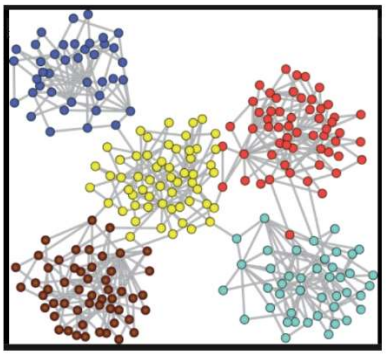
Arkansas Tech University
Graduate College, Information Technology
INFT 6903 Data Science, Summer 2022
Quiz 7 / Due date: July 19th, 2022 before midnight. Total points: 100. Good luck!

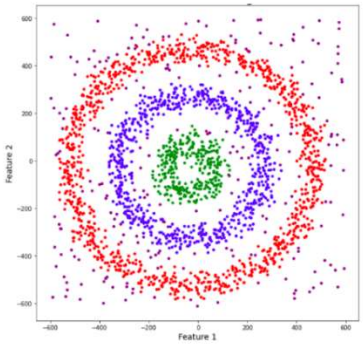
---

1. (*10 points*) The following figure shows the clustering performance of the social network/media data. How many cluster can we use for this result for the best scenario?

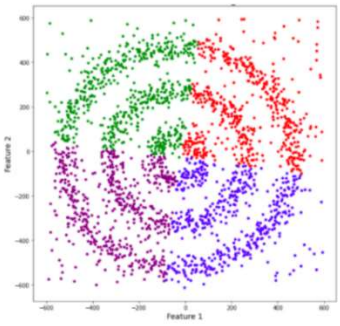


A) 3 B) 5 C) 4 D) 2

2. (*10 points*) Following figures show the clustering results of the Clustering algorithm 1 and the Clustering algorithm 2. Please match the algorithms with the figures.



Clustering algorithm 1

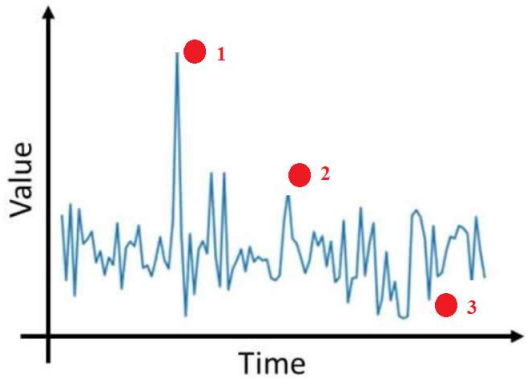


Clustering algorithm 2

	Clustering algorithm 1	Clustering algorithm 2
A)	DBSCAN	kmeans
B)	kmeans	DBSCAN

3. (*10 points*) TRUE or FALSE → All clustering algorithms show the same performance in data science.

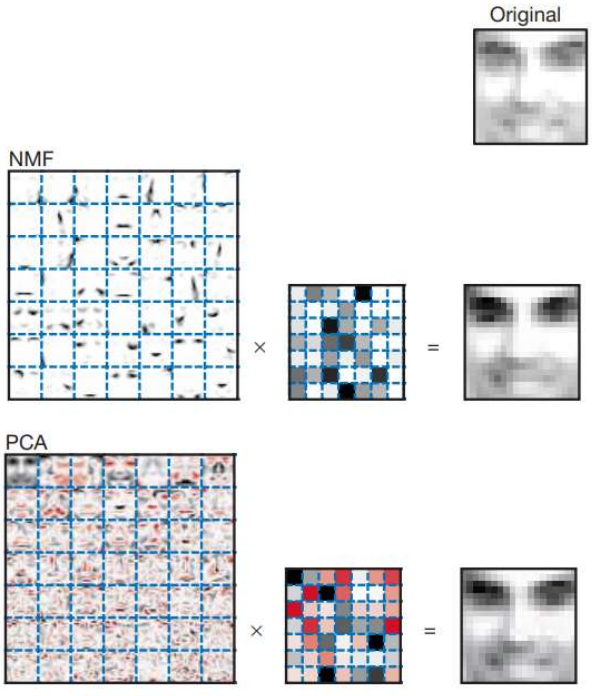
4. (*10 points*) Which one of the following red points can be defined anomaly (outlier) for the following analog signal?



A) 1 B) 2 C) 3

5. (*10 points*) Please see the following figure that has nonnegative matrix factorization (NMF) and principal component analysis (PCA) feature & coefficient matrices (From D. D. Lee, H. S. Seung, Learning the parts of objects by nonnegative matrix factorization, Nature, 401, pp. 788-791, 1999).

As we can see from the figure, non-negative matrix factorization (NMF) learns a parts-based representation of faces, whereas principal components analysis (PCA) learns holistic representations. Please use this result and make a comment in terms of shapes of the feature matrices of NMF and PCA in one or two sentences (*Hint*: What do these shapes look like?)
(*Note*: Feature matrices are two big matrices below for NMF and PCA. You can also see their labels as NMF and PCA on them).



6. (*12 points*) Suppose that our text mining algorithm detects and recognizes several words from any text documents like websites, chatbots, text documents. etc. Which one of the following text mining algorithms/applications/conversations can be considered the best trained&tested algorithm?

Algorithm 1:

Person: What is your favorite food?
Computer: I am an intelligent software and I don't eat any food.

Algorithm 2:

Person: Do you have a coffee table?
Computer: Our company is in Russellville.

Algorithm 3:

Person: What time is it?
Computer: Total cost is $5.

A) Algorithm 3 B) Algorithm 2 C) Algorithm 1

7. (*12 points*) Which one of the following algorithms performs well-trained&tested text conversation (chatbot) with a customer?

Algorithm 1:

Customer (Person): Hllo
Computer : Did you mean Hello?

Algorithm 2:

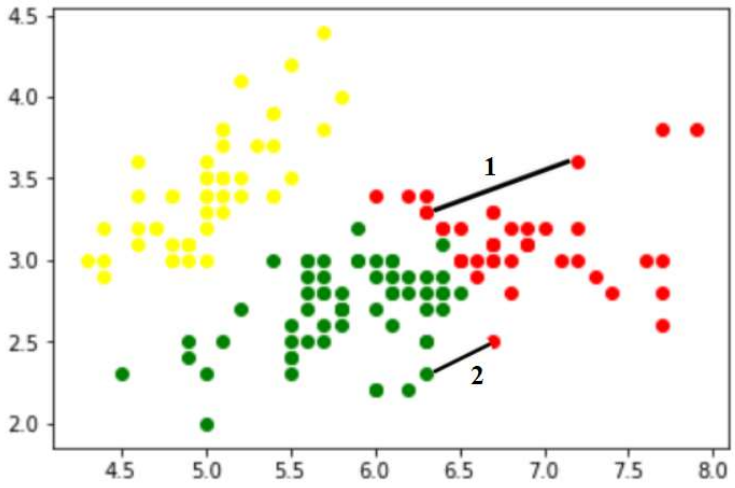
Customer (Person): I want to open a new bank account.
Computer : How much milk do you need to bake a cake?

Algorithm 3:

Customer (Person): What is your working hours for today?
Computer : Can I offer you the cheapest product?

A) Algorithm 1 B) Algorithm 2 C) Algorithm 3

8. (*13 points*) The following graph shows the social network clustering analysis. Please match 'Intra-cluster Distance' and 'Inter-cluster distance'.



A) 1 → Inter-cluster distance and 2 → Intra-cluster distance
B) 1 → Intra-cluster distance and 2 → Inter-cluster distance

9. (*13 points*) Please again see the graph of the Question 8. What is our aim to have a better clustering performance?

A) Minimizing both 1 and 2
B) Maximizing 1 and minimizing 2
C) Maximizing 2 and minimizing 1
D) Maximizing both 1 and 2