Jake Stephens

Assignment #6

August 17, 2021

**Question #1:**

1. SVM with Linear Kernel

Accuracy: 1.0

Confusion Matrix:

[[35 0]

[ 0 35]]

1. SVM with Gaussian Kernel

Accuracy: 1.0

Confusion Matrix:

[[35 0]

[ 0 35]]

1. SVM with Polynomial Kernel

Accuracy: 1.0

Confusion Matrix:

[[35 0]

[ 0 35]]

**Question #2:**

1. KNN Classifier

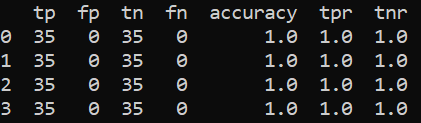
Accuracy: 1.0

Confusion Matrix:

[[35 0]

[ 0 35]]

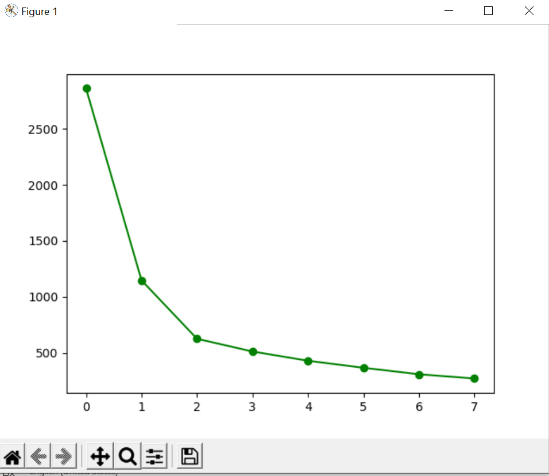




The SVM and KNN classifiers seems to be very accurate for predicting the class, as all methods of classification were 100% accurate.

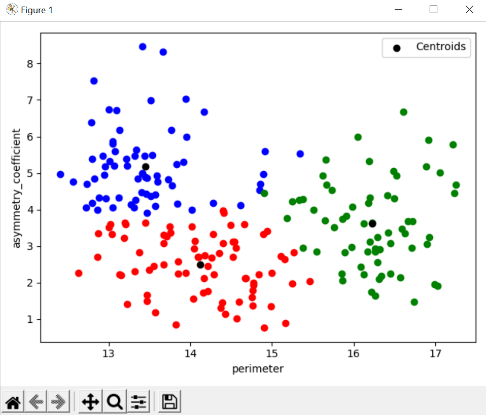
**Question #3:**





I chose the best k to be 3. Since after that the slope of the line planes off and decreases about the same amount for every added cluster.





I picked at random, asymmetry\_coefficient and perimeter, and I can see the clustering looks as you would expect in the graph. However, I learned later that these 2 columns are not at all the columns you want to look at for determining the class. It is interesting that the blue cluster seems to be pretty dense around the centroid, but the red and green clusters seem to be a bit more spread out.

1. For cluster\_1\_red the most prominent class is class 1

For cluster\_2\_blue the most prominent class is class 3

For cluster\_3\_green the most prominent class is class 2

1. The accuracy of my new classifier is awful, it is under 50% for all the runs I have done. The average seems to be around 30%. I believe this is due to the fact that the 2 columns chosen at random have little to no correlation with the class column, and it is very hard to determine what class a data point belongs to just from these 2 columns. Had I chosen 2 different columns at random, such as length\_of\_kernel and width\_of\_kernel, these 2 features may have been able to give me better results at finding the correct class.
2. With only looking at classes 2 and 3, my classifier seems to be a little better, being that its around 40% accurate. This tells me that the 2 random columns chosen, asymmetry\_coefficient and perimeter, may not have more of a correlation with classes 2 and 3 as they do with class 1, this is because it is closer to a 50-50 chance of getting the correct class. Whereas when I was trying to classify with all 3 classes it was around 30% or 1/3 chance. This proves to me further that these 2 columns have little to do with the class assignment.