**Report – Lab 9**

The dataset was split in an 80:20 (train: test) ratio and Mean Squared Error (MSE) score was used to evaluate the performance of the models.

**Linear Regression**

On running a 10-fold cross validation, the mean MSE score was as follows:

A screenshot of a computer

Description automatically generated

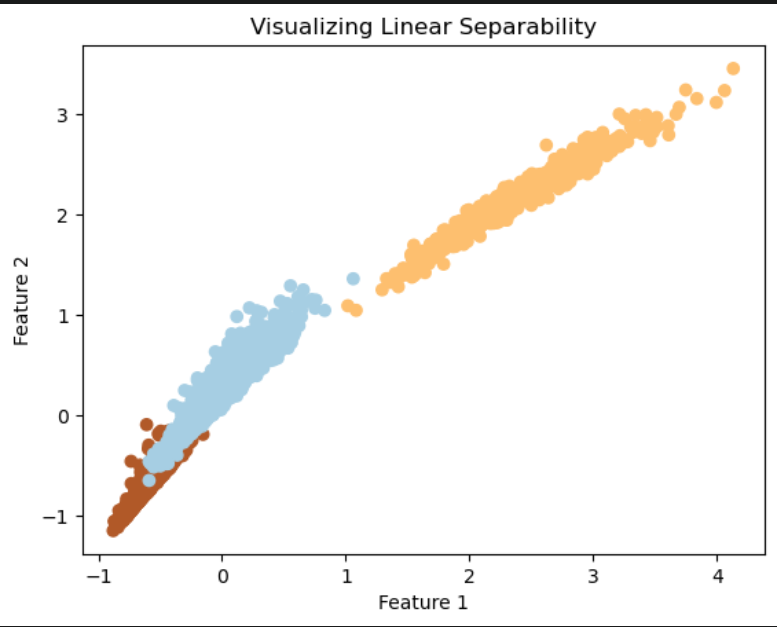
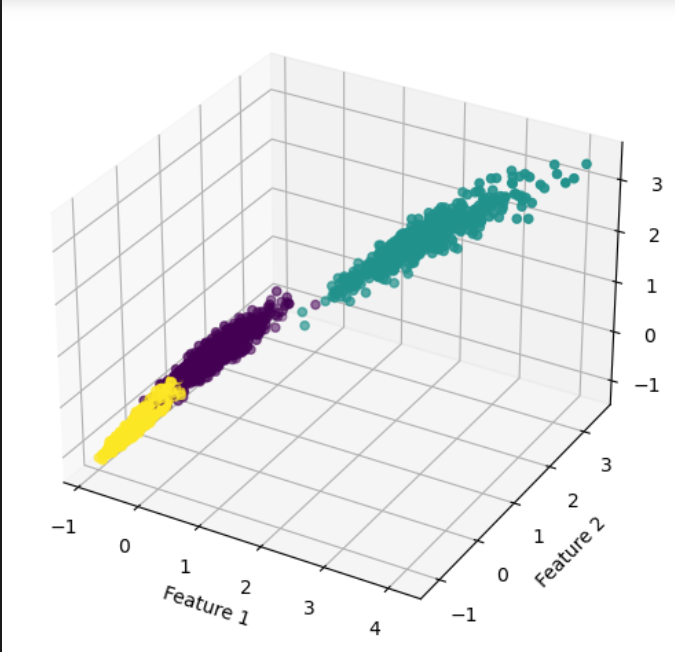
On evaluating the model on train and test data, the MSE score was as follows:

A screenshot of a computer

Description automatically generated

Since the MSE scores are very close in both the above cases, we can conclude that the linear regression model generalized well to unseen data.

On visualizing the data in 2D and 3D space, we can come up with a hypothesis that the data is linearly separable.

**Logistic Regression**

On running a 10-fold cross validation, the mean MSE score was as follows:

A screenshot of a computer

Description automatically generated

On evaluating the model on train and test data, the MSE score was as follows:

A screen shot of a computer code

Description automatically generated

The difference in the MSE scores isn’t much but is greater than the Linear Regression Model. This shows that Linear Regression model generalizes to unseen or new data better than the Logistic Regression model.

The MSE score of the Logistic Regression model is lower than the Linear Regression Model. This shows that on average, the model's predictions are closer to the actual or true values. In other words, the Logistic Regression models makes more accurate predictions compared to the Linear Regression Model.

The Coefficient of Determination represents the goodness of fit of the model. The closer the value is to 1, the better the fit. The Logistic Regression model achieves a higher score for this measure compared to Linear Regression model. This indicates a better fit of the model to the data.

**Hyper Parameter combinations**

The tables below show how different values of the hyperparameters of Logistic Regression model affect MSE Score:

1. Penalty

|  |  |
| --- | --- |
| Penalty | Mean MSE Score |
| **‘none’** | **0.0465** |
| ‘l2’ | 0.0452 |

1. Number of Iterations

|  |  |
| --- | --- |
| Number of Iterations | Mean MSE Score |
| 10 | 0.0468 |
| **50** | **0.0452** |
| 100 | 0.0452 |
| 200 | 0.0452 |
| 300 | 0.0452 |

1. Solver

|  |  |
| --- | --- |
| Solver | Mean MSE Score |
| ‘liblinear’ | 0.0465 |
| ‘lbfgs’ | 0.0452 |
| **‘saga’** | **0.0439** |

1. Class Weight

|  |  |
| --- | --- |
| Class Weight | Mean MSE Score |
| **None** | **0.0452** |
| ‘balanced’ | 0.053 |

1. C: Inverse of Regularization Strength

|  |  |
| --- | --- |
| C | Mean MSE Score |
| 0.001 | 0.0811 |
| 0.01 | 0.0478 |
| 0.1 | 0.0452 |
| **1** | **0.0452** |
| 10 | 0.0465 |
| 100 | 0.0452 |