Logistic Regression

Objective

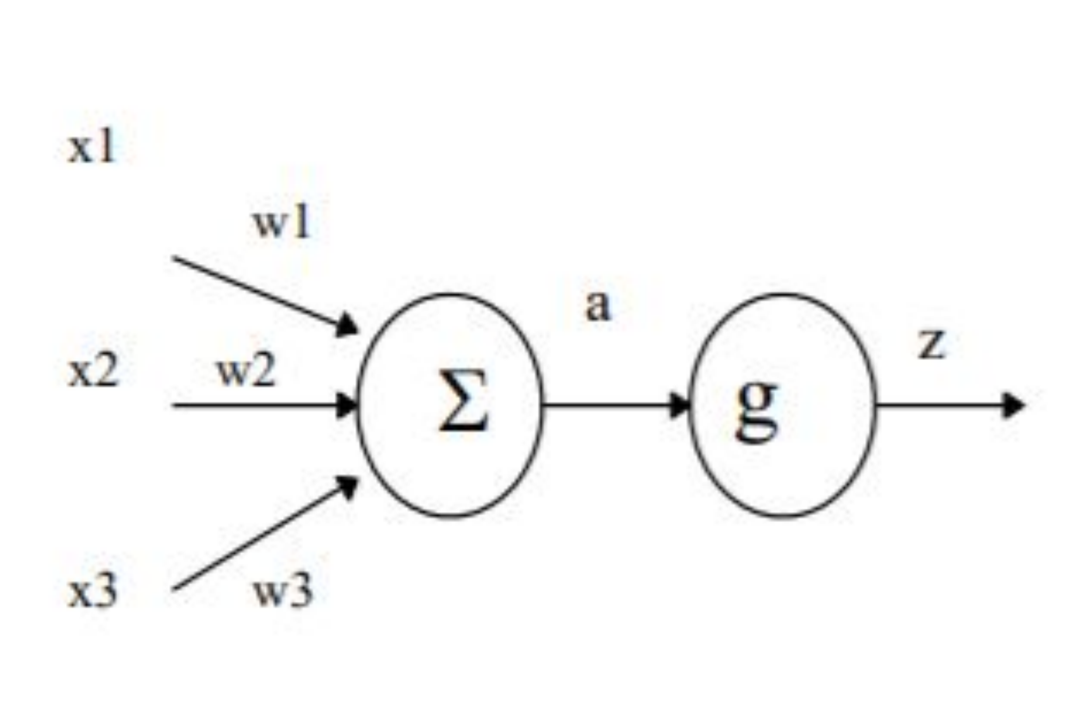
- The previous activity uses a step function as the activation function. (Binary classification)
- In this activity, the activation function is a sigmoid function. (Less binary, there is a gradation)
- The objective is to make a machine learning algorithm to classify the ripeness of a Mango fruit.

Summary of my Life

- Survive :)

Review: Perceptron

- To distinguish classes with features, weights are used to help in differentiating the classes.
- The sum of the product of weights **(x_j*w_j)** of each sample is determined as **a**.
- **a** is inputted on an activation function **g** that tries to identify the class with the result **z.**



Schematic Diagram on the Perceptron Algorithm

Review: Perceptron

- The weights are modified depending on the sum of the difference of the determined value **z** and the intended value per sample and the learning rate.
- This is iterated until a certain error is reached (0.01) or a number of iterations is reached.
- Using 2 features separates classes with a line while 3 features separates with a plane and so on.

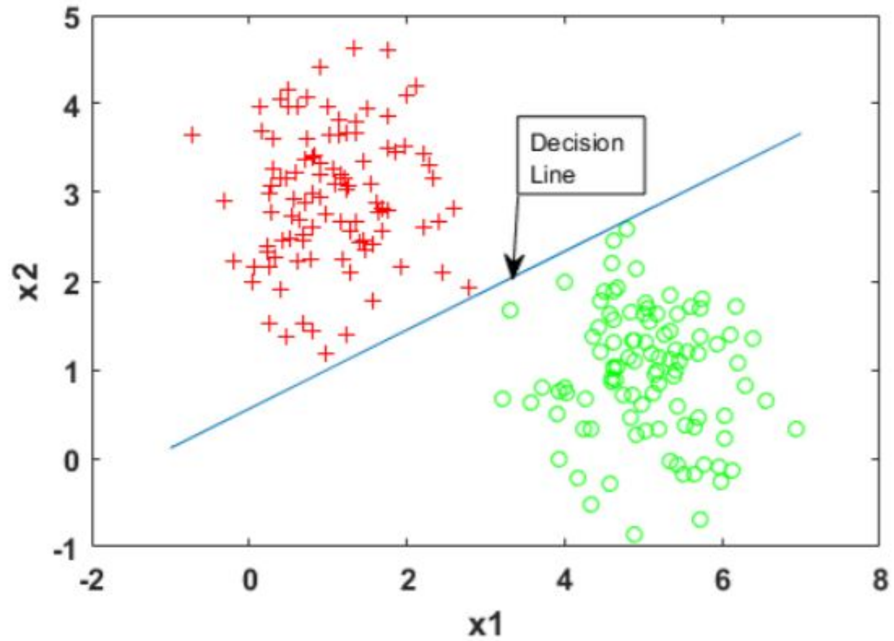


Figure 3: The two classes shown are linearly separable. A decision line can separate the two classes in feature space.

What’s New: Logistic Regression

- The difference of a step function from a sigmoid function is that step function is used for binary decisions (yes or no) while sigmoid function is used for probabilistic decisions (most likely, less likely)

- Graphically the sigmoid function has a smoother transition from one state to another than the step function.

Step Function

$$z = g(a) = \begin{cases} 1 & \text{if } a \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

Sigmoid Function

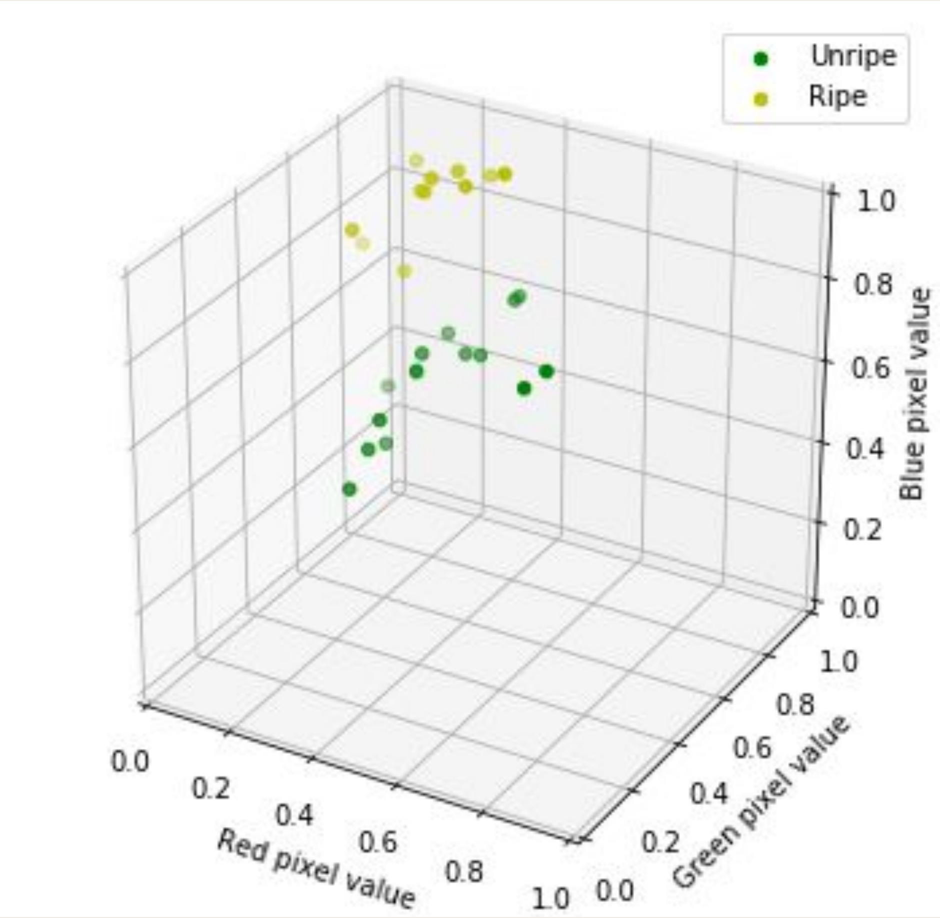
$$\phi(a) = \frac{1}{1+\exp(-a)} = p \quad .$$

Step 0: Get Data

- I used Mangoes classified into two classes (Ripe - yellowish, Unripe - greenish)

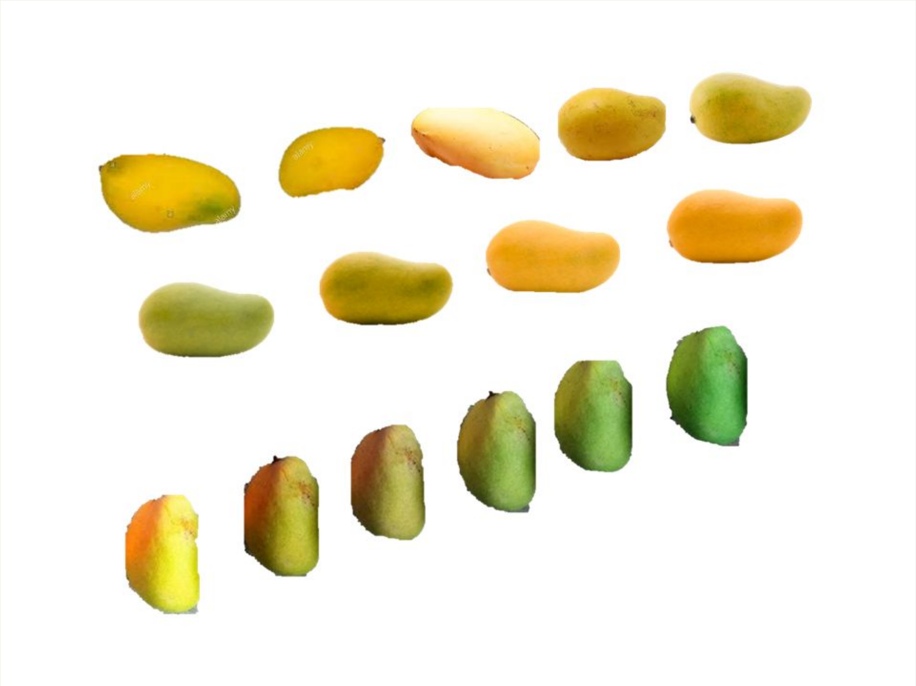
- I had 3 features corresponding the average RGB values of the mango fruit normalized from 0-255 to 0-1.

- These mangoes are going to be the training data where ripe mangoes are classified as d = 1 while d = 0 for unripe mangoes.

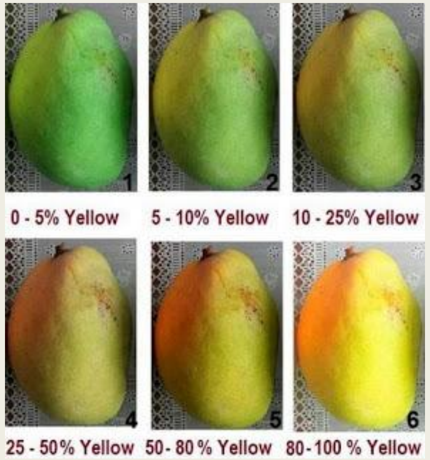


Step 0: Test Data

- To test the algorithm. I used images not used on the training data that are fully ripe, unripe and are those on transition. We will test the algorithm how accurate it is with these mangoes.



Test Data Matrix



Step 1: Initializing Weights/ Variables

- Weights are initialized randomly.
- **X_0** = 1 (bias used)
- Activation function used (**g**)= Sigmoid function
- No. of tries = Until the sqrt of sum of squares of (**d**-**z**) is less than 0.01 for all weights.
- Learning Rate set to 1.5 (since (**d**-**z**) per sample wit sigmoid is much smaller than using step function)

Step 2: Results after iterations

Weight values obtained:

W_bias = -37.5157

W_red = -38.4309

W_green = -2.9554

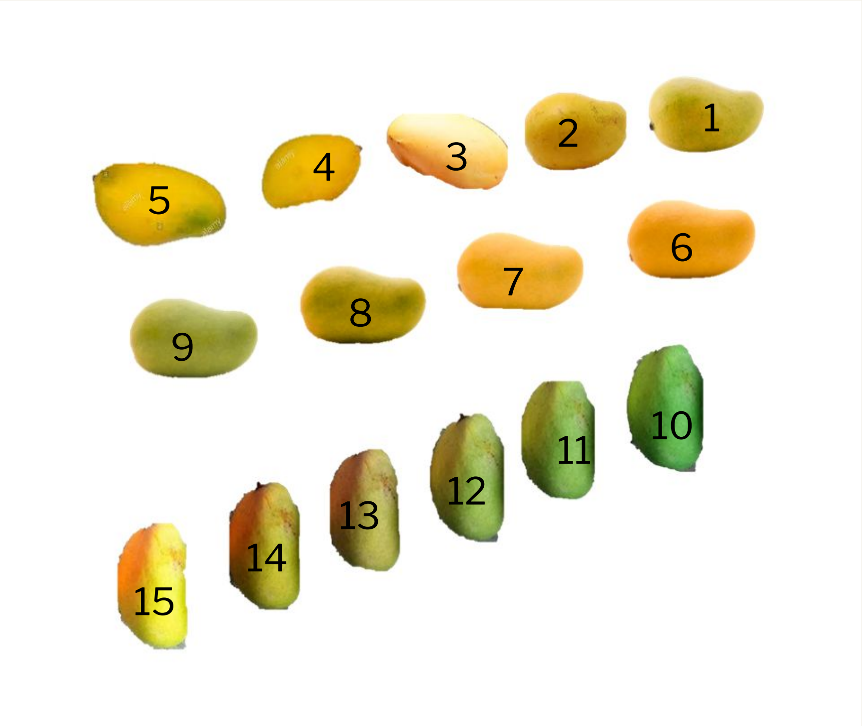
W_blue = 70.163

- The most prominent feature was the Blue pixel value has the highest contribution. The higher the blue value the less amount of blue the pixel has (blue -> white) and vice versa.. It makes sense since green has more blue than yellow (green -> yellow).

- The next feature with high contribution is the red pixel value. It makes sense since green has less red than yellow.

Step 2: Test Data Results

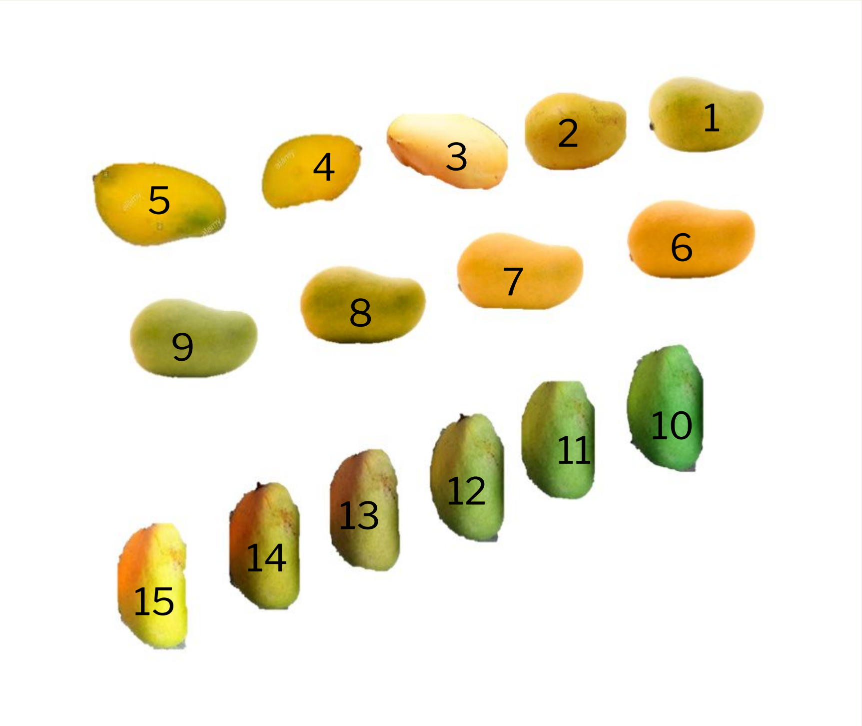
The table shows the mangoes which are color coded qualitatively by its class (yellow - ripe, green - unripe, l.green - between) The results show that the algorithm does not see the fruits that are in the middle of ripeness since the slope of the sigmoid function is steepest at 0.5. Mangoes that are dominantly yellow are regarded as certainly ripe by the program while those dominantly green are classified as certainly unripe.



Mango No.	Z-value	Mango No.	Z-value
1	0.997772	9 (Bitter)	0.00777046
2	0.999979	10 (0-5%)	8.70086e-11
3	0.999976	11 (5-10%)	1.95621e-07
4	1	12 (10-25%)	6.92874e-06
5	1	13 (25-50%)	0.000720742
6 (V Sweet)	1	14 (50-80%)	0.0876307
7 (Sweet)	1	15 (80-100%)	0.999999
8 (Tart)	0.986389		

Step 2: Test Data Results

The 14th mango had a yellow-green color (slightly greener) with a 9% chance of being ripe while the 8th mango (slightly yellow) has a 98.6% chance of it being ripe. For future works, adding more samples to the training data and having an expert that can accurately tell which fruits are ripe or unripe much better than me for comparison with the program's results.



Mango No.	Z-value	Mango No.	Z-value
1	0.997772	9 (Bitter)	0.00777046
2	0.999979	10 (0-5%)	8.70086e-11
3	0.999976	11 (5-10%)	1.95621e-07
4	1	12 (10-25%)	6.92874e-06
5	1	13 (25-50%)	0.000720742
6 (V Sweet)	1	14 (50-80%)	0.0876307
7 (Sweet)	1	15 (80-100%)	0.999999
8 (Tart)	0.986389		