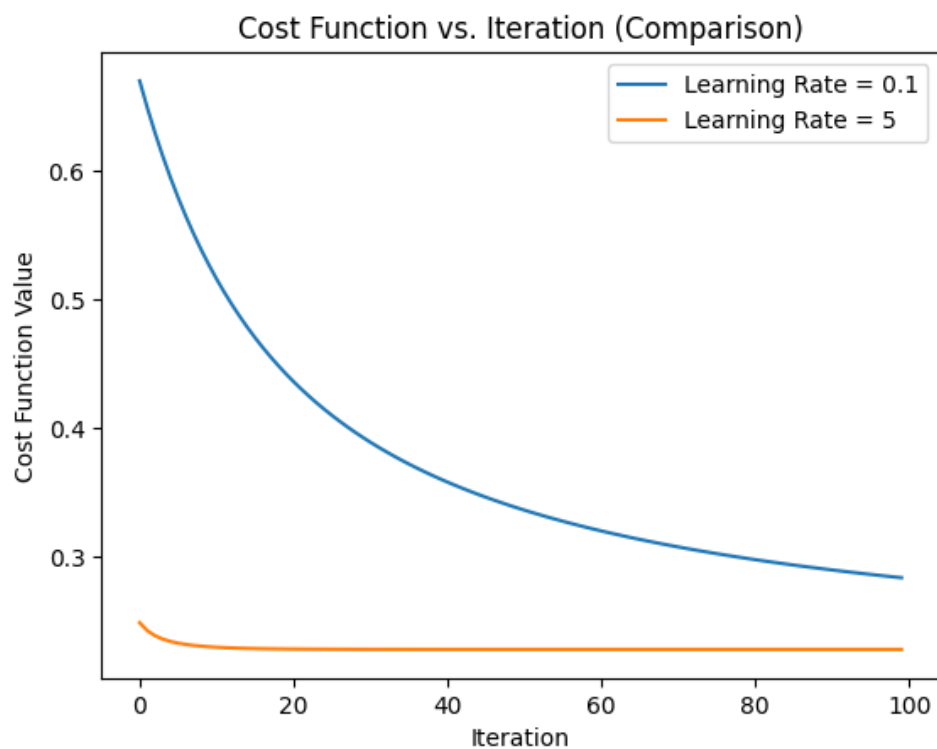


4. Train your model with a learning rate of 0.1 and 5. Plot the cost-function v/s iteration curve for both learning rates on the same graph. For this task, only train your model for 100 iterations.

Cost Function vs Iteration (learning rate – 5)



Comparison between the two learning rate plots wrt to learning rates



5. Find the confusion matrix for your training dataset. Using the confusion matrix to calculate the accuracy, precision, recall, F1-score.

Metrics for Learning Rate 0.1:

Accuracy: 0.89

Precision: 0.9333333333333333

Recall: 0.84

F1 Score: 0.8842105263157894

Metrics for Learning Rate 5:

Accuracy: 0.88

Precision: 0.8958333333333334

Recall: 0.86

F1 Score: 0.8775510204081632

Submitted By:

Abhishek Mallick

21051706

ML_CS-3

<https://github.com/Abhishek-Mallick/Machine-Learning-KIIT>