GRADING CRITERIA - AML

① Projects / Assignments / Quiz - 20%

② Viva - 30% - Last week of Module

③ End-term Exam - 50% → Module End

PROJECTS → Part-1 → Dim. Red. and Clustering

→ Part-2 → Recommender Systems & Time-Series

↳ CBT / Term-Paper ?

"Clustering using DBSCAN"

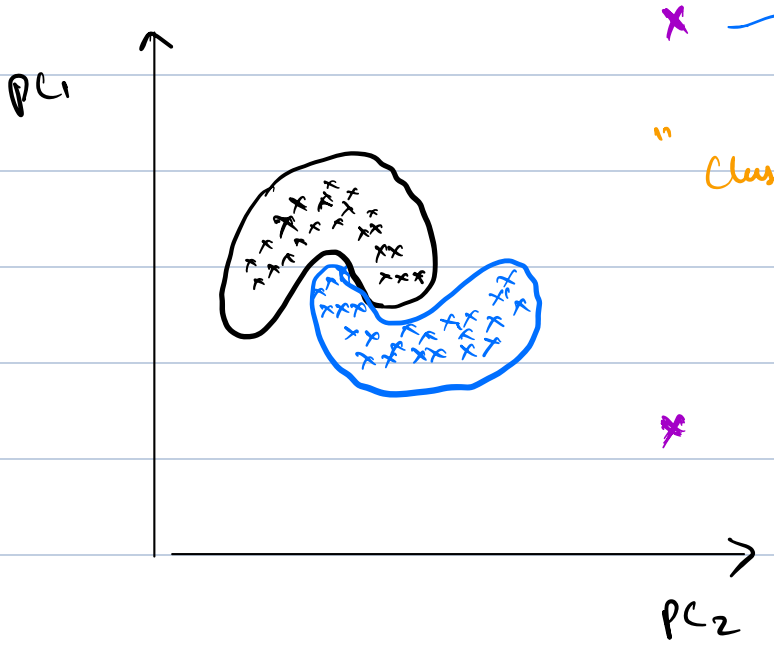
- limitations of GMM:

↳ Elliptical Clusters

↳ Not robust to outliers

↳ Initialize 'k' clusters or 'k' guassians

k = 2 (k means)



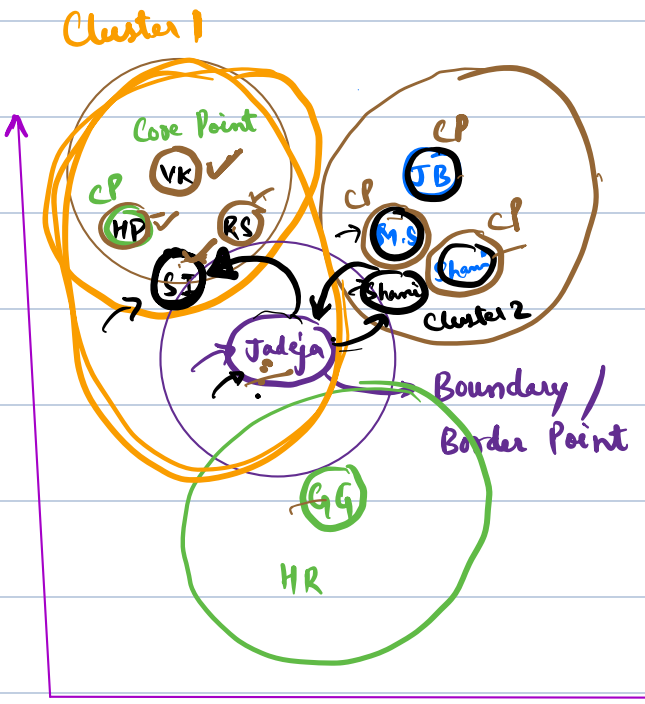
X → k = 3 (guassian)

" Clusters - 2 "

DBSCAN - Density based Spatial Clustering of Applications with 'Noise'

Parameters

- Radius - eps (epsilon)
- Min_points



radius → 1 metre

min-pts → 3 points/friends

VK → min-pts ✓

≥ 3

Min/points → No

Any core point → No

1

Min-points = 3

Points = 2 (< 3) X

↓ NOISE Data point

Steps of DBSCAN

① For each data point, create a circle of radius - eps.

↳ if inside the circle, there are more than or equal to 'min-points', then ⟶ core point

② If for a point, inside the circle there are less than min-points, then it's not a 'CORE POINT'.

But, if there is another C.P. inside the circle, then this data point ⟶ "Boundary / Border Point"

And if so, include in the same cluster as previously selected random point

③ If inside the circle, there are

(i) less than min-points

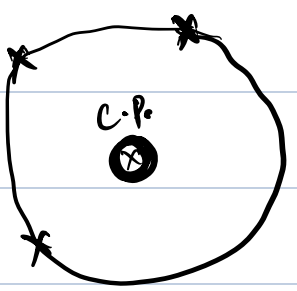
(ii) No core-point touching the circle

Data point ⟶ NOISE point (Outliers)

④ Repeat Steps 1, 2 and 3.

⑤ Till all points are covered / assigned.

NOTE: Border Points do not extend the clusters ahead.



min-points = 4

points = 3 touching + 1 itself

3 points

Limitations of k-means:

① Computationally expensive

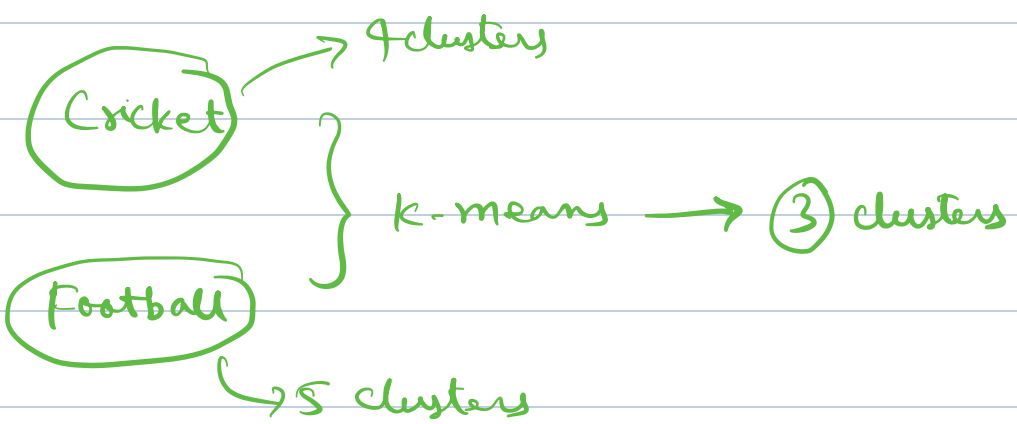
② Hard Clustering

③ Not robust to outliers

④ Spherical clusters

⑤ Almost equal sized clusters

⑥ Initializes 'k' centroids / clusters.



DBSCAN

↳ radius, min-points

↳ specific clusters

Limitations of GMM:

1. Not robust to outliers
2. Assumes elliptical shapes
3. Initialize 'k' distributions
4. Assumes 'Gaussian' Distribution

Significance of DBSCAN:

1. No need to specify the no. of clusters.
2. Detects / Handles Outliers as NOISE
3. Shape / Size can vary
4. Clusters are formed organically (checks density)

→ 20 rows
2 columns

```
[ [1,2],
  [2,1],
  [3,3],
]
```

Implementation of Clustering Algos →
"Make_moons"

Homework :

→ Take 1 popular dataset from Kaggle → for clustering

→ Implement all clustering algorithms.

Anamoly Detection → Detect / Handle.

→ "IQR" ✓