IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – VII
Clustering

Student's Name: **Mobile No:**

Roll Number: **Branch:**

1

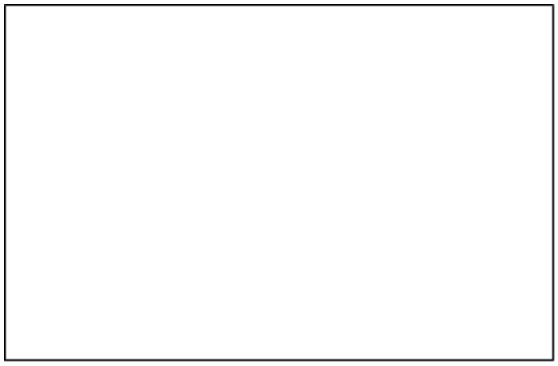

Figure 1 Eigenvalue vs. components

Inferences:

1. Does the eigenvalue increase or decrease corresponding to each component increase or decrease successively?
2. Justify the observed trend.
   Note: The plot above is for illustration purposes. Replace it with the plot obtained by you. Label x-axis as components and y-axis as Eigenvalues.

2 a.

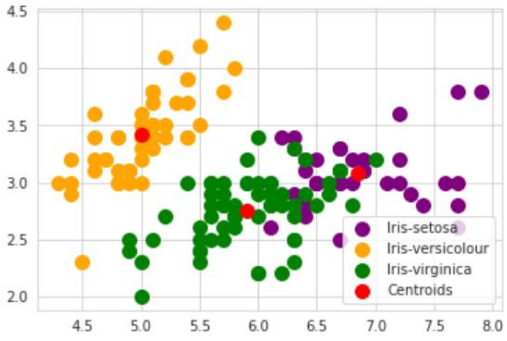

Figure 2 K-means (K=3) clustering on Iris flower dataset

Inferences:

1. Inferring from the clusters formed in the above plot, comment on the clustering prowess of the algorithm.
2. K-means algorithm assumes cluster boundaries to be circular in 2D. From the output, does the boundary seem to be circular?
   Note: The plot above is for illustration purposes. Replace it with the plot obtained by you.

b. The value for distortion measure is

c. The purity score after examples are assigned to the clusters is

3

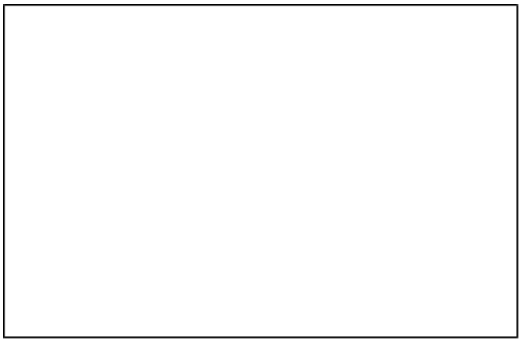

Figure 3 Number of clusters(K) vs. distortion measure

Inferences:

1. Does the distortion measure increase or decreases with an increase in K?
2. Justify the observed trend.
3. From the number of species in the given dataset, intuitively what should be the number of optimum clusters? Does the elbow and distortion measure plot follow the intuition?
   Note: The plot above is for illustration purposes. Replace it with the plot obtained by you. Label x-axis as distortion measure and y-axis as number of clusters (K).

Table 1 Purity score for K value = 2,3,4,5,6 & 7

K value	Purity score

Inferences:

1. The highest purity score is obtained with K =.
2. Infer whether increasing the value of K increases/decreases the purity score.
3. State a suitable reason why increasing the value of K increases/decreases the purity score.
4. Is there any observable relationship between purity score and distortion measure?

4 a.

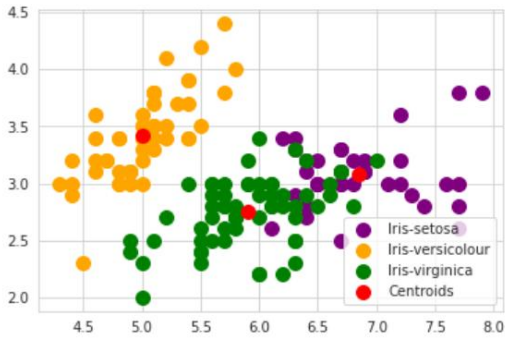

Figure 4 GMM (K=3) clustering on Iris flower dataset

Inferences:

1. Inferring from the clusters formed in the above plot, comment on the clustering prowess of the algorithm.
2. GMM algorithm assumes cluster boundaries to be elliptical in 2D. From the output, does the boundary seem to be circular?
3. Is there any observable difference between clusters formed using K-means in 2.a and GMM in 4.a?
   Note: The plot above is for illustration purposes. Replace it with the plot obtained by you.

b. The value for distortion measure is

c. The purity score after examples are assigned to the clusters is

5

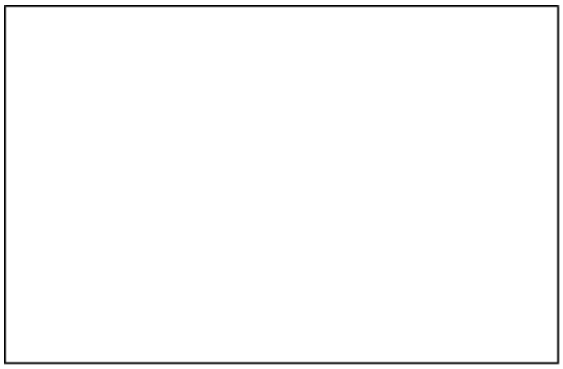

Figure 5 Number of clusters(K) vs. distortion measure

Inferences:

1. Does the distortion measure increase or decreases with an increase in K?
2. Justify the observed trend.
3. From the number of species in the given dataset, intuitively what should be the number of optimum clusters? Does the elbow and distortion measure plot follow the intuition?
   Note: The plot above is for illustration purposes. Replace it with the plot obtained by you. Label x-axis as distortion measure and y-axis as number of clusters (K).

Table 2 Purity score for K value = 2,3,4,5,6 & 7

K value	Purity score

Inferences:

1. The highest purity score is obtained with K =.
2. Infer whether increasing the value of K increases/decreases the purity score.
3. State a suitable reason why increasing the value of K increases/decreases the purity score.
4. Is there any observable relationship between purity score and distortion measure?
5. Compare K-means and GMM based on inferences in Q3 and Q5.

6

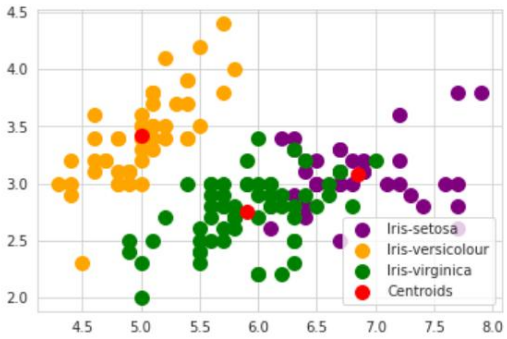

Figure 6 DBSCAN clustering on Iris flower dataset

Inferences:

1. Inferring from the clusters formed in the above plot, comment on the clustering prowess of the algorithm.
2. Is there any observable difference between clusters formed using K-means in 2.a, GMM in 4.a and DBSCAN in 6.a?
   Note: The plot above is for illustration purposes. Replace it with the plot obtained by you. 1 plot has been given, plot for each combination.

b.

Eps	Min_samples	Purity Score
1	5	
	10	
4	5	
	10	

Inferences:

1. For the same eps value, does increasing min_samples increase or decrease purity score?
2. For the same min_samples, does increasing eps value increase or decrease purity score?

Guidelines for Report (Delete this while you submit the report):

- **The plot/graph/figure/table should be centre justified with sequence number and caption.**
- **Inferences should be written as a numbered list.**

- **Use specific and technical terms to write inferences.**
- **Values observed/calculated should be rounded off to three decimal places.**
- **The quantities which have units should be written with units.**