Logistic Regression

Kenneth M. Leo

Score: 10/10

What I used

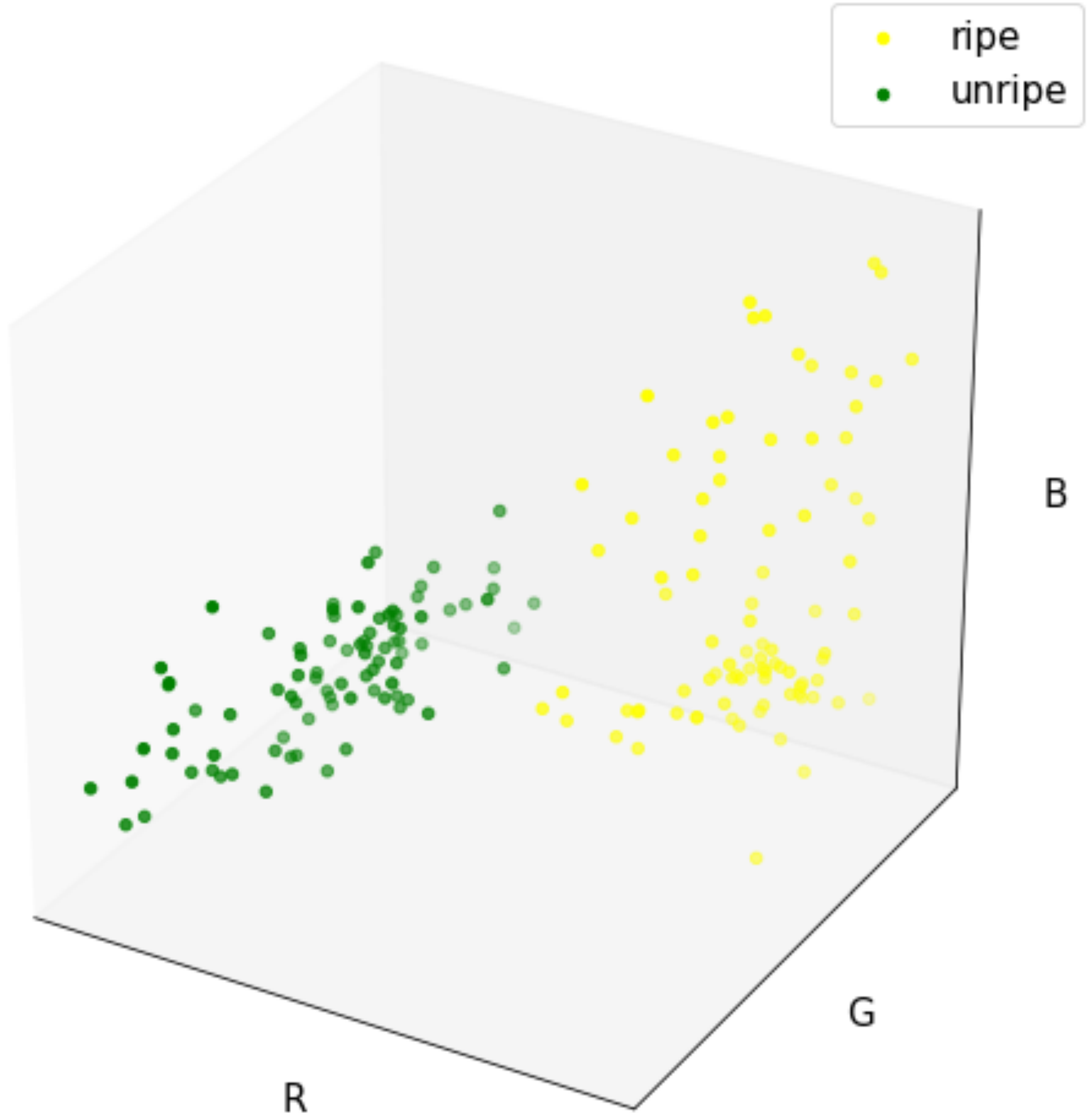
For this activity, I used ripe and unripe bananas. My dataset was from a paper titled "*Ripeness Classification of Bananas Using an Artificial Neural Network.*" It contains photos of ripe, unripe, and mid-ripe bananas. Here are sample photos.

Unripe banana

Ripe banana

What I did

First, I extracted the color features (R,G,B) of each banana and stored it in a dataframe. I also used 0-1 values for each color instead of the usual 0-255. The figure on the right shows 160 data points (used as the training dataset) plotted, 80 ripe and 80 unripe bananas. We see that unripe and unripe bananas cluster together which means it is valid to use the RGB colors as a feature of the fruit.



Logistic Regression

Instead of binary classifying the fruit as either ripe or unripe, I used the logistic function (shown below) as the activation function to give the probability of how ripe the fruit is (1 = ripe, 0 = unripe).

$$\phi(a) = \frac{1}{1 + e^{-a}} = probability$$

I trained the dataset to get the weights of each feature.

$$\mathbf{x} = [x_0, x_{red}, x_{green}, x_{blue}] \qquad \mathbf{w} = [w_0, w_{red}, w_{green}, w_{blue}]$$

The weight of will then be multiplied to the transpose of the feature values (x^T) of a test fruit to find a and then plug a to the activation function to find the probability.

Results and Discussion

The figures and numbers below the test bananas used and their corresponding ripeness probabilities. I made sure that these photos were not used for my training dataset.

probability: 0.0259 or 2.59%

probability: 0.6602 or 66.02%

probability: 0.9785 or 97.85%

We see that my dataset was able to somehow correctly predict the ripeness of the bananas. The green banana have very low ripeness probability while the yellow one have a high ripeness probability.