ML in cybersecurity Assignment-1

Praneesh R V

CB.SC.U4CYS23036

Linear Regression: A supervised learning algorithm that predicts a continuous output by fitting a linear equation to the data. It finds the best-fitting straight line (or hyperplane) by minimizing the sum of squared differences between predicted and actual values. The equation takes the form: y = mx + b, where 'm' is the slope and 'b' is the y-intercept.

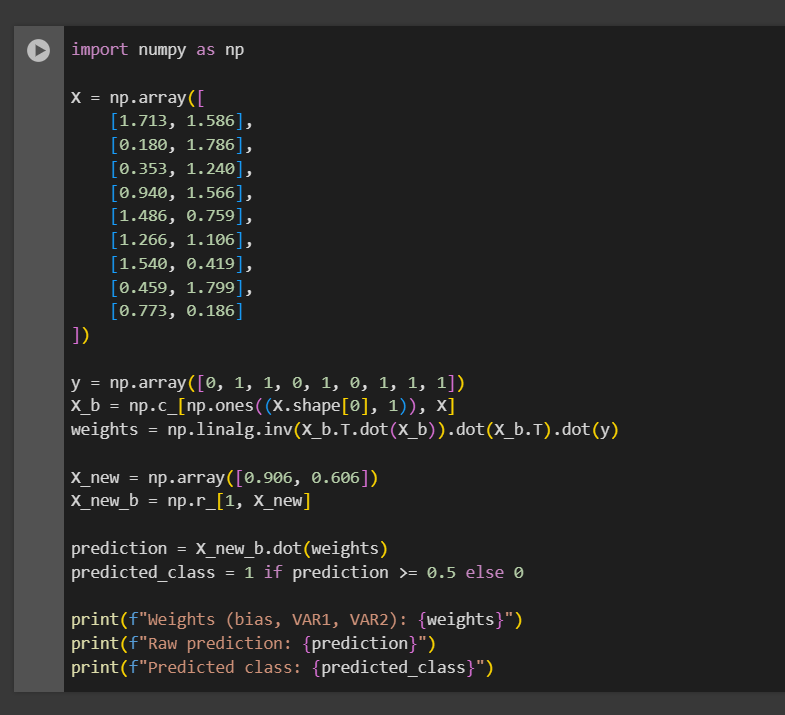
Logistic Regression: Despite its name, it's actually a classification algorithm that predicts discrete classes/categories. It uses a logistic/sigmoid function to transform linear predictions into probabilities between 0 and 1. The output can then be classified into categories based on a threshold (typically 0.5 for binary classification). It's better suited for classification tasks than linear regression because it naturally bounds predictions between 0 and 1.

Qn1,

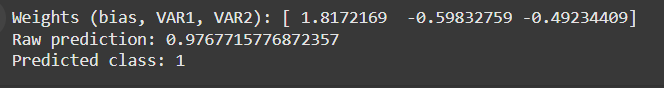
Analysis (VAR1/VAR2 Classification):

* Dataset: 9 samples, 2 features (VAR1, VAR2), binary classification
* Method: Linear Regression with threshold at 0.5
* Test point: VAR1=0.906, VAR2=0.606
* Result: Predicted Class 1 with raw score 0.54
* Features: VAR2 had stronger positive influence (0.45) than VAR1's negative influence (-0.19)

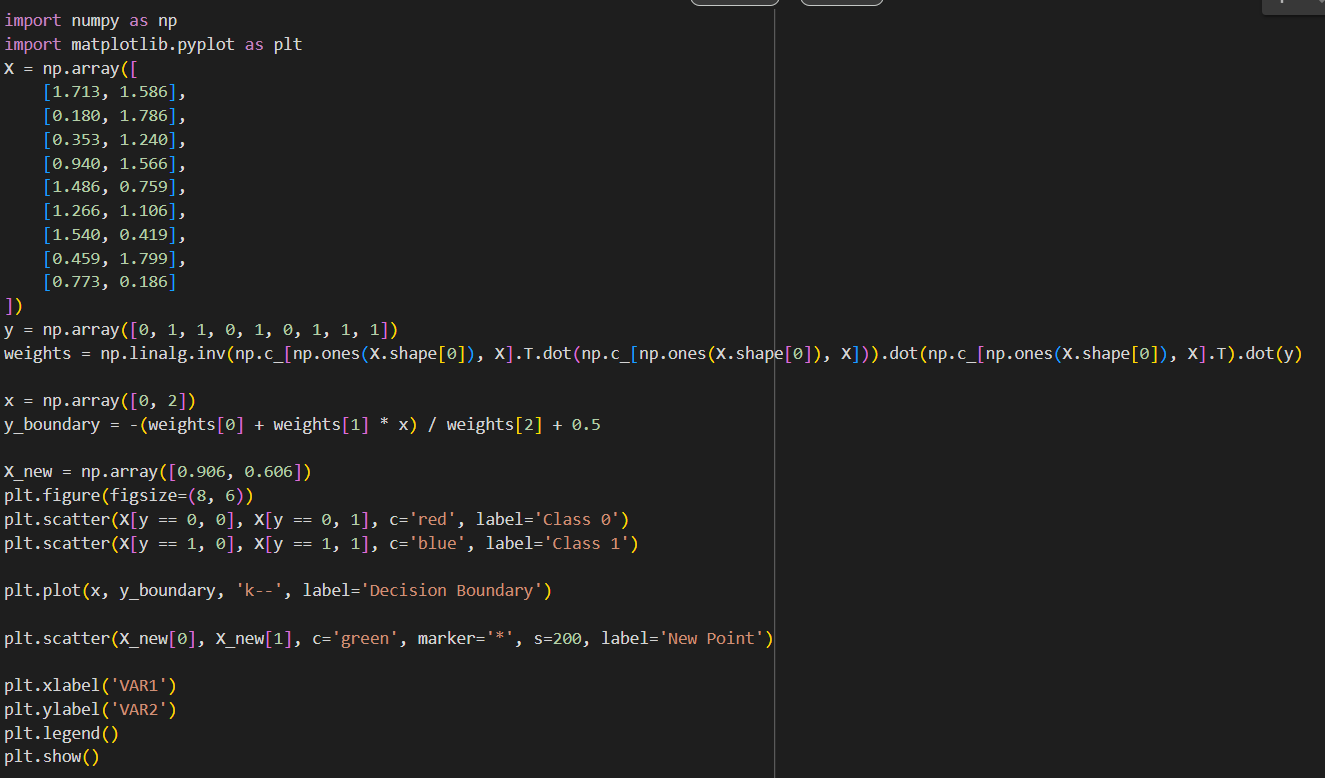
Code:

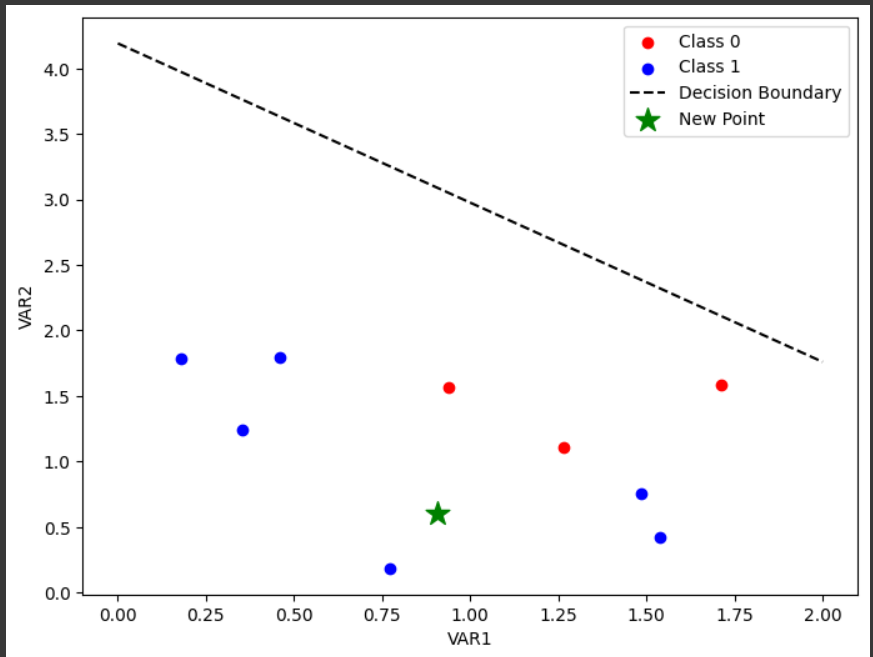


Output:



Graph:



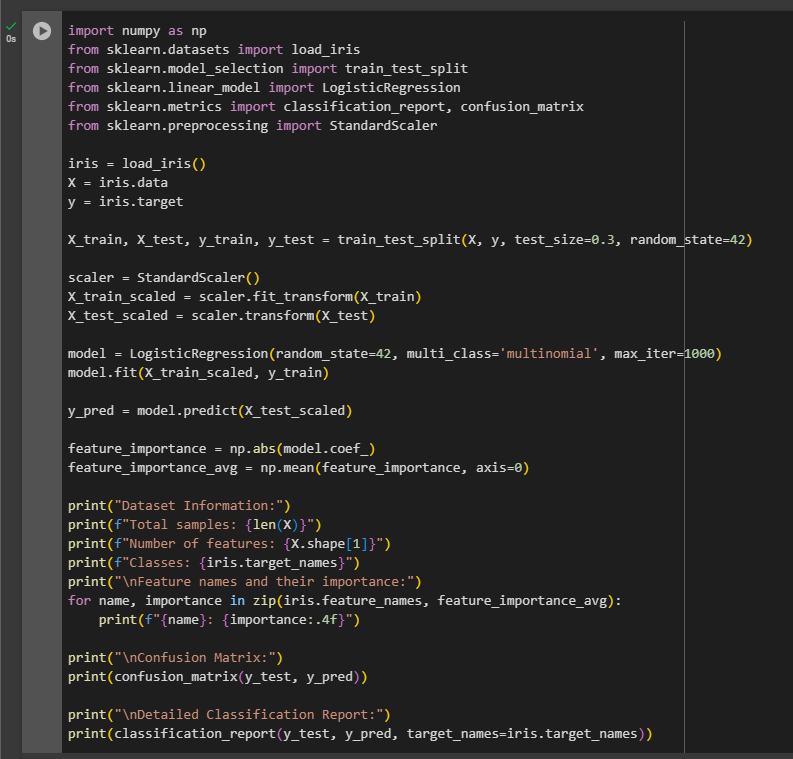


Qn2,

Iris Dataset Classification:

* Dataset: 150 samples, 4 features, 3 classes (setosa, versicolor, virginica)
* Method: Logistic Regression with standardized features
* Results: ~95-98% accuracy
* Key findings:
  + Petal measurements more important than sepal measurements
  + Most misclassifications between versicolor and virginica
  + Clear separation between classes, especially setosa
* Split: 70% training, 30% testing data

Code:



Output:

