

Marisa Paone  
MET CS677  
8/10/2023  
Homework #6

Question 1. 50/50 spits

**Please see main.py**

1. *What is your accuracy and confusion matrix for linear kernel SVM?*

**My accuracy is 94.3%**

**Confusion matrix:**

**[[35 0]**

**[4 31]]**

2. *What is your accuracy and confusion matrix for Gaussian kernel SVM?*

**My accuracy is 98.6%**

**Confusion matrix:**

**[[35 0]**

**[1 34]]**

3. *What is your accuracy and confusion matrix for polynomial kernel SVM (deg 3)?*

**My accuracy is 97.14%**

**Confusion matrix:**

**[[34 1]**

**[1 34]]**

Question 2.

**Please see main.py**

1. *What is your accuracy and confusion matrix for your classifier?*

**My classifier was Naive Bayesian and my accuracy is 92.86%**

**Confusion matrix:**

**[[33 2]**

**[3 32]]**

2. *Discuss your results:*

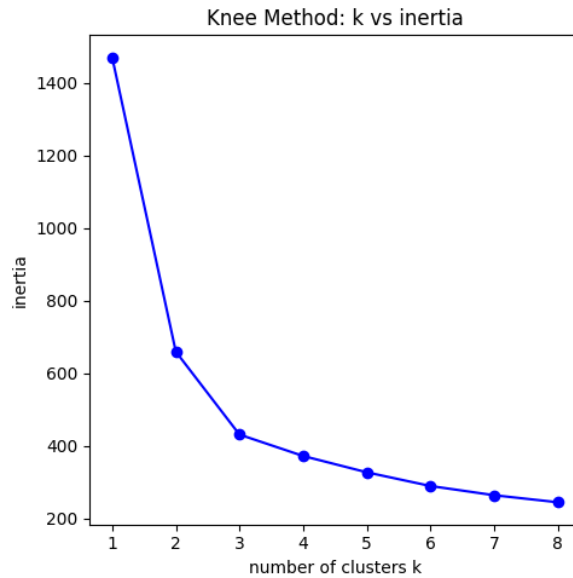
**Gaussian SVM was the most accurate at 98.6% followed by Polynomial (deg 3) SVM at 97.14% (The best TPR and TNR follows this as well). The least accurate was my classifier (naive bayesian). However I don't think that enough rows were in the training set (70 rows is not much).**

<u>Model</u>	<u>TP</u>	<u>FP</u>	<u>TN</u>	<u>FN</u>	<u>Accuracy</u>	<u>TPR</u>	<u>TNR</u>
Linear SVM	35	0	31	4	94.3%	0.897	1
Guassian SVM	35	0	34	1	98.6%	0.972	1
Polynomial SVM	34	1	34	1	97.14%	0.971	0.971
Naïve Bayesian (My classifier)	33	2	32	3	92.86%	0.917	0.941

Question 3.

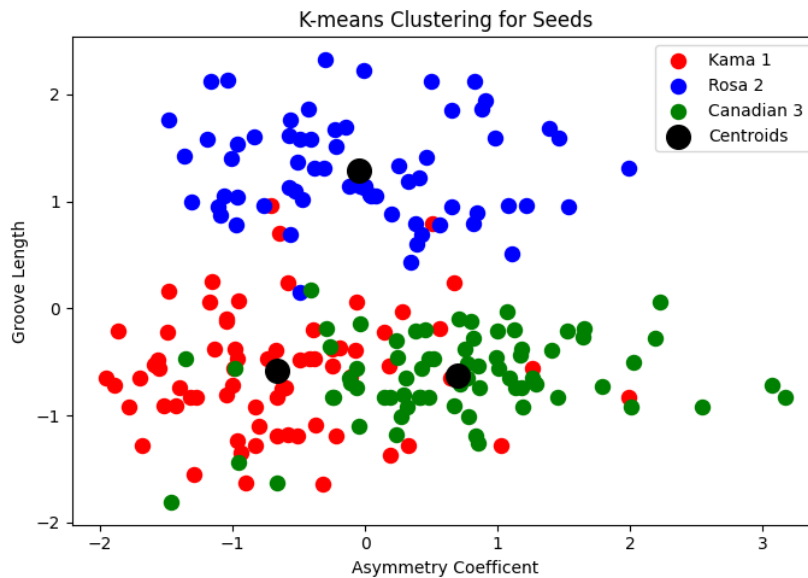
**Please see main.py**

1. Compute and plot distortion vs.  $k$ . Find the best  $k$ .



**The best  $k$  was 3 (the slope started to become less dramatic here).**

2. rerun your clustering with the best  $k$  clusters. Are there any interesting patterns?



Each time running the program there are different features selected. This time I ran it, the asymmetry coefficient and the groove length were the mapped features. As far as patterns go, the kama, rosa and Canadian seeds are separated into clusters for the most part. You can easily differentiate one from the other. There are some kama seeds that are spilling into the canadian cluster and there are some canadian seeds spilling into the kama cluster. There are also some

points from the kama cluster close to the Rosa seed but these are very few. An interesting pattern that I saw was the centroids make an isosceles triangle (or at least close to it).

3. For each cluster, print out its centroid and assigned label:

**For the two features here are the centroids and assigned labels:**

Kama Centroid: [-0.6619186699835916, -0.585893107814457]

Rosa Centroid: [-0.04521936049468102, 1.2923078690131353]

Canadian Centroid: [0.6948044822497021, -0.6248085634590778]

**For all features here are the centroids and their assigned labels:**

Kama Centroid: [-0.14111949 -0.17004259 0.4496064 -0.25781445 0.00164694 -0.66191867 -0.58589311]

Rosa Centroid: [ 1.25668163 1.26196622 0.56046437 1.23788278 1.16485187 -0.04521936 1.29230787]

Canadian Centroid: [-1.03025257 -1.00664879 -0.9649051 -0.89768501 -1.08558344 0.69480448 -0.62480856]

4. What is the overall accuracy of this new classifier when applied to the complete dataset?

**The accuracy of KNN using k = 1 for the two random features was 85.23%.**

**The accuracy of KNN using k = 1 for all seven features was 91.90%.**

5. What is your accuracy and confusion matrix using the same two labels you used for SVM? How does your classifier from task 4 compare with any classifiers listed in the table for question 2 above?

**My accuracy of KNN using k = 1 for just class 1 Kama and class 3 Canadian for just the two features was 80.71%.**

**Confusion matrix:**

**[[58 12]**

**[15 55]]**

**My accuracy of KNN using k = 1 for the two classes and all seven features was 84.29%.**

**Confusion matrix:**

**[[69 1]**

**[21 49]]**

**My classifier from task 4 (85.23%) is the most similar to the naive bayesian classifier that had an accuracy of 92.86%. This KNN classifier is the least accurate out of the ones in my table. My classifier had even lower accuracy when I took out class 2 Rosa (as shown above ~ 80.71%).**