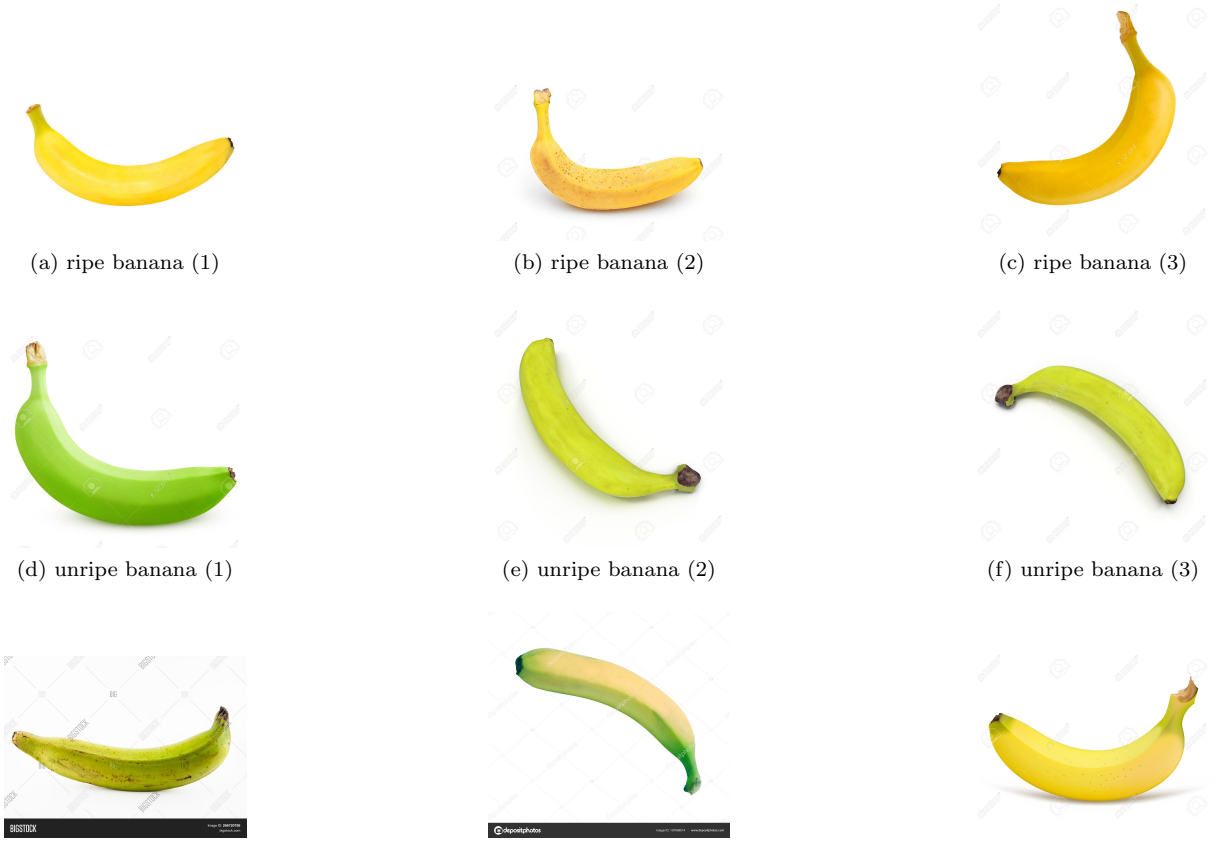
Activity 14 Logistic Regression

Kyngzer-Rem Vargas*

National Institute of Physics, University of the Philippines Diliman, Quezon City, 1101
*Corresponding author: kmvargas2@up.edu.ph

Objective(s)

1. Obtain a data set of fruit that changes color as it ripens.
2. Plot fruit feature data points into the feature space.
3. Apply logistic regression to train an artificial neuron to give the degree of ripeness of that the fruit. Plot the Test with images not yet seen by the neuron and comment if the output agrees with the visual appearance [1, 2].



(a) ripe banana (1) (b) ripe banana (2) (c) ripe banana (3)

(d) unripe banana (1) (e) unripe banana (2) (f) unripe banana (3)

(g) unknown (1) (h) unknown (2) (i) unknown (3)

Figure 1: Sample of the test Images used [3]

Results

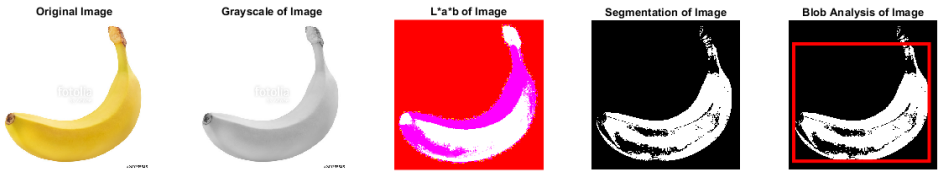
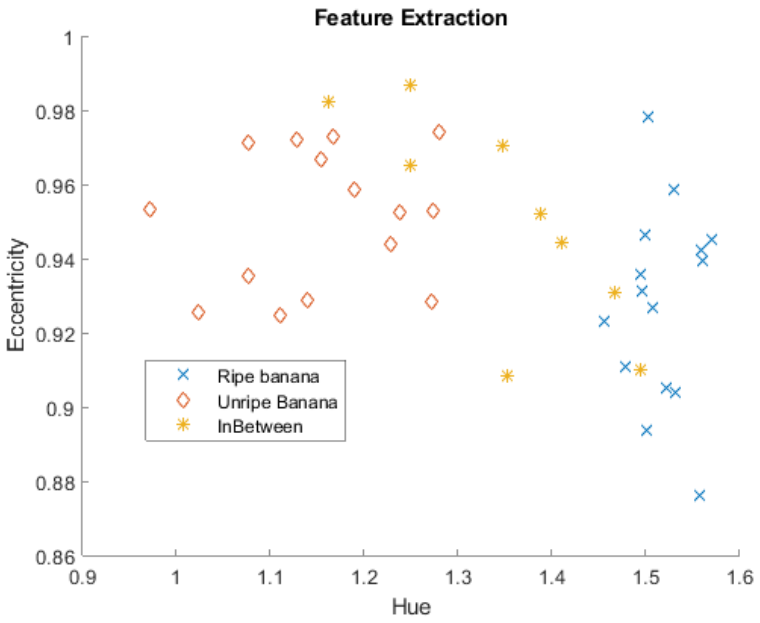
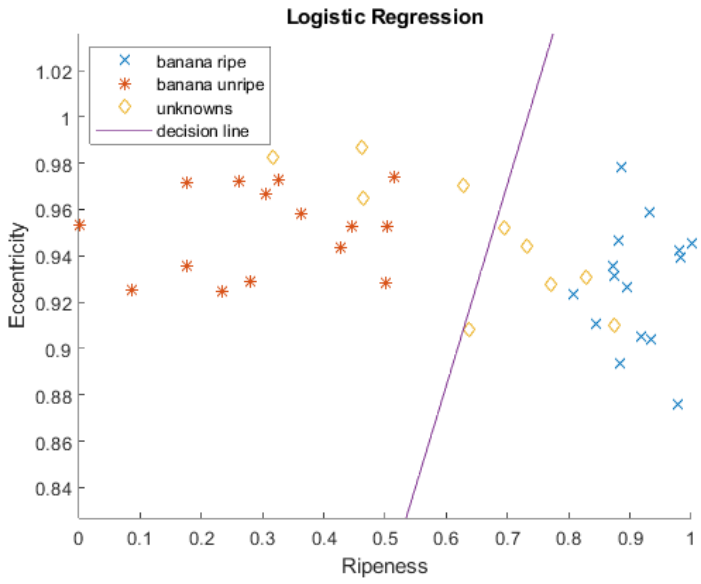


Figure 2: Sample of post processing and feature extraction of data. [4]



(a) Feature extraction



(b) Logistic regression

Figure 3: Decision from fruit feature data.

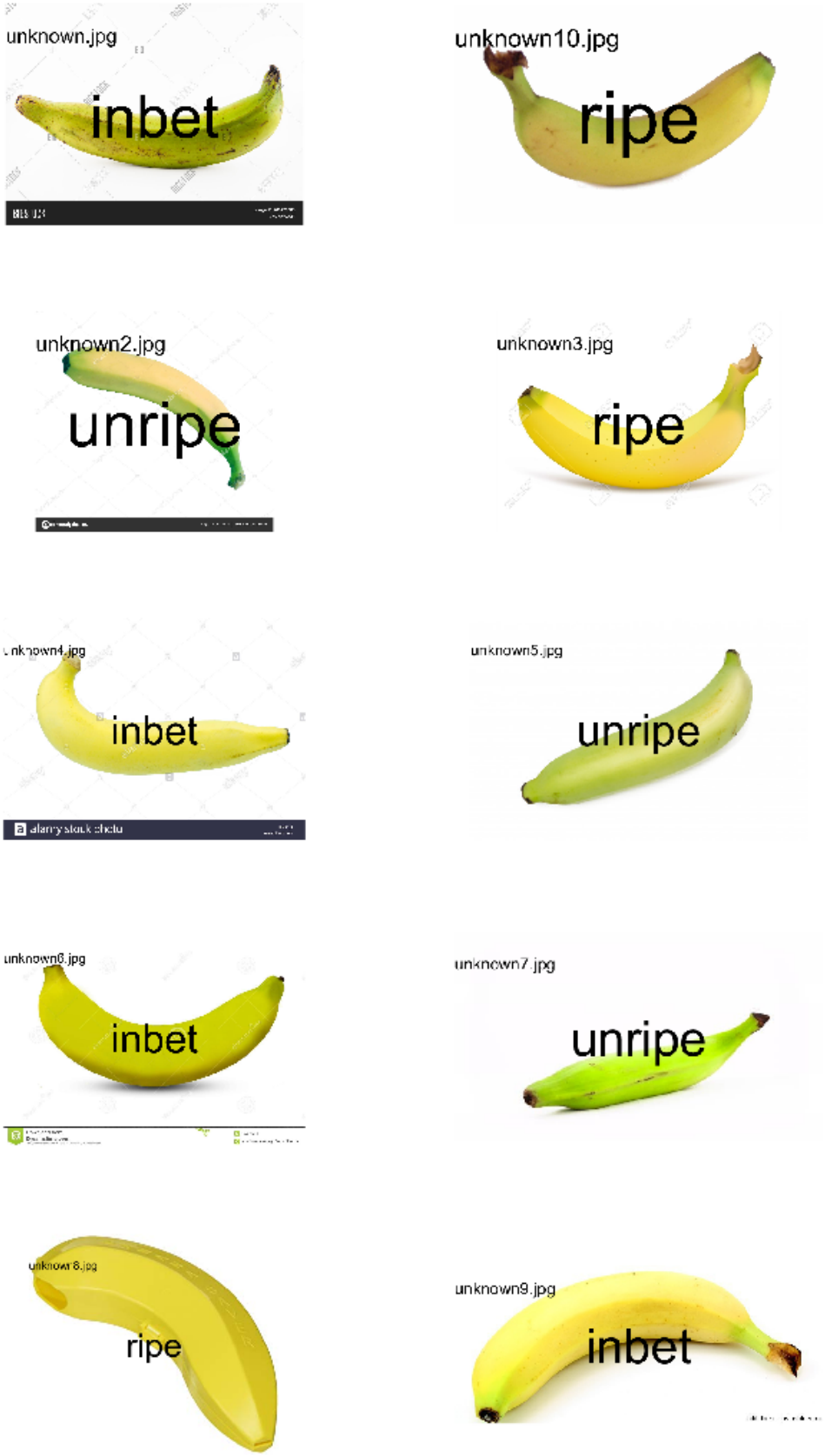


Figure 4: Logistic regression applied to images not yet seen by the neuron.

Table 1: Probability values from the images not yet seen by the neuron.

Image	Probability	Prediction
1	0.20	'inbet'
2	0.97	'ripe'
3	0.01	'unripe'
4	0.99	'ripe'
5	0.58	'inbet'
6	0.01	'unripe'
7	0.54	'inbet'
8	0.00	'unripe'
9	0.91	'ripe'
10	0.77	'inbet'

Comment(s)

Post Processing. Figure 2 shows a brief preview of the post processing that occurs for one image. The main tools used were regionprops and non-parametric segmentation which was used for the blob analysis. The L*a*b* values of the image were also obtained to determine its hue for an accurate measure of the object's color which was used to determine the fruit's ripeness. The feature data points were obtained from post processing such that the eccentricity and hue of the object can be extracted from the image. These data points were then used for the perceptron algorithm with logistic regression as the activation function.

Decision Line. Figure 3 shows the feature space and the respective feature data points obtained from the image data sets. The decision line was obtained by adjusting each weight for each class feature and it was calibrated such that the labels assigned to each class can be predicted perfectly. The error was minimized such that it is below 0.01 and then the weights are then converted into an alternative form of the equation of the decision line.

Logistic Regression. The probability values from the logistic regression was used in order to determine if the banana was ripe, unripe, or in-between. In correspondence to the ripeness values, if the probability is close to zero then it is unripe, if the probability is close to one then it is ripe, and anything in between will be in between. Figure 4 shows if the probability values in Table 1 are correct with respect the fruit's visual appearance and it can be seen that the predictions are correct.

Self-Evaluation

I would rate myself a 10. The objectives for this activity was met. The image data sets were properly processed such the each of the image's features are extracted via regionprops and hue values from its L*a*b values. The decision line was successfully plotted onto the feature space such that it gave a clear division between each class found on the feature space. Things I would like to improve on the activity is the volume of the images of the untrained data set to see how good the classification performs and improve on it.

References

[1] M. Soriano, A14-logistic regression.pdf.

[2] M. Soriano, Machine learning intro.pdf.

[3] G. Search, Google images (2019), last accessed 07 November 2019, https://www.google.com/search?q=images.

[4] TheMathWorksInc., regionprops (2006), last accessed 07 November 2019, https://www.mathworks.com/help/images/ref/regionprops.html.