Assignment-2

Logistic Regression

Eshita Yadav

Information Technology (IT-6)

School of Computer Engineering

Bachelor of Technology

Roll: 2206340

Email: 2206340@kiit.ac.in

Answer the following questions based on your observations.

1. Use logistic regression to find the 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?

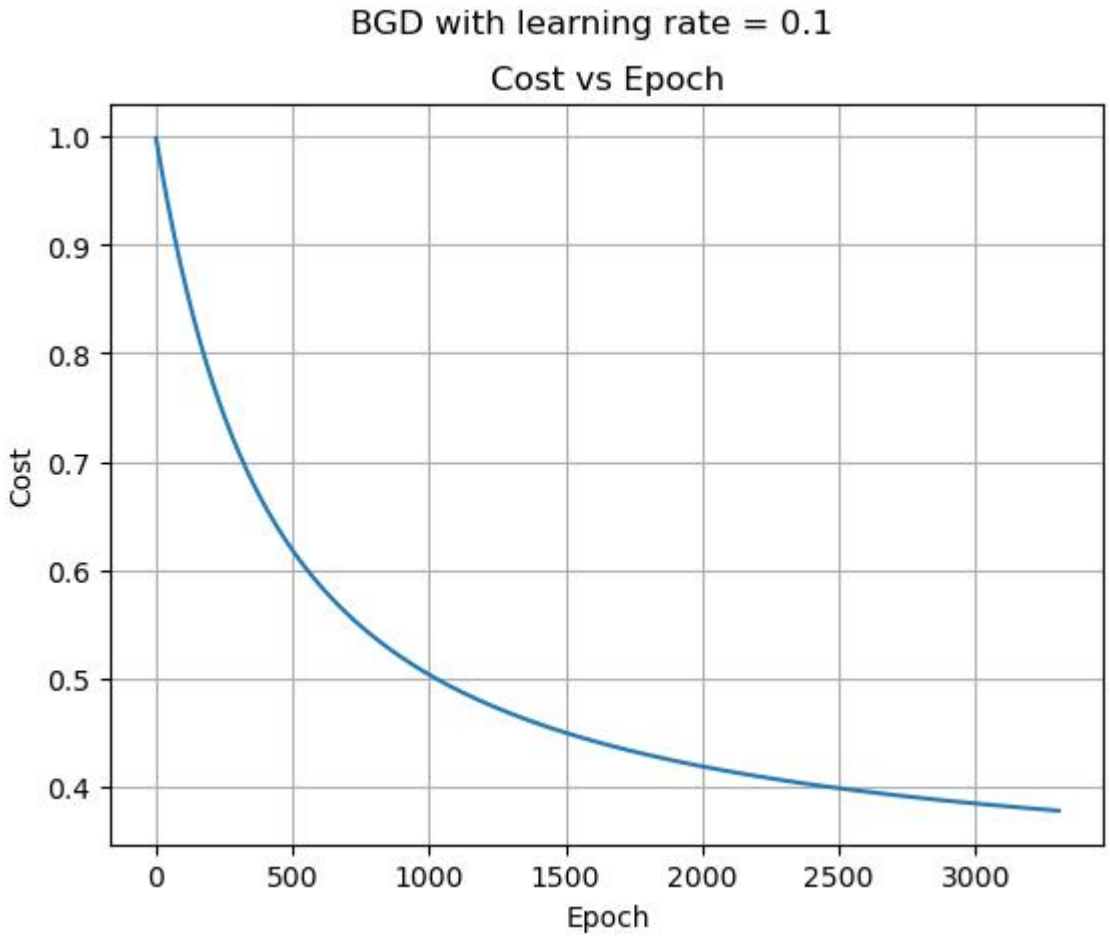
   Ans: A logistic regression model was built from scratch using batch gradient descent with a learning rate of 0.1. Convergence was achieved before the maximum limit of 5000 epochs, at 3304 epochs.

   - **Learning Rate (α)**: 0.1
   - **Epochs**: 3304
   - **Final Cost Function Value**: 0.37787475909925594
   - **Parameters**:
     - $w = \begin{bmatrix} 6.91479057 \\ -6.96696498 \end{bmatrix}$
     - b = 0.4863303181144615

   The cost function reduced steadily without fluctuations, confirming stable convergence.

2. Plot cost function v/s iteration graph for the model trained in question 1. Plot the line as shown here - https://pythonguides.com/matplotlib-plot-a-line/#Matplotlib_plot_a_line_chart. Do not use scatter plots for this.

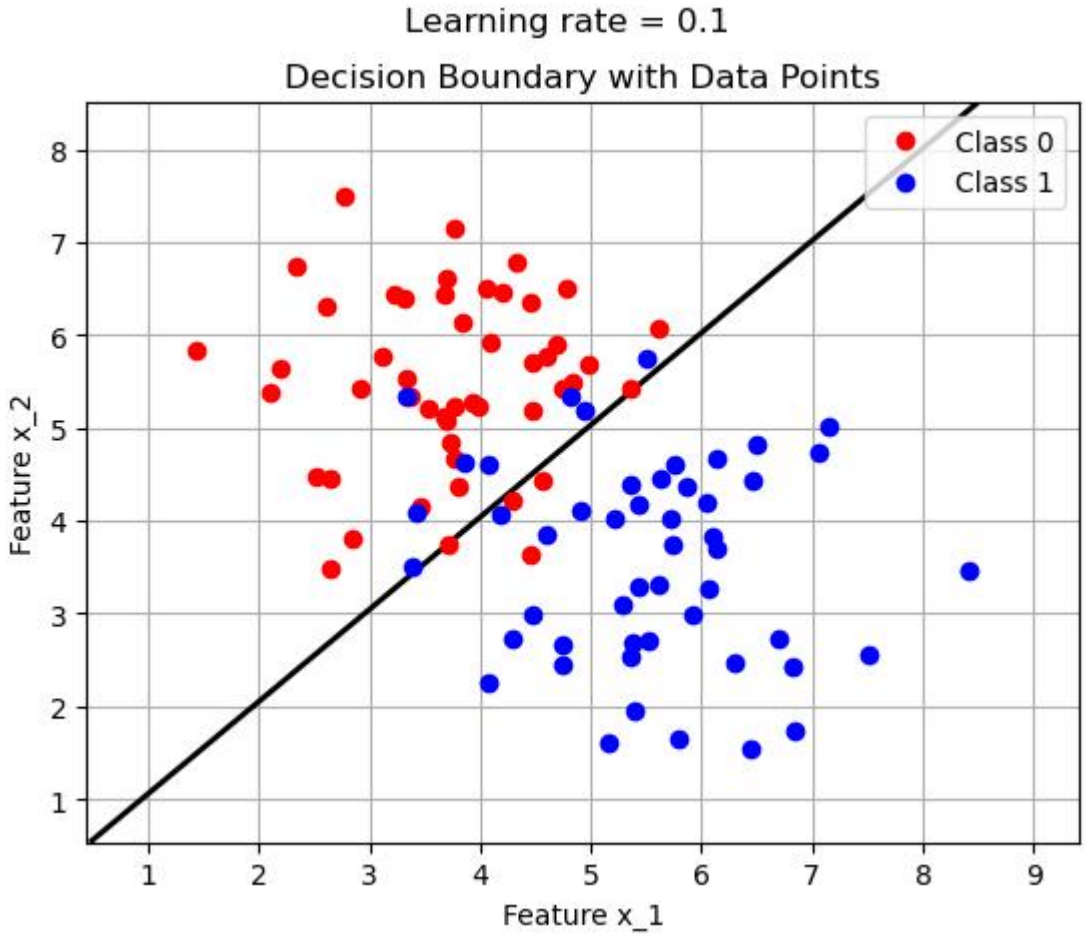
Ans:



The plotted line graph illustrates a smooth and steady decline in cost over 3300 epochs. Starting from a cost of 1, the curve flattens as convergence is approached, showing effective optimization by batch gradient descent.

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.

Ans:



Th

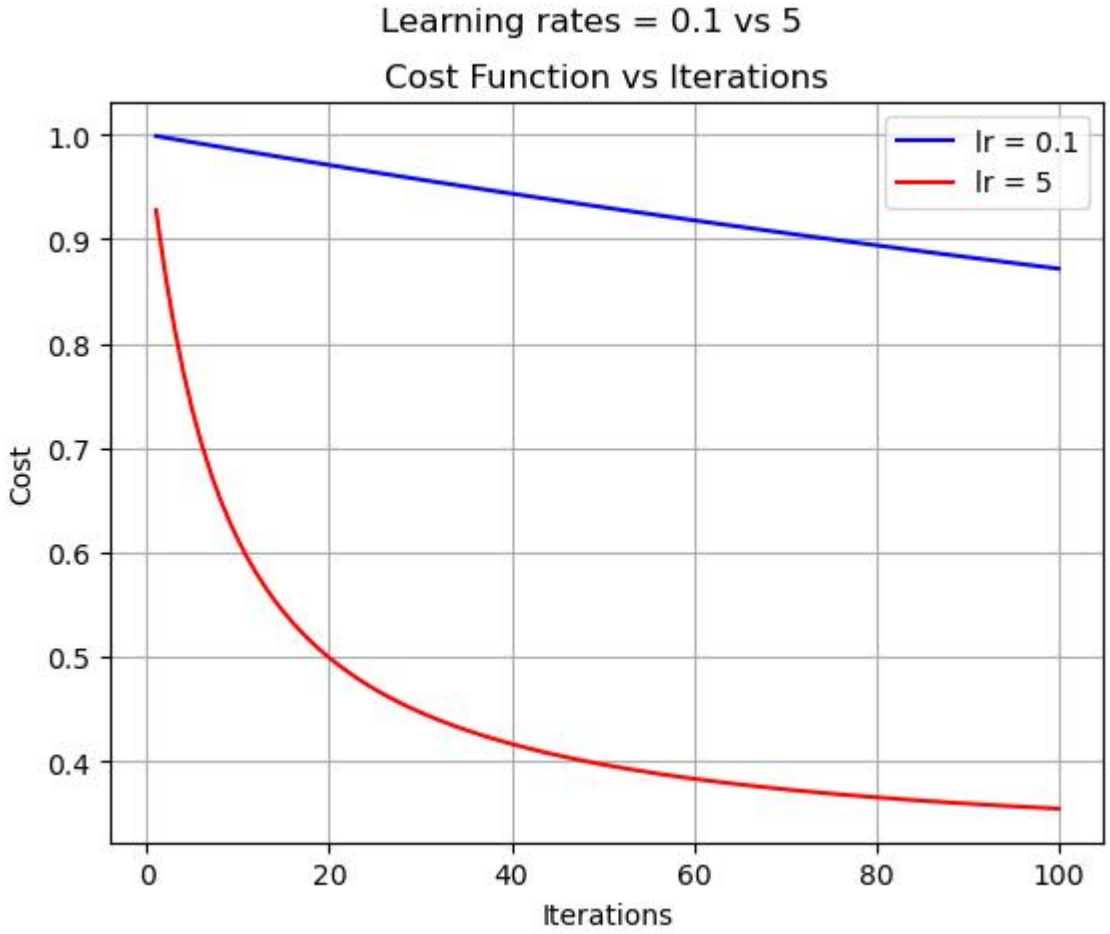
The plot shows two distinct classes:

- **Red** for Class 0
- **Blue** for Class 1

The graph displays two classes in distinct colors: red for Class 0 and blue for Class 1. A black line represents the learned decision boundary, which effectively separates the classes. Although a few points overlap, the overall separation reflects solid model performance.

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.

Ans:



Both models are trained for 100 iterations.

The graph shows two lines:

- **Learning Rate α = 0.1 (Blue Line)**: Cost declines gradually and nearly linearly, indicating slow but stable learning.
- **Learning Rate α = 5 (Red Line)**: Cost drops steeply at first and flattens quickly, achieving faster convergence without instability.

Despite common assumptions, a learning rate of 5 yielded better convergence here, whereas 0.1 was too slow within the limited iterations.

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

Ans:

Confusion matrix: $\begin{bmatrix} \mathbf{45} & \mathbf{30} \\ \mathbf{2} & \mathbf{50} \end{bmatrix}$

Accuracy: 95%

Precision: 94.34%

Recall: 96.15%

F1-Score: 95.23%

These metrics show high overall performance with balanced handling of both false positives and false negatives, indicating the model's reliability.

Dataset and Repository Links:

- Dataset:
    - Independent/Predictor Variable: https://drive.google.com/file/d/1eEBjqnIMlViYOwoOBqZ11tMbn9cZLLEo/view?usp=drive_link
    - Dependent/Response Variable: https://drive.google.com/file/d/1hnowuwHw_qgRl1a0LLa6y0eItyMU_lZ0/view?usp=drive_link
- GitHub Repository: https://github.com/Eshitacodes/Scratched-LogiReg