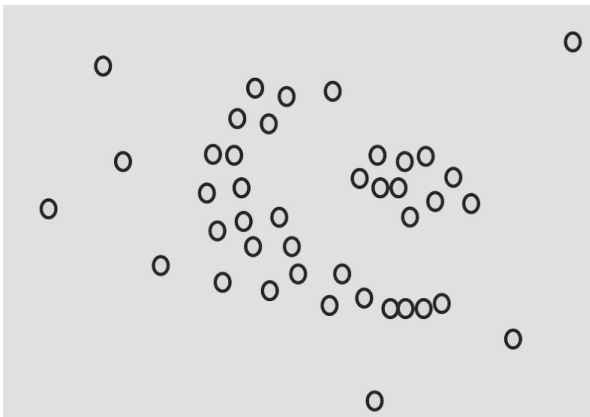


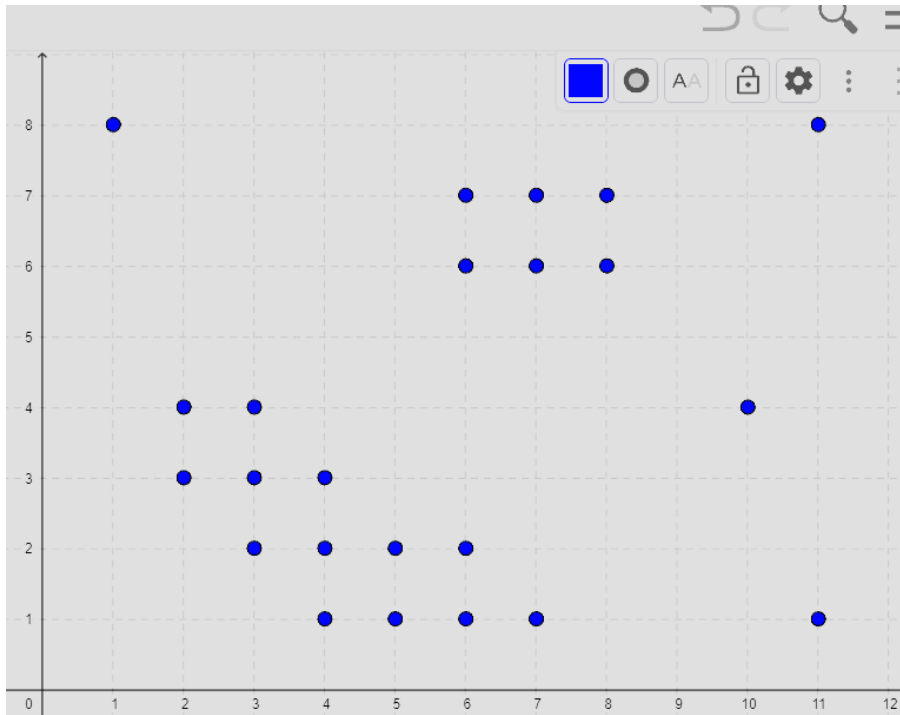
1. Nêu các bước thực hiện của thuật toán k-means
2. Cho các cặp điểm x,y có tọa độ như sau: (0.5; 0.5), (1; 0.5), (1; 1.5), (1.5; 1), (2.6; 2), (3; 2), (2.4; 2.5), (2.5; 3)
 - a. Áp dụng thuật toán k-means với số centroid bằng 2.
 - b. Khởi tạo tâm centroid tọa độ ngẫu nhiên, chạy thuật toán k-means 3 iterations và trả ra tọa độ 2 centroid.
3. Vẽ 3 trường hợp khác nhau có tập dữ liệu khi tiến hành phân cụm không hoạt động/ không tốt nếu sử dụng k-means. Nêu lí do. Đề xuất thuật toán phù hợp cho tập dữ liệu đó và giải thích tại sao thuật toán đó lại phù hợp
4. Thuật toán GMM
 - a. Viết Loss function cho thuật toán
 - b. Từ loss function chỉ ra hướng tiếp cận của thuật toán GMM để optimize loss này (E-M trong slide). Giải thích được ý nghĩa của E-step và M-step để làm gì (Có sử dụng công thức toán)
5. Nêu các bước thực hiện của thuật toán GMM
6. Kẻ bảng so sánh k-means và DBSCAN theo từng aspect
7. Cho hình sau:



Nêu cách thức hoạt động của dbscan theo ý hiểu (có thể dùng thêm toán để giải thích hoặc không).

Tham khảo link sau: <https://medium.com/mllearning-ai/dbscan-demystified-understanding-how-this-parameter-free-algorithm-works-89e03d7d7ab#:~:text=DBSCAN%20stands%20for%20Density-Based,their%20distance%20to%20other%20points.>

8. Cho các điểm trên hệ tọa độ OXY sau:



Áp dụng thuật toán DBSCAN với radius (epsilon) = 1.5 và min points = 3. Xác định các cụm và các noise points.