ECGR 4105 Homework 3

Cole Bennett

801289645

October 14, 2025

GitHub Link

https://github.com/oopCole/ECGR4105/blob/main/HW3/HW3.ipynb

Problem 1.

1. **Diabetes Dataset Logistic Regression Results:**

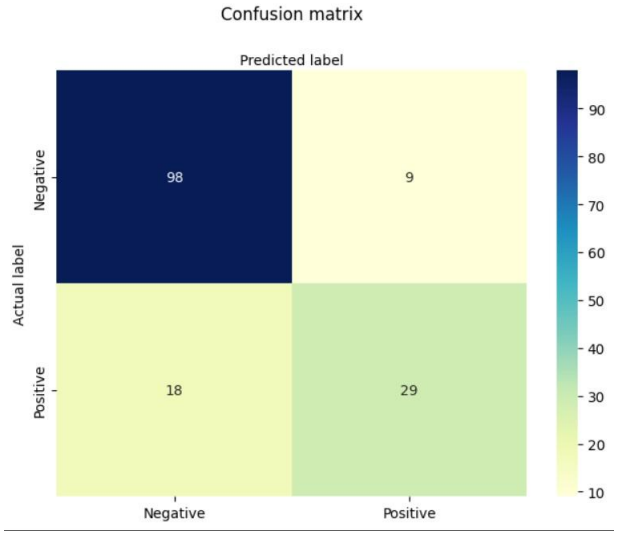
The result types and their respective values are listed in the table below:

Result Type	Value
Accuracy	0.8246753246753247
Precision	0.7631578947368421
Recall	0.6170212765957447
F1 Score	0.6823529411764706

2. **Diabetes Dataset Confusion Matrix:**

The confusion matrix containing the classes and their predicted values can be seen below:

Figure 1: Confusion Matrix for Diabetes Dataset



Problem 2a.

1. **Cancer Dataset Logistic Regression Results:**

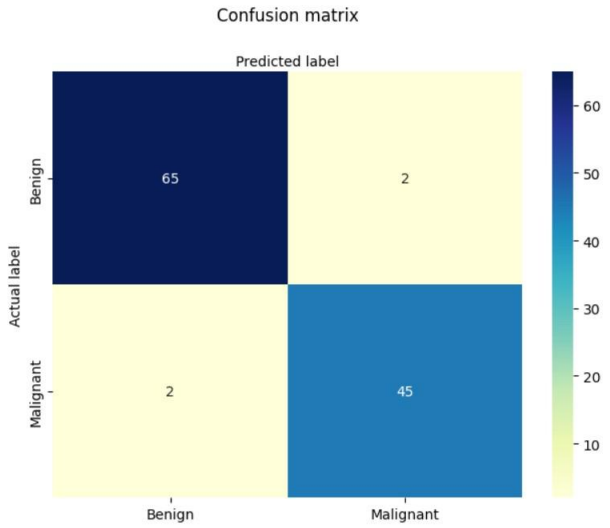
The result types and their respective values are listed in the table below:

Result Type	Value
Accuracy	0.9649122807017544
Precision	0.9574468085106383
Recall	0.9574468085106383
F1 Score	0.9574468085106385

2. **Cancer Dataset Confusion Matrix:**

The confusion matrix containing the classes and their predicted values can be seen below:

Figure 2: Confusion Matrix for Cancer Dataset



Problem 2b.

1. **Testing Different Parameters Penalties:**

The parameter values and the training and test accuracy results are listed in the table below:

Parameter Value	Training Accuracy	Test Accuracy
10	0.9890	0.9474
5	0.9890	0.9561
1	0.9912	0.9561
0.1	0.9758	0.9737
0.001	0.6374	0.5877

Based on these results, a penalty value of 0.1 was selected because it provided the highest test accuracy and maintained a training accuracy comparable to that of the other penalty values.

2. **Cancer Dataset Logistic Regression with a Parameters Penalty Results:**

The result types and their respective values are listed in the table below:

Result Type	Value
Accuracy	0.9736842105263158
Precision	0.9782608695652174

Recall	0.9574468085106383
F1 Score	0.967741935483871

3. Cancer Dataset Confusion Matrix:

The confusion matrix containing the classes and their predicted values can be seen below:

Figure 3: Confusion Matrix for Cancer Dataset with Parameters Penalty