Reading Assignment 2

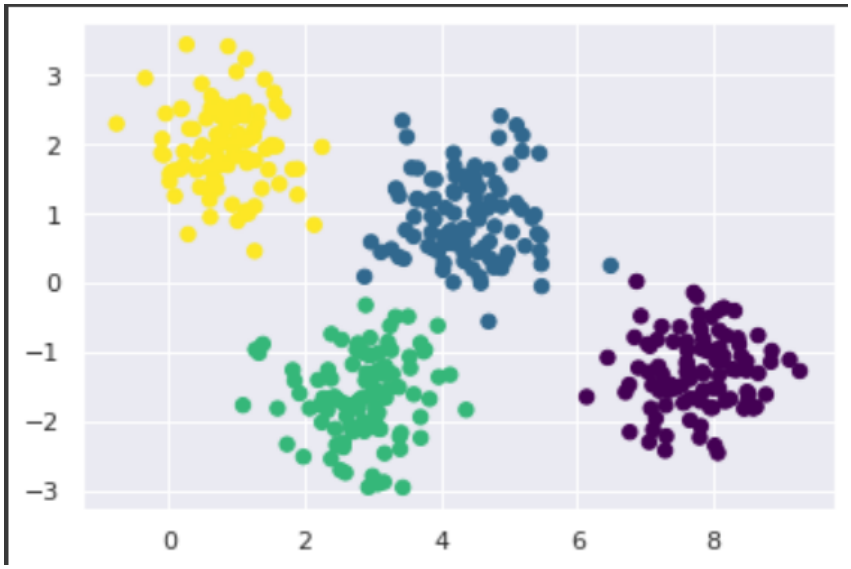
Clustering:

In simple words, clustering is the process which is used segregate and distribute the points into different clusters where each of the points in a particular cluster share the same trait.

Gaussian Mixture Model:

Gaussian Mixture is one of the unsupervised clustering techniques where a probabilistic assignment is taken into consideration while forming Gaussian clusters containing many data points. Each of the cluster are separated by a separate Gaussian distribution.

Although it requires a significant number of data-points to form clustering with better accuracy, it is one of fastest algorithms for mixture models and it won't bias the means towards zero.



Expectation Maximization:

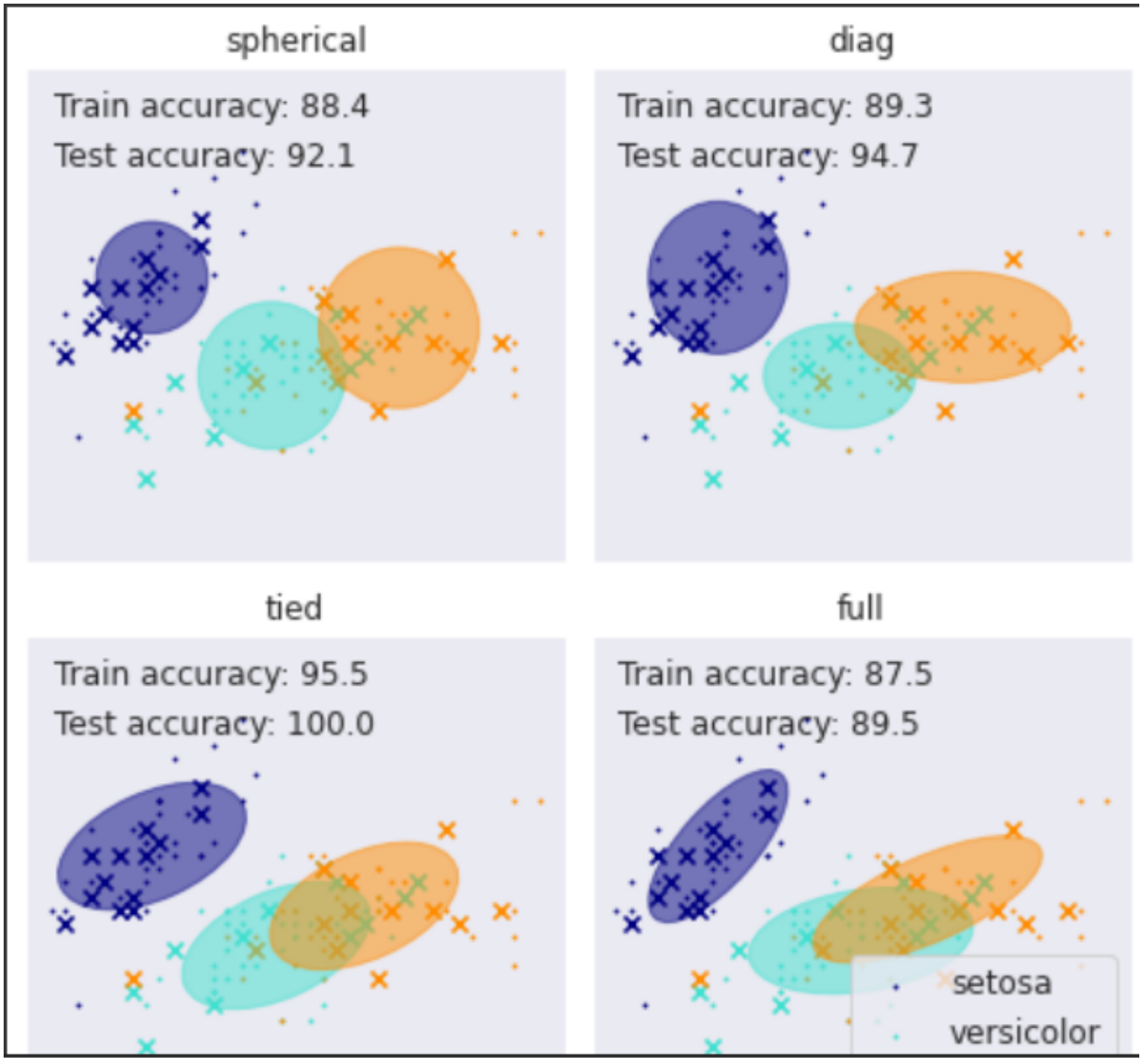
This theory uses an algorithm to determine and estimate maximum likelihood for the latent variables. In the first stage,

1. It performs estimation for the latent variables,
2. Optimize the built model
3. Repeat Steps and 2

Iris Dataset Observations:

For the iris dataset, we have applied 4 different types of Gaussian mixtures. They are:

Spherical(Shape: Circle)
Diag(Shape: Horizontal Ellipse)
Full(Shape:Diagonal Ellipse)
Tied (Shape: Diagonal Ellipse)



After analyzing the outputs of each of the 4 models, we can conclude that in this dataset, 2 clusters out of 3 are overlapped, a lot are tied and full types, while they are the least overlapped in the Spherical ones.

Research Question Answer:

Q: Do social media platforms "care" if they are spreading misinformation?
A: In 2016, Facebook implemented a new feature called “Time-Spent” algorithm, which showed posts on the basis of the total time spent by all the people on that video. If many people watch the whole video , other people would get the video recommendations keeping in mind this factor and it doesn’t matter if the people liked the video or not. But since then, facebook has made several updates trying to negate this belief and improve the whole system.

Similarly, most of the social media websites aim for its own progress to make sure their platform is growing everyday and for that, they will share the content which the users will like to receive , not caring about the truthfulness of the information being shared on their sites. But on the other hand, some websites ,nowadays, are using some policies to mitigate the negative effects of the above problem. In addition, they also ask users to never blindly trust the information being shared on such sites.

Colab for clustering references: colab