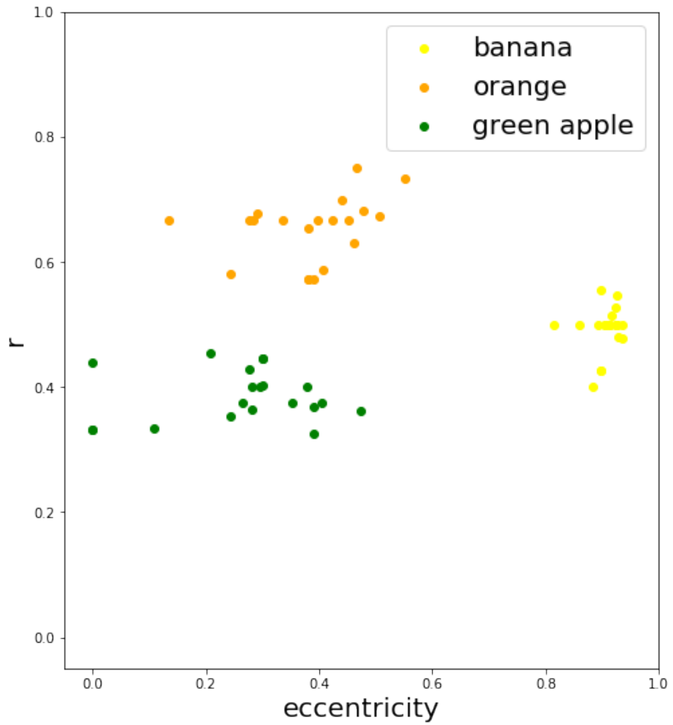
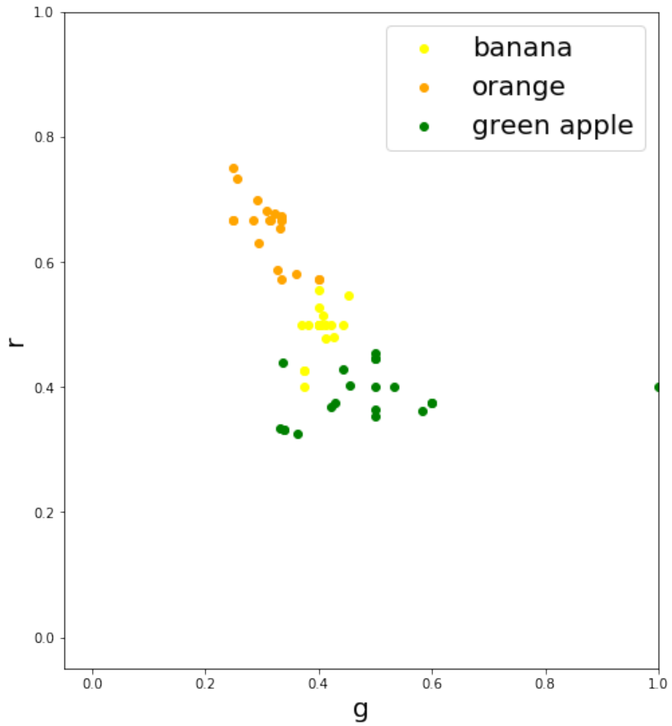
Perceptron

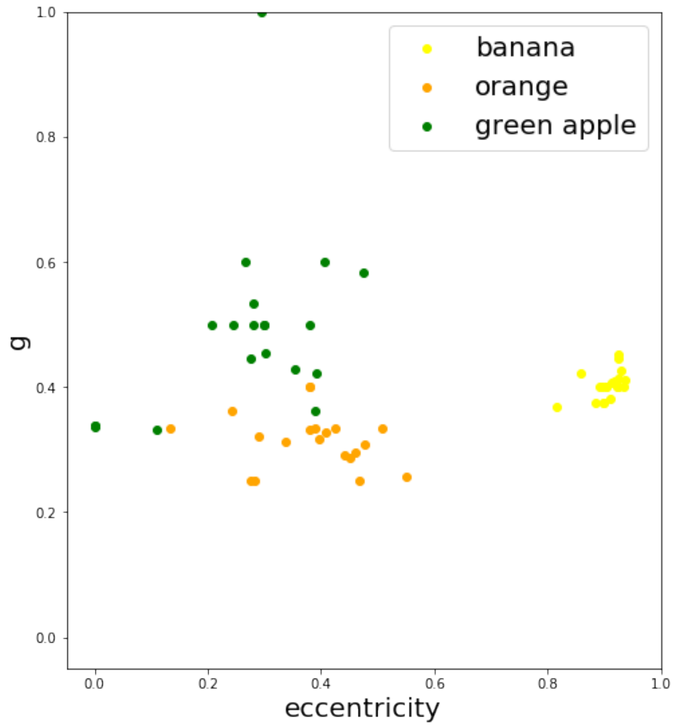
Kenneth M. Leo

Score: 11/10 (I did 3 features and applied perceptron to all three of them

Previously…







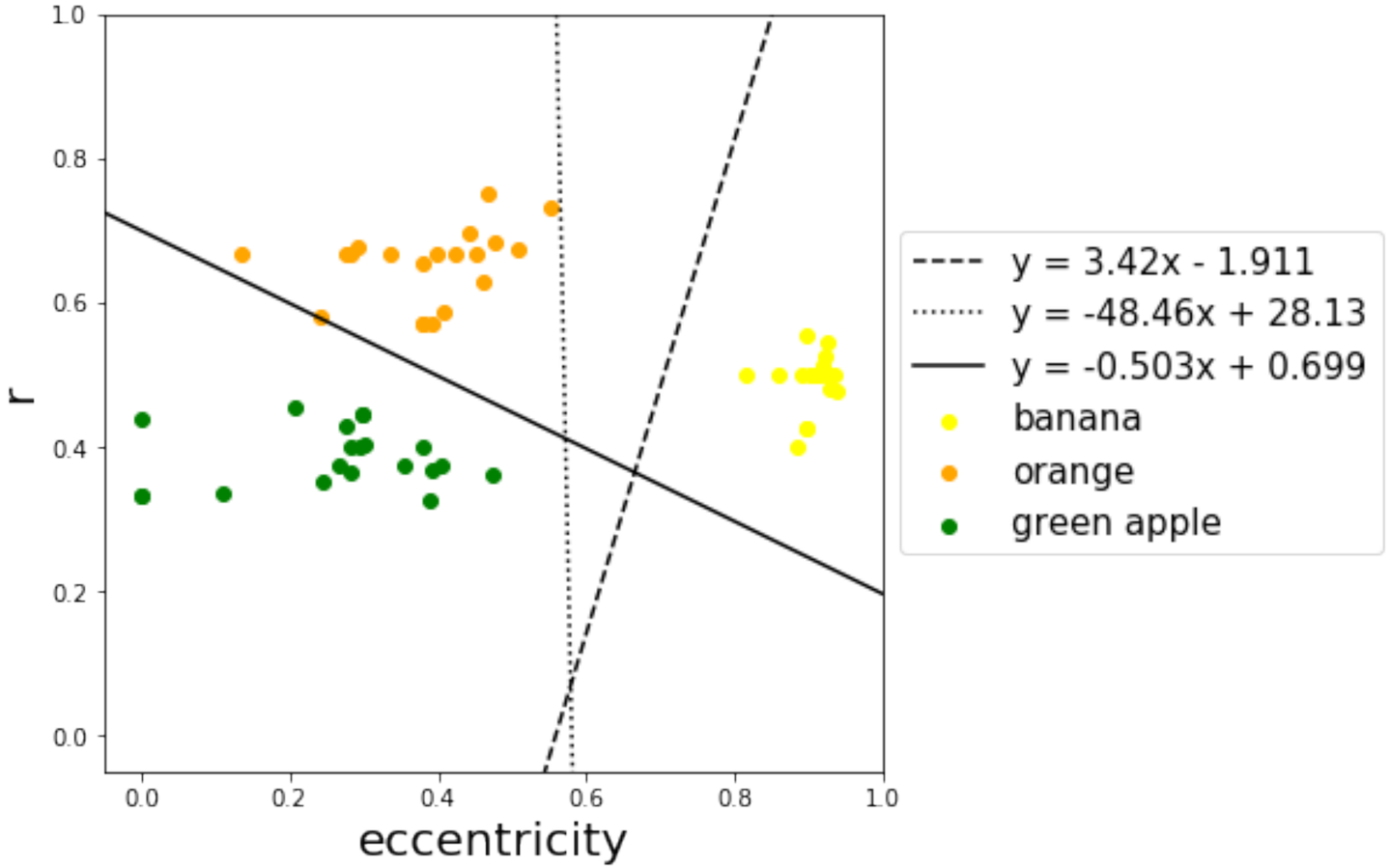
From the previous activity, these are the features and the classes I used. We can see that obvious clustering occurs in the eccentricity vs. r plot but it is not that obvious in the other two plots.

Perceptron Algorithm

The snapshot below shows how I implemented the perceptron algorithm. The X and D arrays are the lists of features, and their corresponding class labels.

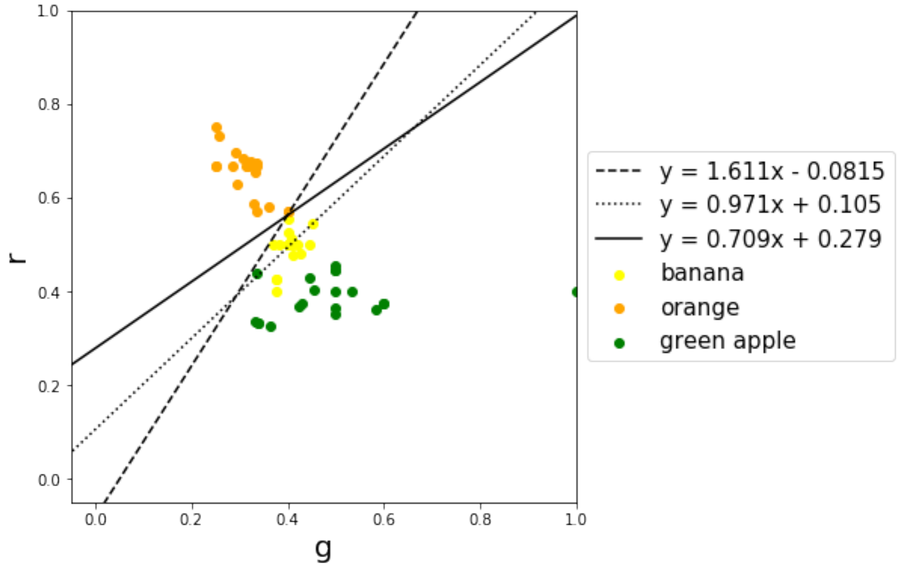
```
def perceptron(X,D):
    w = np.zeros(3)
    w[0],w[1],w[2] = random.random(), random.random(),random.random()

    eta = random.random()

    time = 1000 #can be longer
    for t in range(time):
        for i,x in enumerate(X):
            a = np.dot(x.T,w)
            z = g(a)
            delta_w = eta*(D[i] - z)*x
            w += delta_w
    return w

def line_parameters(w):
    A, B, C = w[1], w[2], -w[0]
    m = -A/B
    b = C/B
    return m, b
```

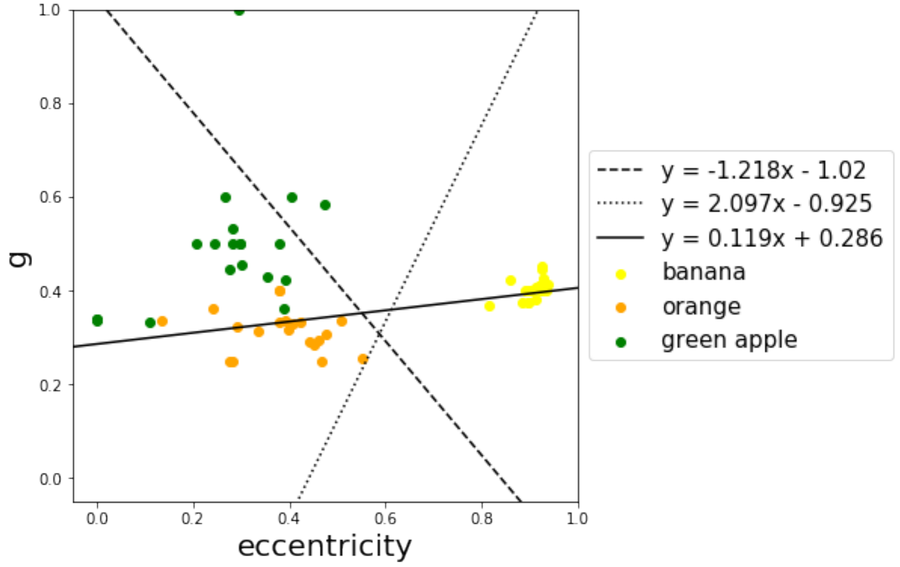
Results



We can see that we were able to get 3 line equations that separate the three classes. The dashed line separates the banana and orange, the dotted line separates the banana and green apple while the solid line separates the orange and green apple. We can see that there are three distinct regions formed by the three lines. For example if an unknown data is plotted to be in the region where the banana class is located, we can say that the unknown data is a banana.

Results





For these two plots, no clear distinctions were made. For the r vs. g plot, we see that each line separates the two classes that they have to separate (line styles similar to the previous slide). But looking at it overall, we cannot see three regions that show where each classes belong. Because of this we cannot completely identify an unknown data as to where it will belong if plotted in the rg plot. On the other hand, the eccentricity vs. g plot shows that it had problems creating the line separating the orange and green apple. This is because the two classes appeared to ‘mix’. But since the banana is obviously distinct compared to the other two, the regions formed are better compared to the rg plot regions.

Conclusion

Based on my observations, using the eccentricity values and red ncc of a fruit is the best way to categorize them. This is because these features had the best ‘region segmentation’ when perceptron algorithm is used.