Faculty of Engineering and Technology

Electrical and Computer Engineering Department

Machine Learning and Data Science - ENCS5341

Assignment #2

Prepared by: Lana Batnij __ 1200308

Instructor: Dr.Yazan Abu Farha

Section: 2

Table of Contents

- Assignment Progress and Results Discussion: 4

1. Part 1: 4

   1.1 Question: 4

   1.2 Question: 5

   1.3 Question: 8

2. Part 2: 9

   2.1 Question 1: 9

   2.2 Question2: 11

   2.3 Question 3: 11

- ***Table of Figures:***

Figure 1: Scatter plot of the three sets - A 4

Figure 2: Scatter plot of the three sets – B........ 4

Figure 3: Polynomial Regression - Degrees 1 and 2........ 5

Figure 4: Polynomial Regression - Degrees 3 and 4........ 5

Figure 5: Polynomial Regression - Degrees 5 and 6........ 6

Figure 6: Polynomial Regression - Degrees 7 and 8........ 6

Figure 7: Polynomial Regression - Degrees 9 and 10........ 7

Figure 8: Validation Error vs Polynomial Degree Graph........ 7

Figure 9: MSE on Validation vs Regularization Parameter 8

Figure 10: Best Regularization Parameter 9

Figure 11: Training and Testing Sets 9

Figure 12: Training and Testing Sets Accuracy 9

Figure 13: Logistic regression model with a linear decision boundary 10

Figure 14: logistic regression model with quadratic........11

- ***Assignment Progress and Results discussion:***

1. *Part 1:*

1.1 Question:

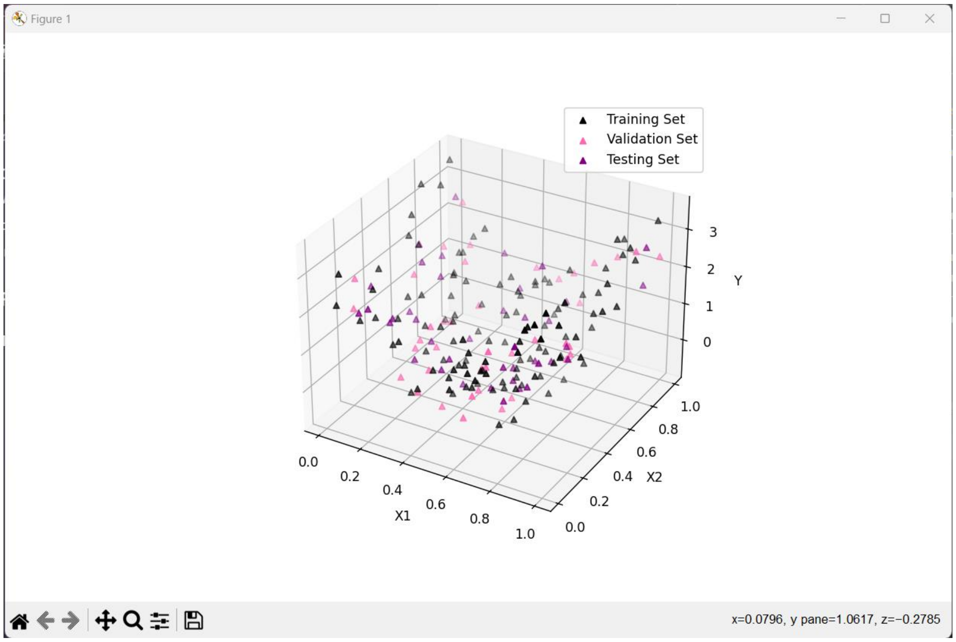

Figure 1: Scatter plot of the three sets - A

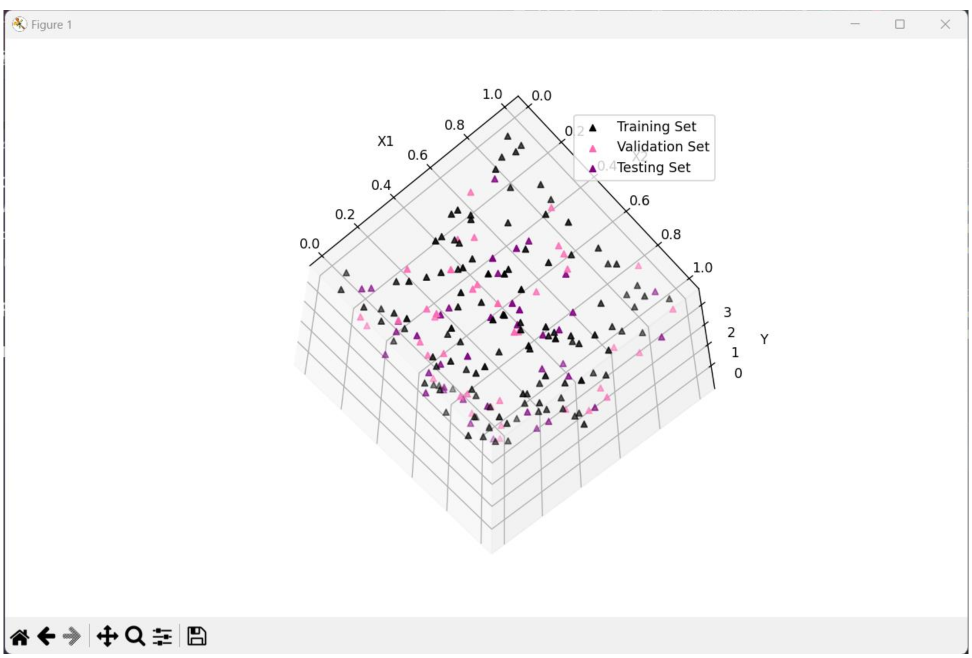

Figure 2: Scatter plot of the three sets – B

The set of data was obtained from "data_reg.csv," and it contains 200 samples with two characteristics (x1 and x2) and a continuous target label (y). A scatter diagram of the training, validation, and testing sets shows the different groups. The training set is the largest since it has 120 examples, while the other has 40.

1.2 Question:

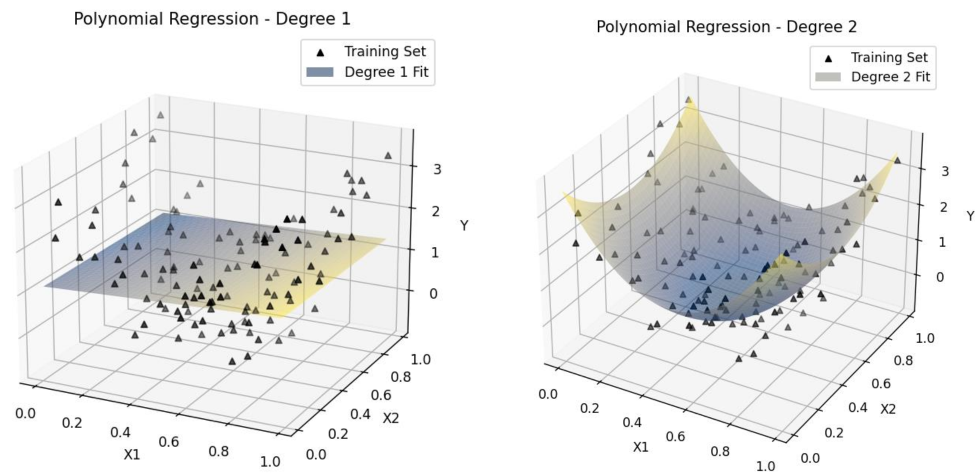

Figure 3: Polynomial Regression - Degrees 1 and 2

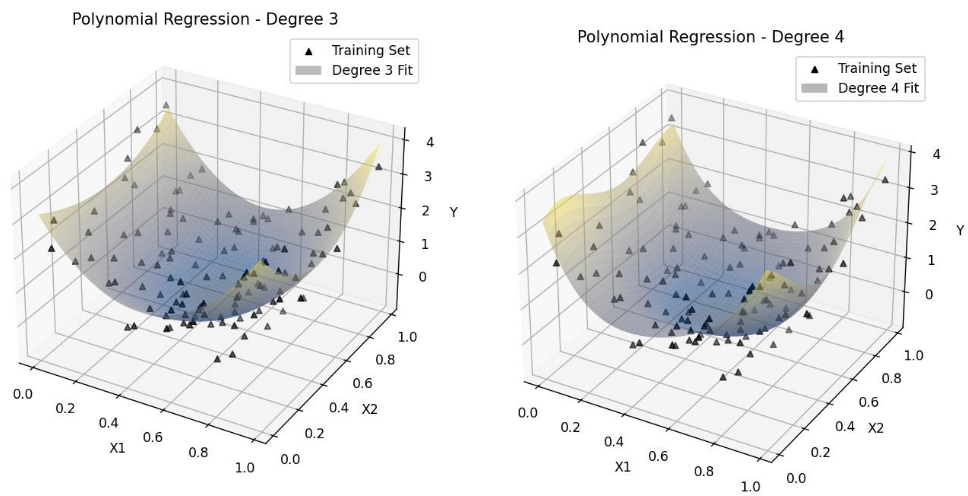

Figure 4: Polynomial Regression - Degrees 3 and 4

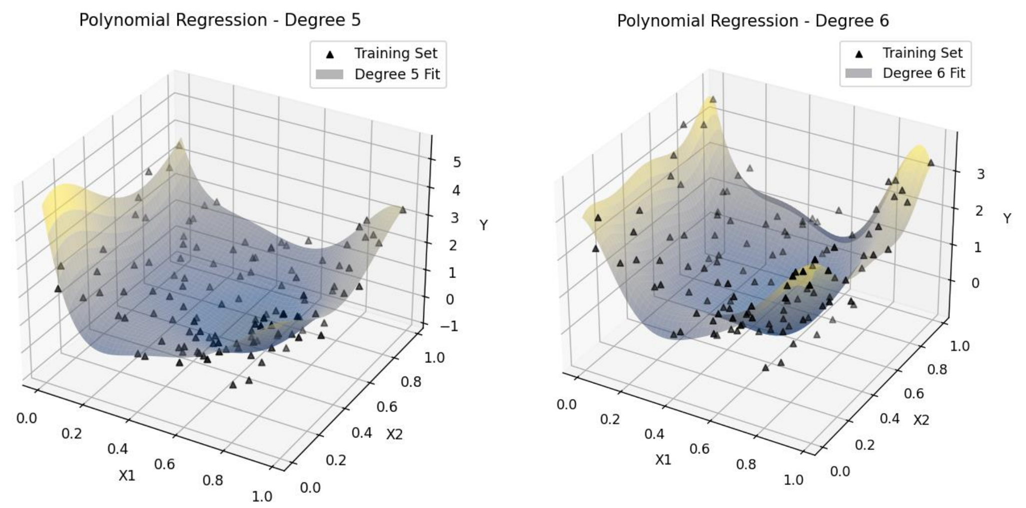

Figure 5: Polynomial Regression - Degrees 5 and 6

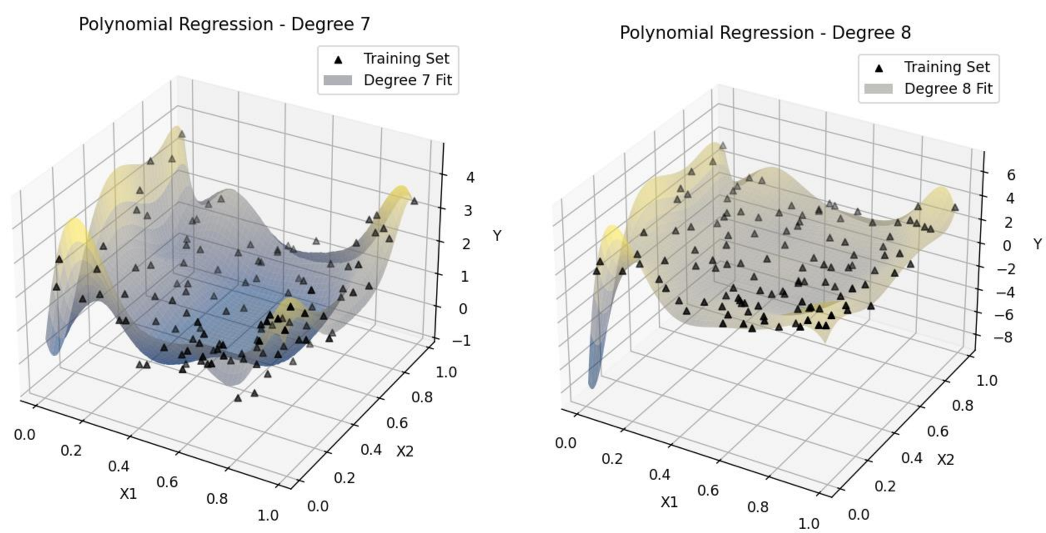

Figure 6: Polynomial Regression - Degrees 7 and 8

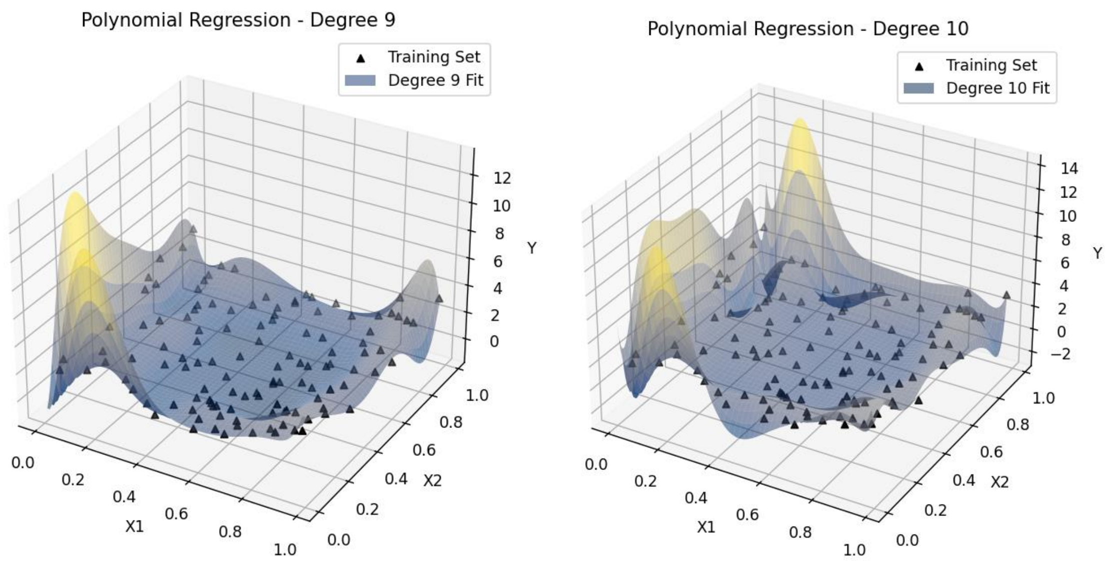

Figure 7: Polynomial Regression - Degrees 9 and 10

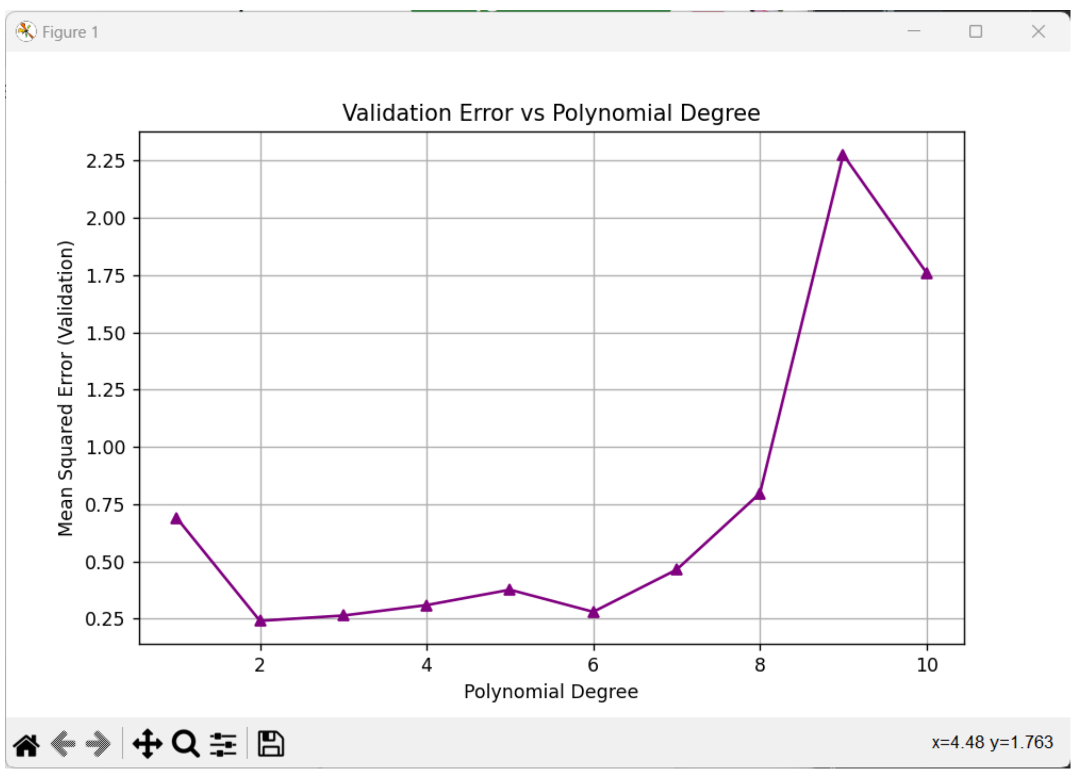

Figure 8: Validation Error vs Polynomial Degree Graph

Polynomial regression, with degrees ranging from one to ten, identifies non-linear connections among features and the objective. This method offers more flexibility when analyzing complex data patterns. The growth of model complexity with increasing degrees is presented by viewing the learned function surfaces for each degree with training examples. The validation error curve is used to determine the best polynomial degree. It would be the degree to which the validation error has been reduced in the context of the data being provided. As Figure 8 shows, degree 2 is the best polynomial since it has the lowest Validation error and a good balance between underfitting and overfitting. Degree 6 has also a low validation error but its learned surface tells that it is very close to being overfitting, so the preferred one is degree 2.

1.3 Question:

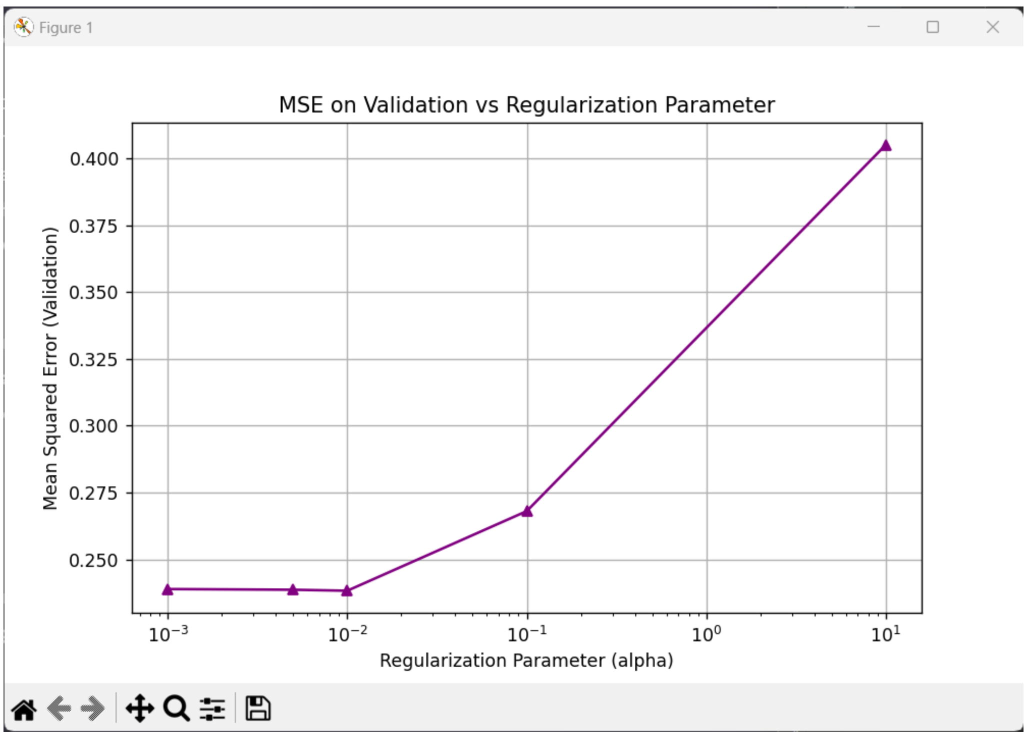

Figure 9: MSE on Validation vs Regularization Parameter

```
main
C:\Users\LENOVO\OneDrive\Desktop\ENCS4320\MLA2\venv\Scripts\python.exe C:\
Best Regularization Parameter (alpha): 0.01
Columns of train_cls_data: Index(['x1', 'x2', 'class'], dtype='object')
```

Figure 10: Best Regularization Parameter

A polynomial model of degree 8 was applied to the training set using Ridge regression, with the best regularization parameter set at 0.01. By displaying the Mean Squared Error (MSE) on the validation set versus the regularization parameter, the plot indicates the relationship between model complexity and generalization performance. The selected regularization parameter finds a balance between preventing overfitting and emphasizing the importance of regularization in improving model stability and dependability.

2. Part 2:

2.1 Question 1:

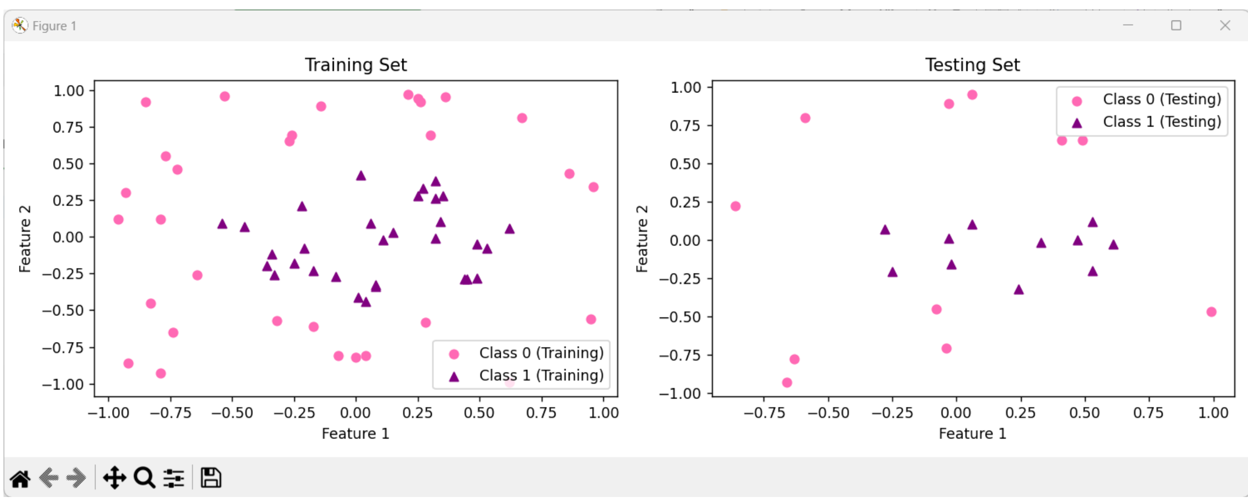

Figure 11: Training and Testing Sets

```
Training Accuracy: 0.6612903225806451
Testing Accuracy: 0.6818181818181818
```

Figure 12: Training and Testing Sets Accuracy

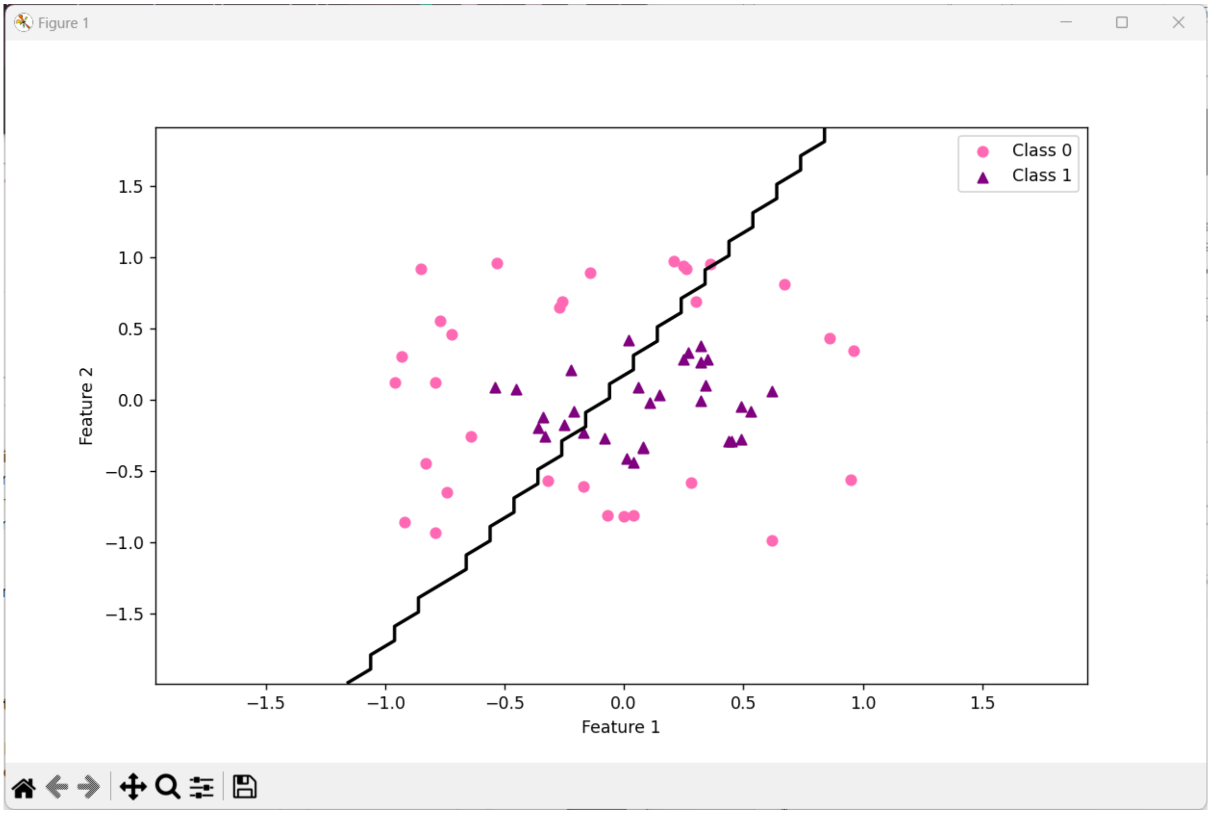

Figure 13: Logistic regression model with a linear decision boundary

The logistic regression model with a linear boundary for decision-making was determined to have a training accuracy of 66.13% and a testing accuracy of 68.18%. On the training set scatterplot, near to linear characteristics of the decision boundary were graphically shown, successfully dividing the dataset into two sections. even with its simplicity, the model obtained an acceptable precision on both the training and testing sets.

2.2 Question2:

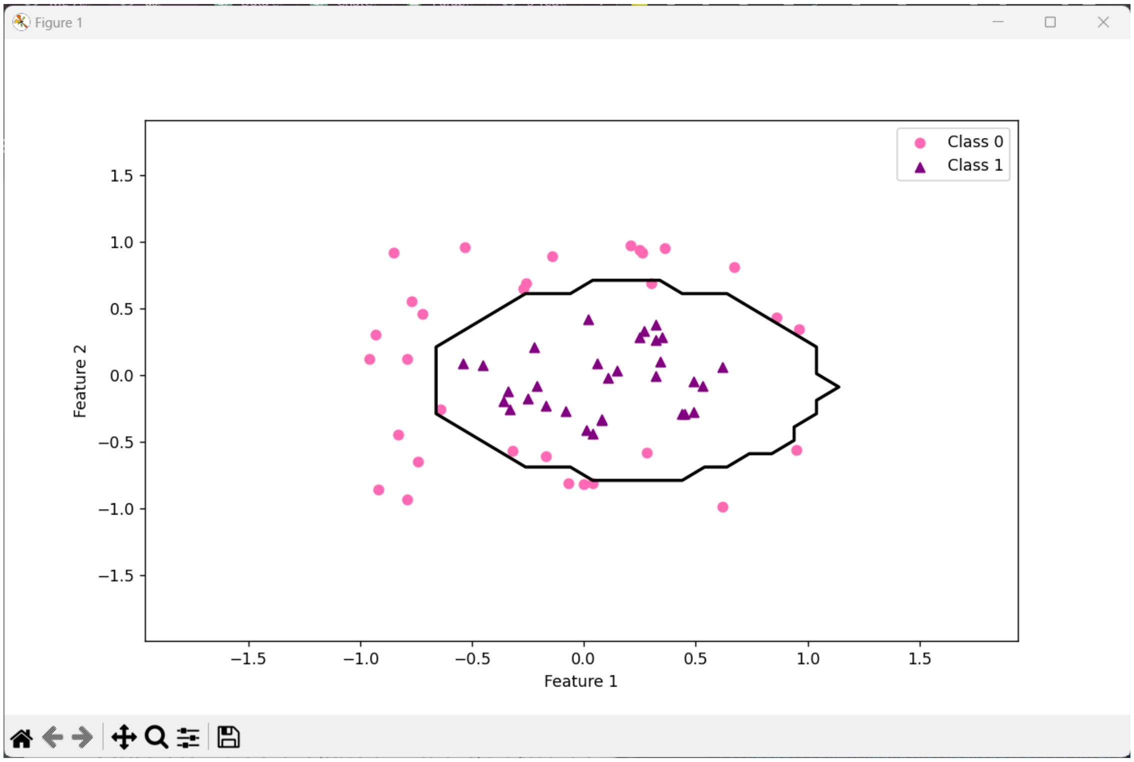

Figure 14: logistic regression model with quadratic

The logistic regression model performed much better after including a quadratic decision boundary, with a training accuracy of around 96.77% and a testing accuracy of roughly 95.45%. On the training set scatterplot, the graphical representation of the decision boundary indicated a more detailed pattern, closed to be a circular shape that efficiently separated the classes, and it enhanced precision.

2.3 Question 3:

Part 1's logistic regression model with a linear decision boundary has a reasonable level of accuracy, with a training accuracy of about 66.13% and a testing accuracy of roughly 68.18%. The related decision boundary, which appears as a nearly straight line dividing the dataset, denotes a degree of simplicity. This simplicity may indicate an underfitting scenario, in which the model struggles to represent the underlying data patterns' complexity.

Part 2's logistic regression model with a quadratic decision boundary, obtains much greater accuracy, with a training accuracy of around 96.77% and a testing accuracy of around 95.45%.

The complex decision boundary, which features an almost circular outline format, indicates a better match to the complex dataset. While high accuracy is obtained.