

CSE 6363 Assignment 1

Introduction:

This report details the implementation and evaluation of linear models for both regression and classification.

PART 1

Regression with linear Models:

All three methods i.e., Fit, Predict and Score were implemented. The results for the models are as follows:

Model Number	Input Features	Output Features	Mean Squared Error
1	Sepal length and Width	Petal length	0.6219551309140072
2	Sepal length and Width	Petal width	0.3551326715740868
3	All features	Petal length	0.018939452794874734
4	All features	Petal width	0.07675422144816718

	Input Features	Output Features	Mean Squared Error
1	Sepal length and width	Petal length and Width	0.6440312507984517

Execution Steps:

- Execute all of the train_regression*.py files to generate weights.
- After weights are generated its evaluated by eval_regression*.py files
- For execution of (1.7 Regression with multiple) run regression_with_multiple.py and evaluate it with eval_regression_with_multiple.py

1. We predicted the dimensions of the iris flower's petals using different input features and the Mean Squared Error (MSE) metric. Comprehensive feature utilization is critical for predictive modeling, with Model 3 providing consistently accurate predictions. Sepal dimensions are also meaningful for petal size prediction. Models 3 and 2 are the most proficient choices for regression tasks due to their consistently low MSE values.

In Part 2 of our Classification Models series, we aimed to categorize iris flowers into distinct groups using Logistic Regression and Linear Discriminant Analysis (LDA). Our primary focus was on assessing the performance of these models through accuracy metrics.

Logistic Regression as the Winner: Among the two classification models, Logistic Regression performed the best. It consistently achieved high levels of accuracy and, in some cases, even achieved a perfect accuracy score of 1.0. This robust performance underscores Logistic Regression's suitability for binary and multi-class classification tasks.

Feature Selection Plays a Crucial Role: Our observations highlighted the pivotal role of feature selection in classification accuracy. We found that including all available features consistently led to superior accuracy, emphasizing the significance of comprehensive feature consideration.

Linear Discriminant Analysis (LDA): Although LDA showed respectable performance, mainly when all features were employed, it consistently trailed Logistic Regression in accuracy. For our dataset and classification task, Logistic Regression presents as the preferred choice.

In summary, our analysis demonstrated the criticality of feature selection in both regression and classification tasks. Models 3 and 2 were the optimal choices for Regression due to their low MSE values. For classification, Logistic Regression exhibited exceptional accuracy across various feature sets, making it the preferred choice for sorting iris flowers into different classes. In Part 2 of our Classification Models series, we aimed to categorize iris flowers into distinct groups using Logistic Regression and Linear Discriminant Analysis (LDA). Our primary focus was on assessing the performance of these models through accuracy metrics.

PART 2

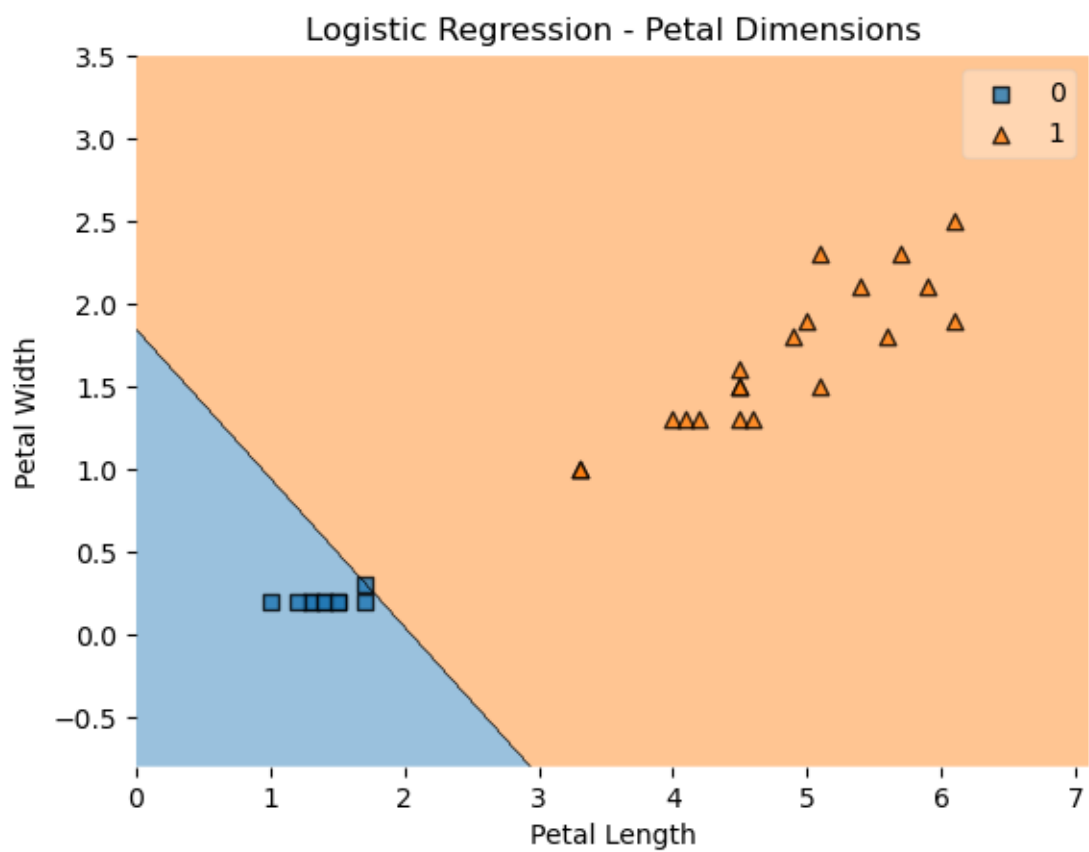
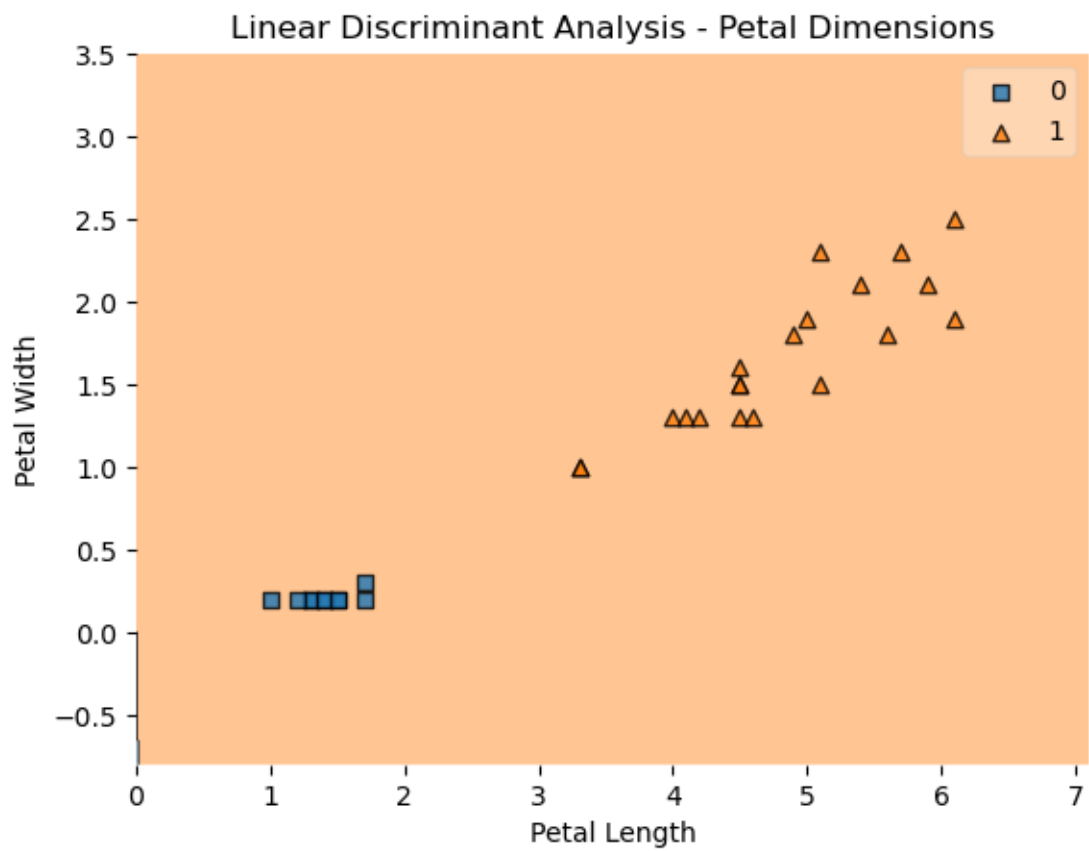
The accuracies of the classifiers on the test data were:

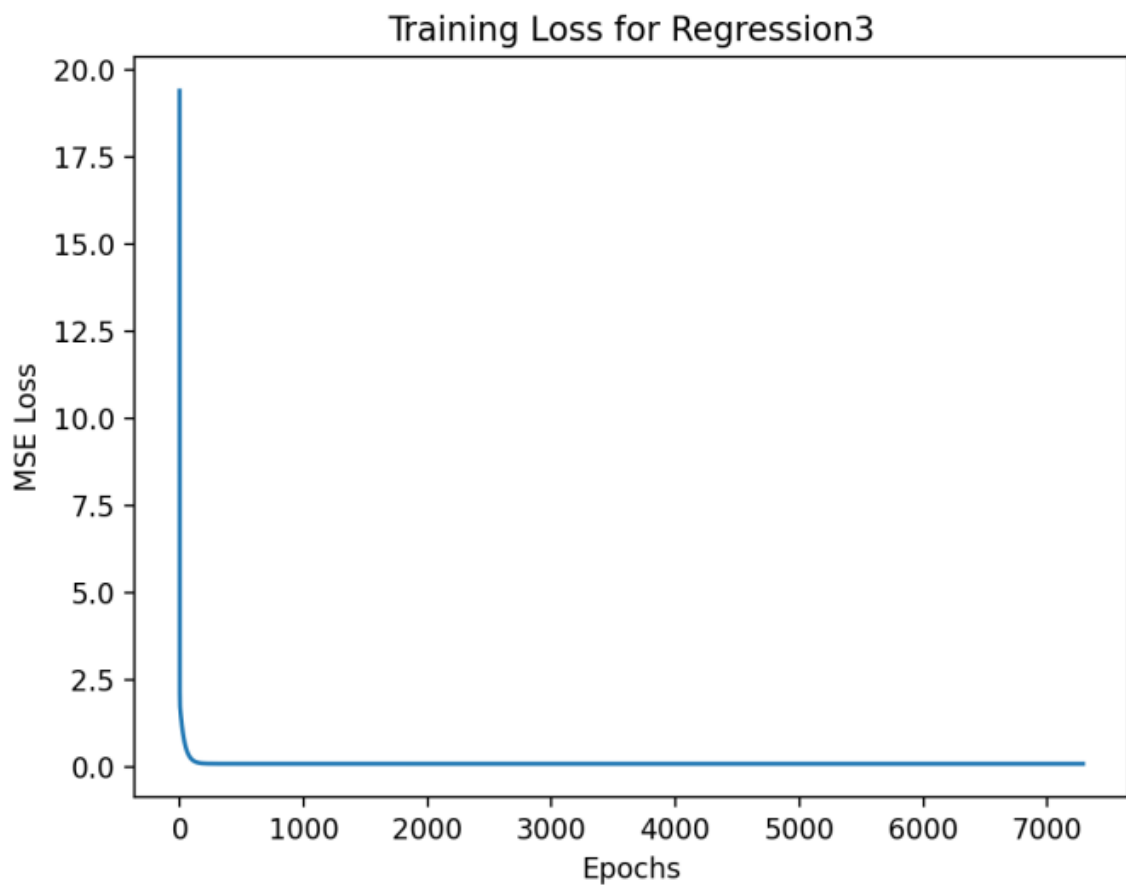
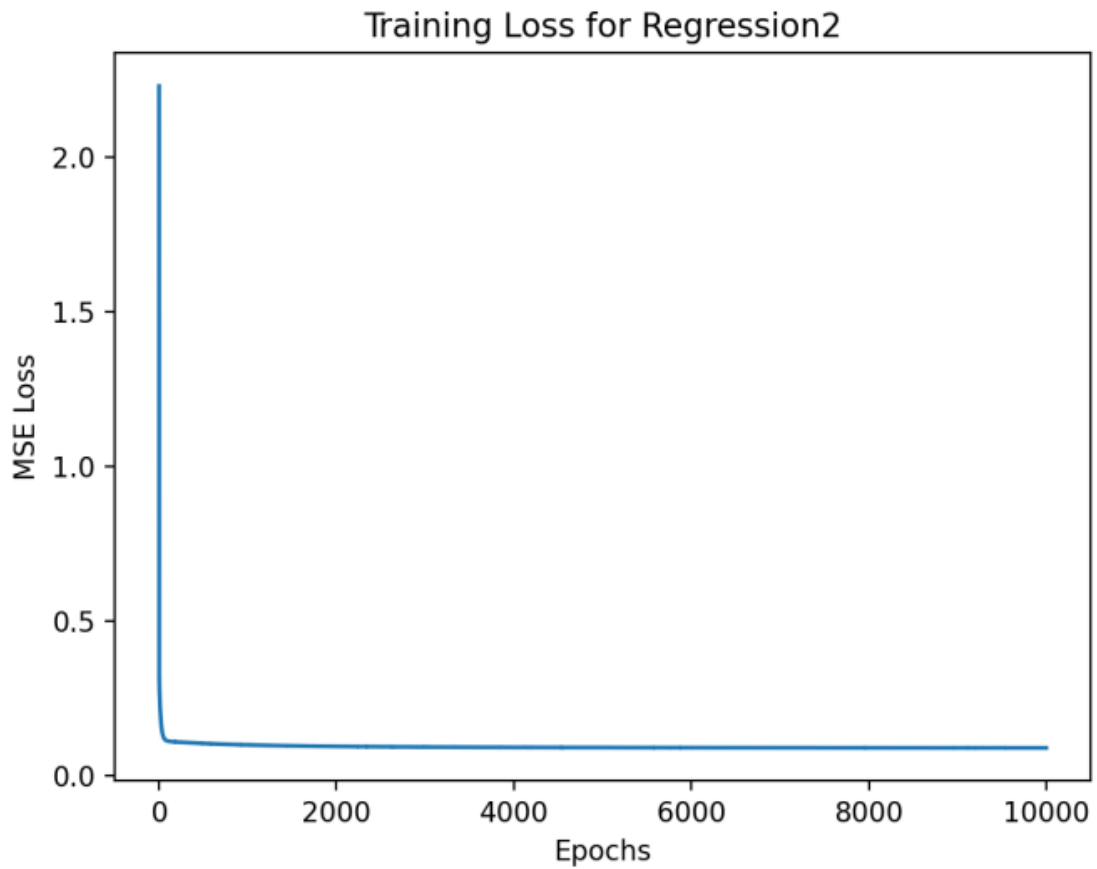
- Logistical Regression:
 - Petal length/width – 1.0
 - Sepal length/width – 0.9666666666666667
 - All features – 1.0
- Linear Discriminant Analysis:
 - Petal length/width – 0.6666666666666666
 - Sepal length/width – 0.7333333333333333
 - All features – 0.8666666666666667

Execution Steps:

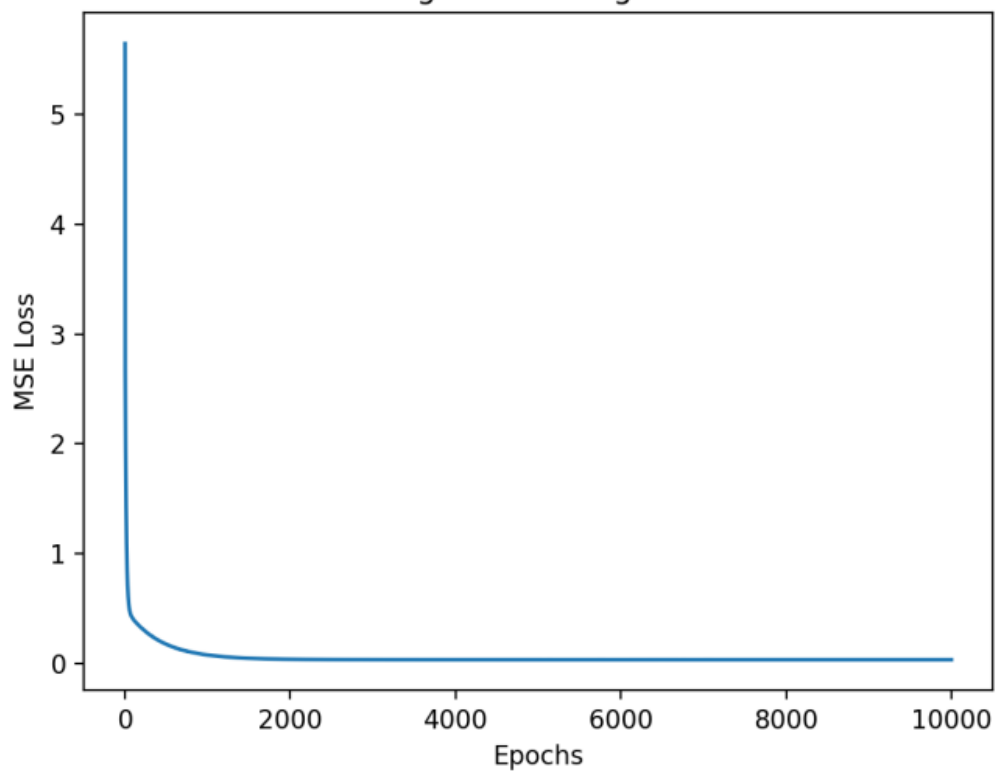
- Execute all of the three eval_classifiers*.py files the same file includes the training and testing as soon as the .npz trained file is generated.
- For execution of graph run the Visulaizations.py

The Visualized graph generated:





Training Loss for Regression4



Loss vs. Epochs

