



# Machine learning

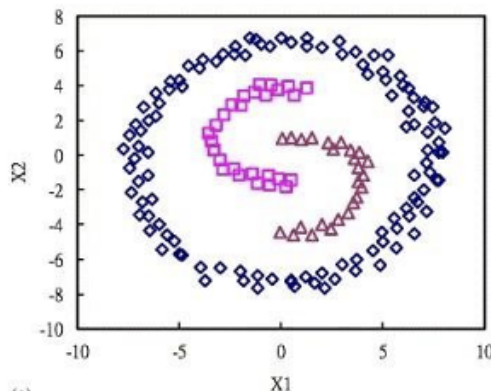
## Fall 2023

### HW4

---

#### PartA - theory questions

1. What does the curse of dimensionality mean in clustering algorithms? What methods exist to address this issue? (Just name one method.)
2. In the k-means clustering algorithm, one of the most important decisions is the choice of the number of clusters, which significantly affects the outcome. Name one algorithm used for this purpose and explain it.
3. Suppose we have an image with 3 channels: green, red, and blue (each pixel of the image is represented by 3 numbers between 0 and 255 for each of these colors). We want to compress this image. How can we use the k-means algorithm for this purpose? Explain what  $k$  represents in your proposed method and how you use the input image.
4. If the data distribution is as follows, which one of the algorithms taught in the class would you choose for clustering them? Explain your reason.



## PartB - practical questions

1. Implement the DBSCAN algorithm.
2. Plot a scatter of the given dataset. Then, use the implemented algorithm to cluster the 2 sets of data available. Finally, display the identified clusters with different colors.