APP PHY 157 WFY-FX-2

LAB REPORT 8

Machine Learning: Perceptron and Logistic Regression

Source code here!

LOVELY L. ANDEO
2020-05405

https://github.com/lovelyandeo/AP157-Computational-Analysis-and-Modeling-in-Physics

Background

Machine learning (ML) is a field within artificial intelligence that utilizes data and algorithms to mimic human learning patterns, with the ultimate goal of achieving accurate predictions or insights when presented with new and unfamiliar data. In this lab report, we will be mainly focusing on some ML algorithms such as Perceptron and Logistic Regression, and see how they perform with our data.

Objectives

In this activity, we aimed to:

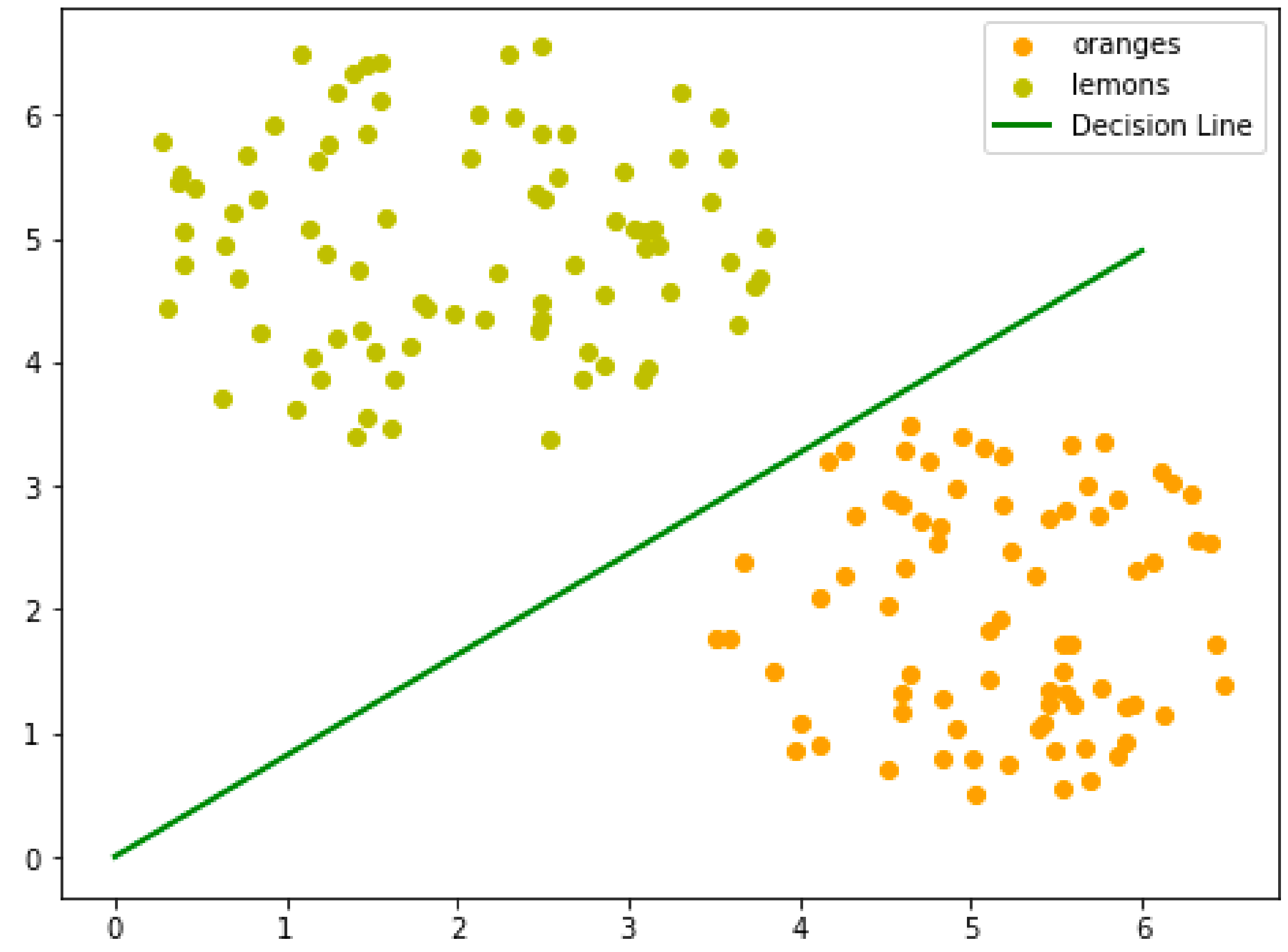
1. Familiarize with the perceptron and logistic regression algorithms
2. Use perceptron to classify two different classes
3. Use logistic regression to identify the ripeness of a fruit

Results and Analysis

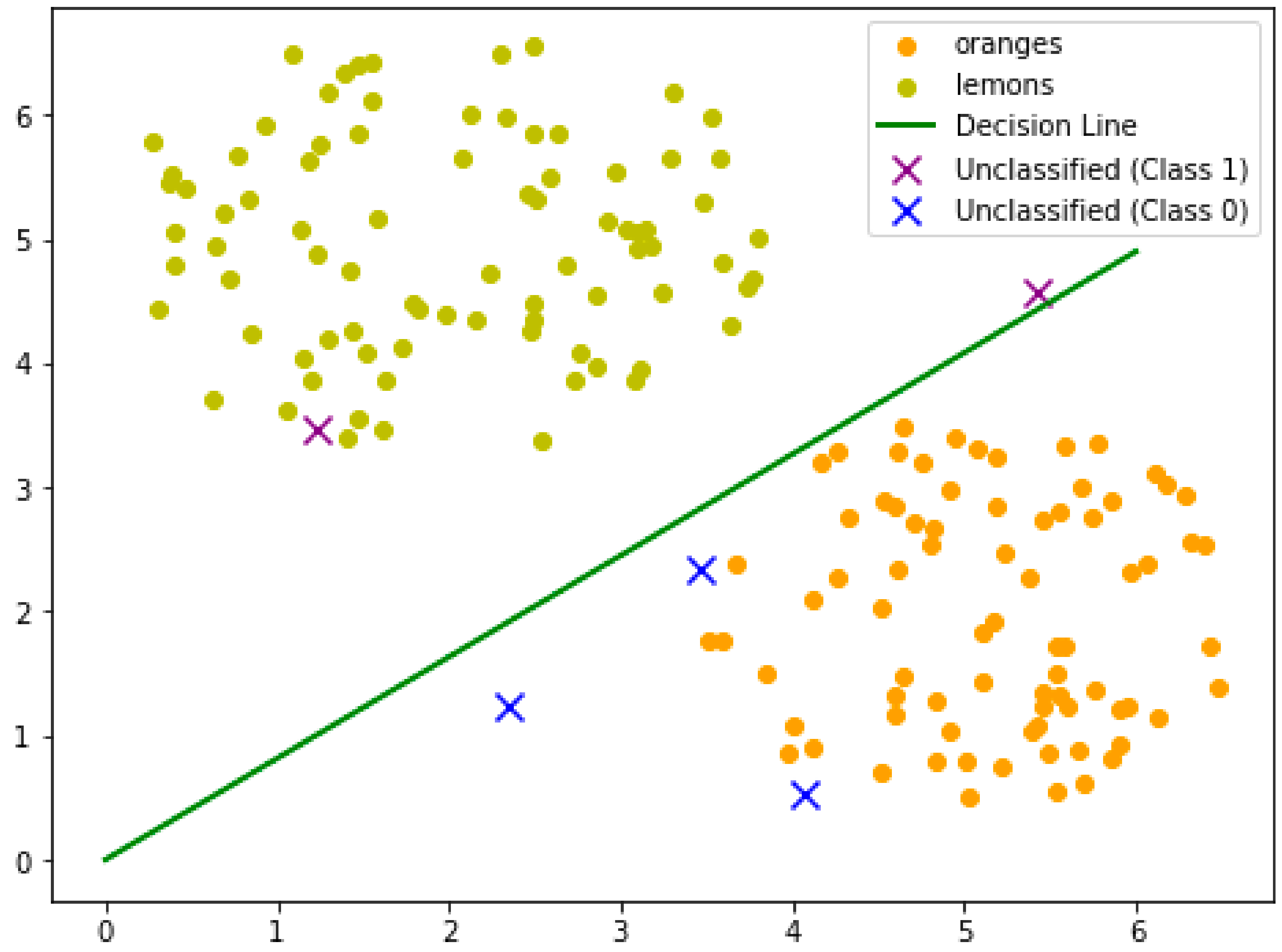
The Perceptron is often regarded as the simplest neural network model, functioning as a binary classifier to determine whether an input belongs to class A or class B.

In this report, I employed the Perceptron algorithm to classify two distinct fruits: oranges and lemons. Specifically, I assigned lemons as class 1 and oranges as class 0. Due to technical limitations and the inability to synchronize a large number of fruit images with GitHub, I opted to generate random fruit features instead that will be used for the classification. This allowed me to showcase the workings of the Perceptron algorithm effectively. The resulting classification of the fruits is depicted in the figure on the right, where the decision line serves as the boundary separating the two classes.

Although I initially intended to extract features from the fruits dataset containing 360 images (found in the Other trials folder), practical constraints led me to adopt the random feature generation approach. Nonetheless, the objective of demonstrating the Perceptron algorithm's functionality was successfully achieved. And overall, this provides insights into the workings of the Perceptron algorithm and showcases its ability to classify fruits based on their features.



Perceptron

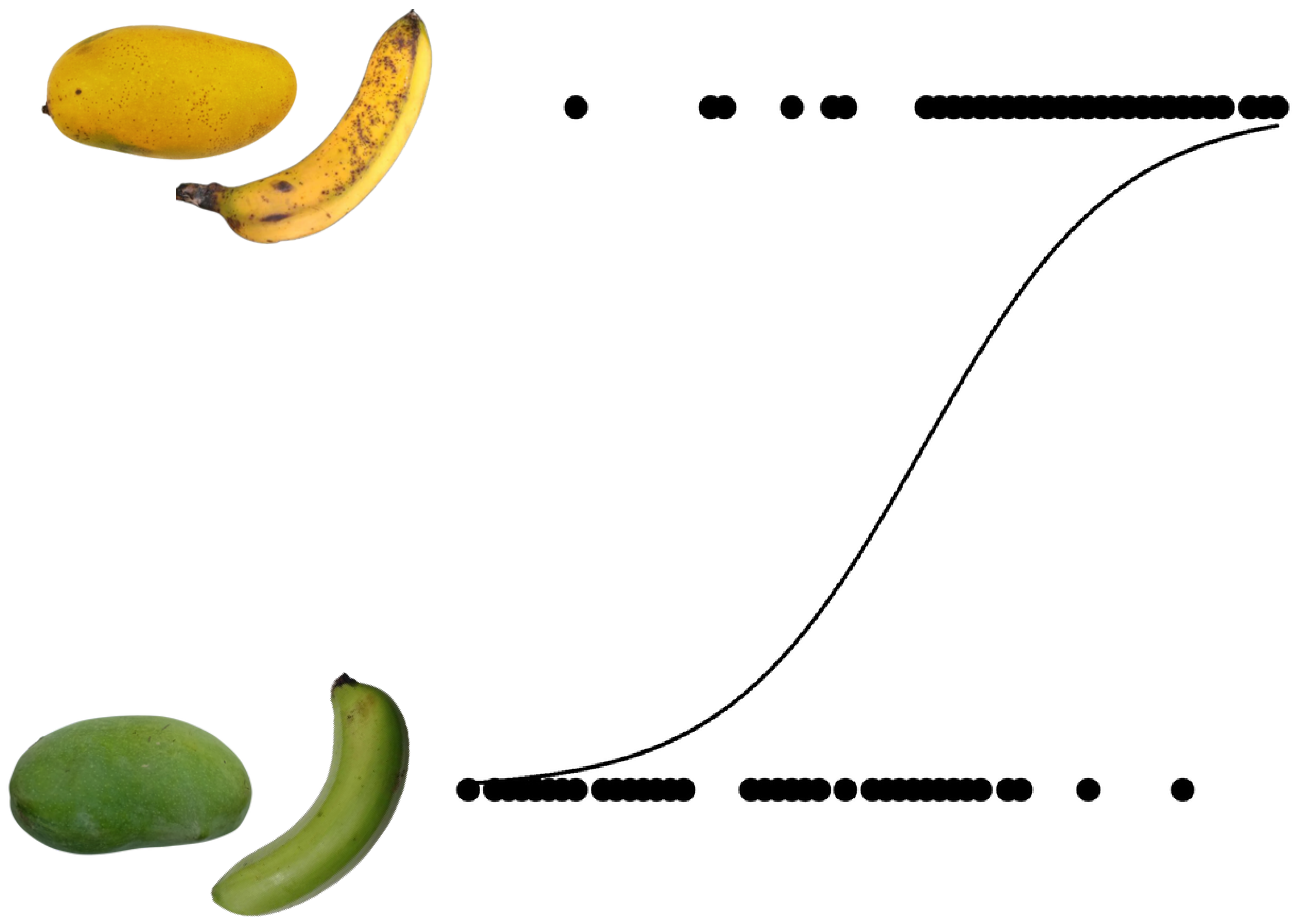


In this next part, we put the Perceptron to the test to assess its capability in classifying various other fruits based on the initial set of features provided. I carefully introduced five distinct fruits, each with its corresponding set of features, and then observed how the algorithm performed in assigning these fruits to their respective classes.

The graph on the left visually presents the outcome of this classification experiment. The x marks scattered across the graph represent the newly introduced fruits. Among them, the purple x marks indicate fruits that fall above the decision line, which means they have been classified as part of the lemon class. On the other hand, the blue x marks correspond to fruits positioned below the decision line, indicating that they belong to the orange class.

This compelling demonstration showcases the Perceptron's adaptability in accurately classifying not only the initial fruits but also a set of entirely new fruits based on their unique features. This robust performance further underscores the efficacy and reliability of the Perceptron algorithm as a powerful tool for fruit classification tasks.

Logistic Regression



Now, we shift our focus to a different model known as Logistic Regression. This machine learning approach is commonly employed to assess the likelihood or probability of a binary event taking place.

For the purposes of this lab report, we leverage Logistic Regression to tackle the task of determining the ripeness level of fruits. Determining the ideal ripeness of a fruit can often prove challenging, making it difficult to determine the optimal time for harvesting or consumption. However, by closely examining the changes in their colors, we can extract valuable insights regarding their ripeness levels. This enables us to estimate the percentage of ripeness of fruit, with 0% indicating complete unripeness and 100% representing full ripeness.

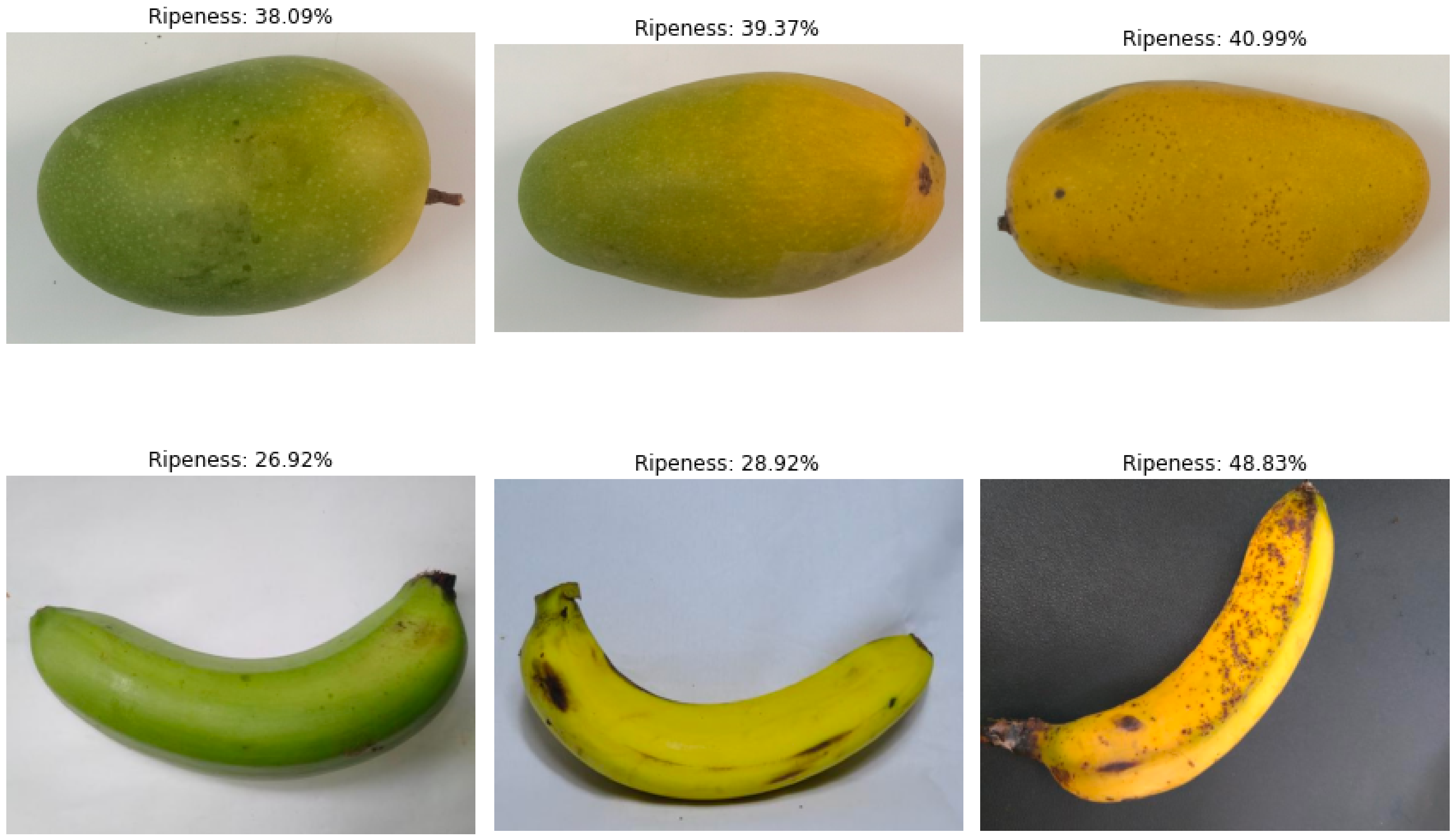
By harnessing the power of Logistic Regression, we aim to unveil the inherent patterns and correlations between the visual cues of fruits and their ripeness levels. In particular, we will be applying this model to mangoes and bananas.

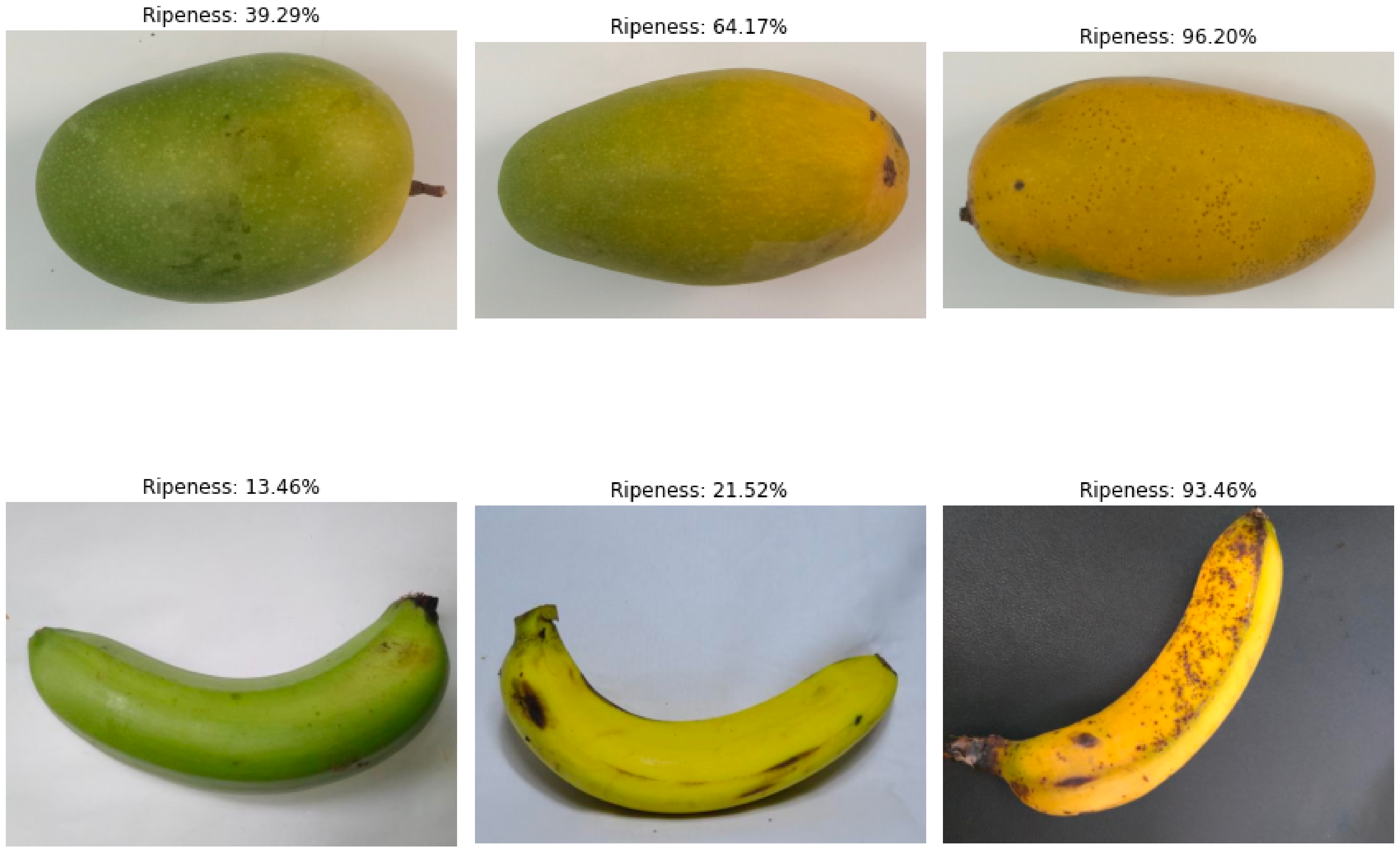
By employing the technique of Logistic Regression, we have conducted an analysis to determine the ripeness levels of fruits, and the results obtained from this analysis are depicted in the images displayed on the right-hand side. While there seems to be a general increasing trend in the percentage values as the fruits visually exhibit more ripeness, it is important to acknowledge that the accuracy of these predictions may not be entirely precise.

One of the primary factors contributing to the slightly diminished accuracy of the model could be attributed to the efficiency of the training set. It is crucial to have a well-curated and diverse training set that adequately captures the nuances and variations in the visual cues of ripeness across different types of fruits. If the training set is insufficient or lacks diversity, it may limit the model's ability to discern and differentiate between the various fruits with higher precision.

Upon closer inspection of the results, it is evident that the performance of the Logistic Regression model is relatively more favorable when applied to the mango dataset compared to the banana dataset. This discrepancy could be attributed to inherent differences in the visual characteristics and ripeness indicators between these two types of fruits.

There are two ways to optimize this result: using a more diverse training set and using a slightlymore robust model.





It's not always easy to find more data, sometimes we have to make do with the data that we have. Hence, I decided to go for a little extra challenge by using a different model. I believe that the art of machine learning lies not only in analyzing the data but also in selecting the most suitable model to extract meaningful insights.

In this specific dataset, I opted for the Multi-layer Perceptron (MLP) model. Known for its interconnected layers of neurons with non-linear activation functions, the MLP exhibits the capability to capture intricate patterns and relationships within the data through forward propagation and backpropagation. With that, it is expected to perform better in determining the ripeness of the fruits that we have.

And as anticipated, the images on the left showcase the impressive performance of the MLP in accurately estimating the ripeness percentages of the fruits, outshining the Logistic Regression model in terms of accuracy and precision. This proves the suitability of the MLP for handling complex tasks and highlights its superiority in delivering robust outcomes.

Key takeaway

- Thoroughly exploring the data is essential in selecting the optimal machine-learning model
- The data plays a crucial role in determining the most effective machine learning model for accurate predictions

Reflection

Exploring machine learning models is always a fun journey, especially when delving into the intricacies of each algorithm and witnessing their unique functionalities. Personally, I find great fascination in these aspects. However, one aspect that often presents a significant challenge for me is the data cleaning and preprocessing stage, which can be quite excruciating. Fortunately, I still managed to effectively showcase the remarkable capabilities of both the Perceptron and Logistic Regression models in this report. With this, I look forward to exploring and learning more about the different machine-learning models.

Self-evaluation

100/100

I believe I was able to deliver what was required for this lab report, as mentioned in the manual.

References

Here are the materials I used as guide to accomplish this activity:

Soriano, M. (2023). ML2 - Perceptron. https://uvle.upd.edu.ph/pluginfile.php/885073/mod_resource/content/1/ML2%20-%20Perceptron.pdf

Soriano, M. (2023). ML3 - Logistic Regression. https://uvle.upd.edu.ph/pluginfile.php/885075/mod_resource/content/1/ML3%20-%20Logistic%20Regression.pdf

Soriano, M. (2023). Machine Learning. https://uvle.upd.edu.ph/pluginfile.php/885060/mod_resource/content/1/ML1%20-%20Machine%20Learning%20Intro.pdf

Klein, B. (2022). Separating classes with dividing lines. https://python-course.eu/machine-learning/separating-classes-with-dividing-lines.php

Logistic Regression Dataset: https://universe.roboflow.com/fruit-ripening/banana-ripening-process