



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

Introdução à Inteligência Artificial

2023-2024

Questões Sobre Algoritmos de Agrupamento(Clustering)

© Ernesto Costa, Penousal Machado, João Correia, Nuno Lourenço, Tiago Martins, Tiago Baptista, Sérgio Rebelo, Pedro Silva, João Macedo, Luís Gonçalo, Jessica Parente, Marcio Lima, Luís Torres 2008-2024

Perguntas Teóricas

1. Diga o que entende por:
 - a. Aprendizagem não supervisionada
 - b. Algoritmo de Clustering
 - c. Centroide
 - d. Métrica de Semelhança no contexto de Clustering.
2. Distinga Clustering Hierárquico de Clustering de Partição (Partition).
3. Diga o que entende por “linkage” no contexto de algoritmos de clustering hierárquico.
4. Distinga Clustering de Partição por densidade e por centroide.
5. Comente a seguinte afirmação: O algoritmo de DBSCAN não define um número fixo de clusters antes da sua execução.
6. Comente a seguinte frase: No algoritmo DBSCAN, os clusters são formados apenas por pontos centrais (core).
7. Comente a seguinte frase: Utilizando o algoritmo k-means para um k predefinido e mantendo o mesmo conjunto de pontos, a cada iteração do algoritmo iremos obter sempre clusters constituídos de pontos diferentes independentemente do número de iterações que efectuarmos.

Perguntas Teórico-Práticas

1. Considerando a distribuição de dados (pontos) e a posição inicial dos centroides (triângulos) apresentada na Figura 1.

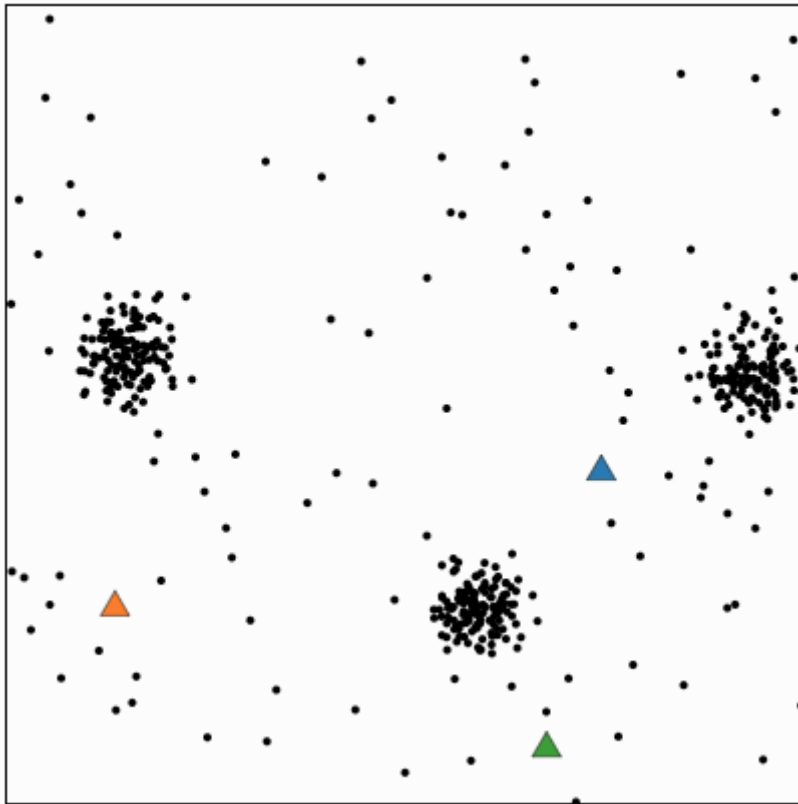


Figura 1.

- a) Enumere os passos da aplicação do algoritmo K-means Clustering.
- b) Aproxime, justificando, a solução final.

2. Considerando a distribuição de dados (pontos) e a posição inicial dos centroides (triângulos) apresentada na Figura 2. Aproxime, justificando a solução final.

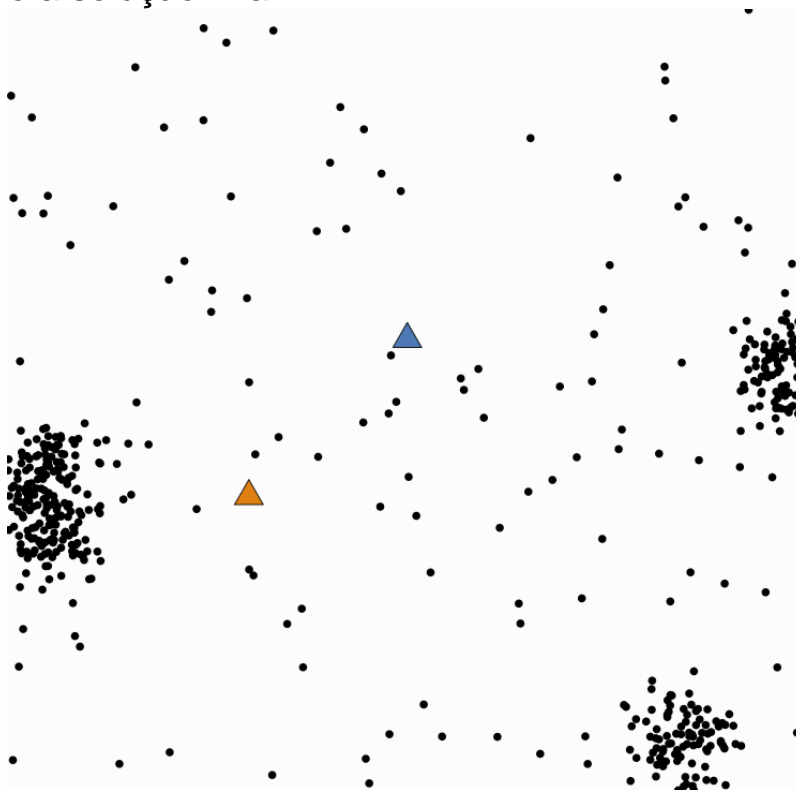


Figura 2.

3. Considere a seguinte tabela de pontos. Aplique o algoritmo de K-means até estabilizar, considerando $K = 2$ com centroides a começar na posição do A e C e utilizando a distância euclidiana.

Ponto	x	y
A	1	1
B	1	0
C	0	3
D	3	4
E	4	5

4. Considerando a distribuição de dados (pontos) da Figura 3, um de ϵ equivalente ao raio da circunferência e um número mínimo de pontos = 3). Aplicando o DBSCAN, identifique, justificando, os pontos alcançáveis (reachable), centrais (core) e *outliers* do cluster.

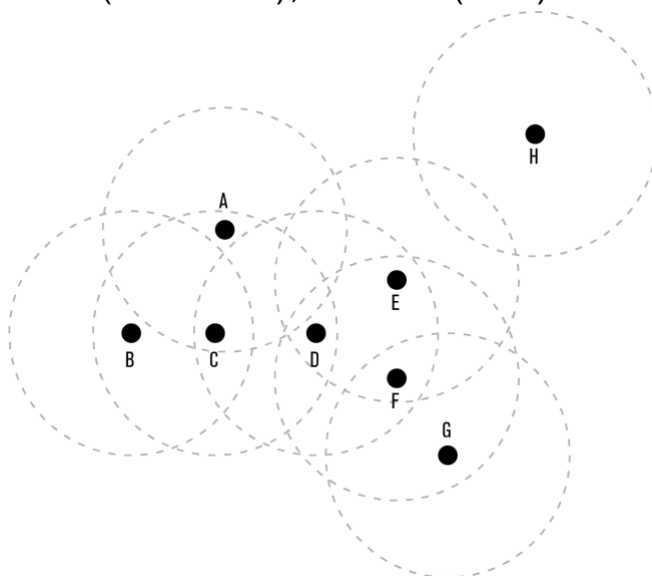


Figura 3.

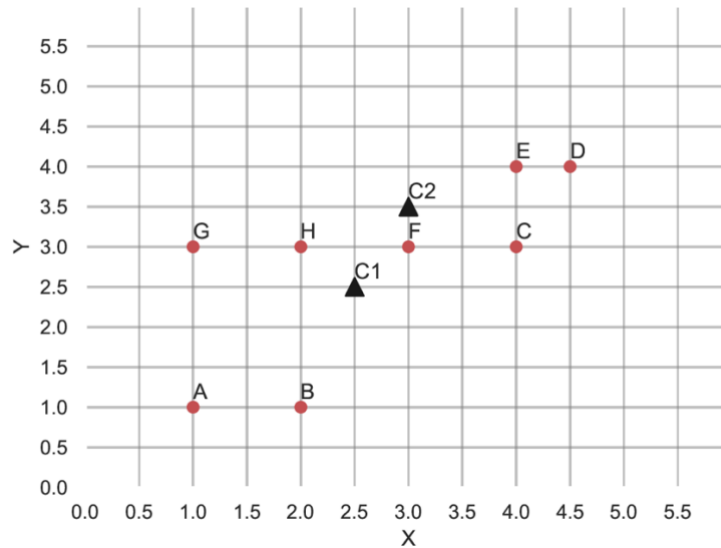
5. Considere a seguinte tabela de pontos. Utilize o algoritmo DBSCAN com número mínimo de pontos = 2 e ϵ de 1 em distância de Chebysheve.

Point	x	y
A	0	1
B	1	2
C	1	0
D	3	1
E	5	2
F	5	1
G	5	0
H	6	1

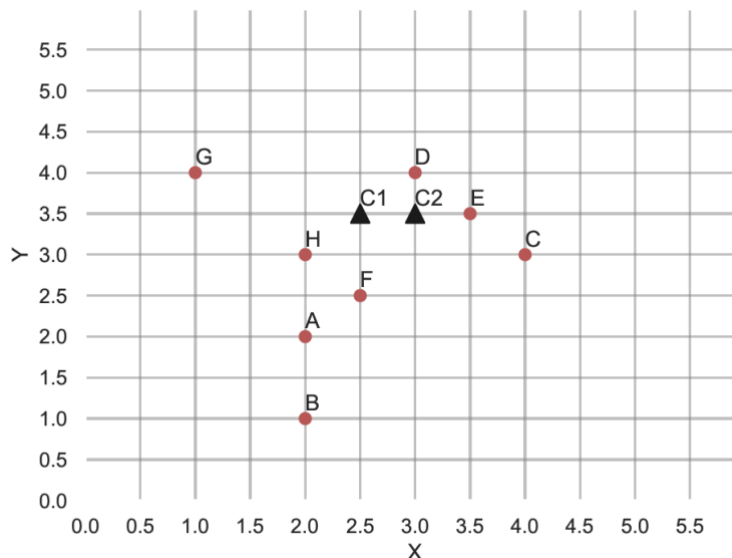
- Indique os clusters que se formam, que pontos são centrais, alcançáveis e outliers.
 - Repita a) mas considere um ϵ de 2 em distância Chebysheve
6. Considere a distribuição de pontos das figuras abaixo **a.** e **b.** .
 Aplique de forma rigorosa o algoritmo K-means até estabilizar a

cada uma delas separadamente, considerando $K=2$ com centros iniciais $C1$ e $C2$ que estão representados pelos triângulos na figura. Deverá indicar as coordenadas finais de cada um dos centros bem como a que cluster é que cada um dos pontos irá pertencer. Deverá considerar a distância Euclidiana para o cálculo de distâncias entre pontos.

a.

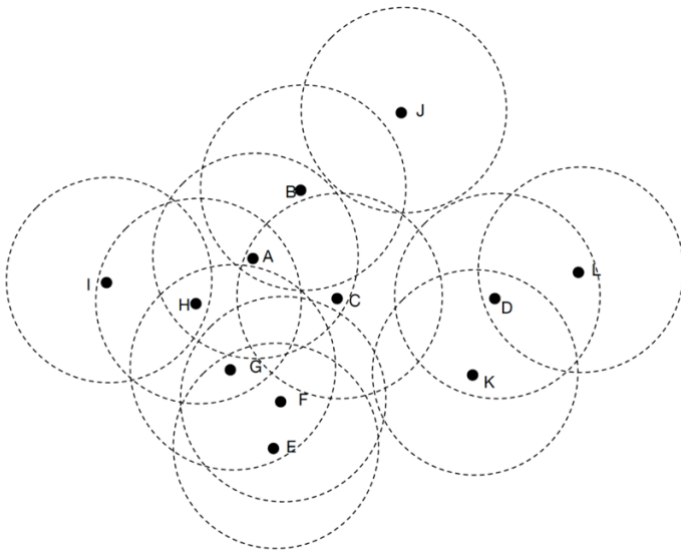


b.

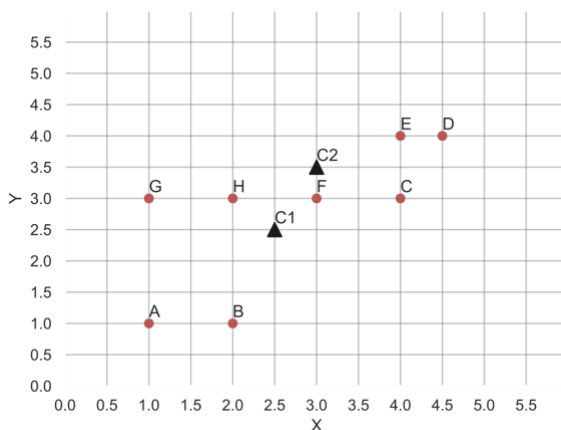


7. Considerando a distribuição de dados (pontos) da Figura 3, um de ϵ equivalente ao raio da circunferência e um número mínimo de pontos = 3). Aplicando o DBSCAN, identifique, justificando, os

pontos alcançáveis (reachable), centrais (core) e *outliers* do(s) cluster(s).

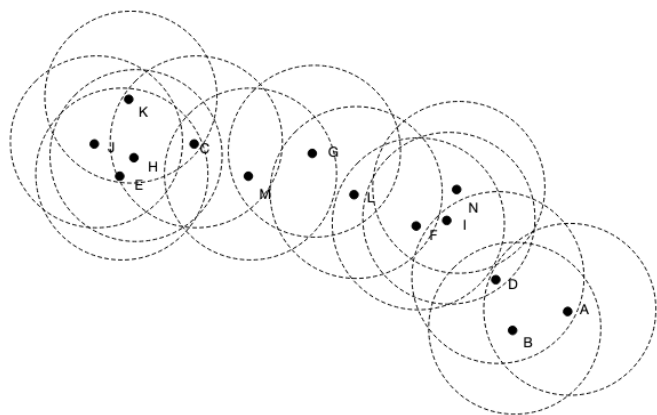


8. Aplique de forma rigorosa o algoritmo K-means até estabilizar a cada uma delas separadamente, considerando $K=2$ com centros iniciais C1 e C2 que estão representados pelos triângulos na figura. Deverá indicar as coordenadas finais de cada um dos centros bem como a que cluster é que cada um dos pontos irá pertencer. Deverá considerar a distância Euclidiana para o cálculo de distâncias entre pontos.



9. Considere a distribuição de dados (pontos) da figura, um ϵ equivalente ao raio da circunferência, um número mínimo de pontos = 4. Considere que um ponto conta para a sua própria vizinhança. Aplique o algoritmo DBSCAN a partir do ponto inicial A

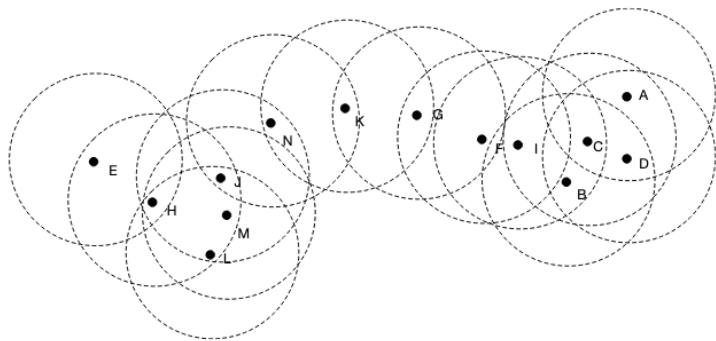
de forma a identificar o(s) cluster(s) existente(s) e os aos pontos alcançáveis (reachable) e centrais (core) do(s) cluster(s) bem como os ``outliers".



Classifique como verdadeira ou falsas as seguintes afirmações:

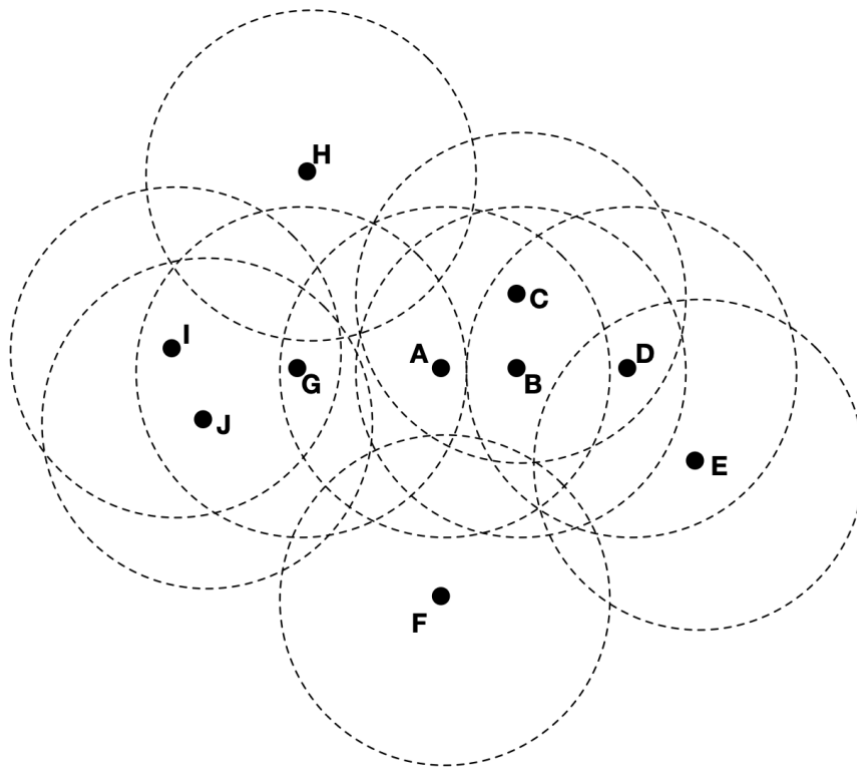
Afirmação	Verd.	Falso
Só há um outlier.		
Há exatamente dois clusters.		
O ponto B é um outlier.		
Os pontos C e M pertencem ao mesmo cluster.		
A e G pertencem a ao mesmo cluster.		

10. Considere a distribuição de dados (pontos) da figura, um épsilon equivalente ao raio da circunferência, um número mínimo de pontos = 4. Considere que um ponto conta para a sua própria vizinhança. Aplique o algoritmo DBSCAN a partir do ponto inicial A de forma a identificar o(s) cluster(s) existente(s) e os aos pontos alcançáveis (reachable) e centrais (core) do(s) cluster(s) bem como os ``outliers".

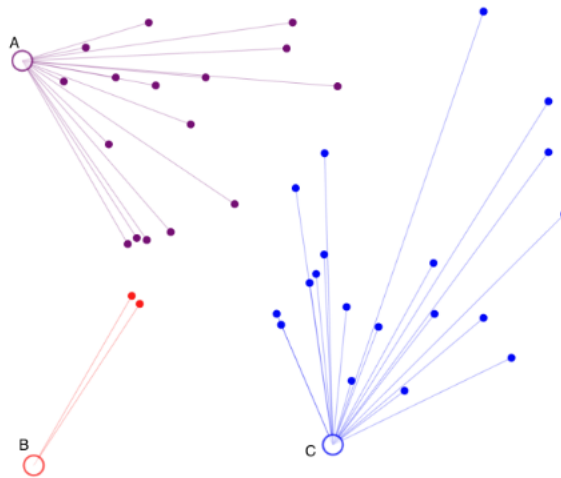


11. Considere a distribuição de dados (pontos) da figura, um épsilon equivalente ao raio da circunferência, um número mínimo de pontos = 3. Considere que um ponto conta para a sua própria vizinhança. Aplique o algoritmo DBSCAN a partir do ponto inicial A

de forma a identificar o(s) cluster(s) existente(s) e os pontos alcançáveis (reachable) e centrais (core) do(s) cluster(s) bem como os "outliers".



12. Considerando a distribuição de pontos, ondes os dados são representados por pontos, os centroides por circunferências e as arestas ligam os pontos ao centroide mais próximo.

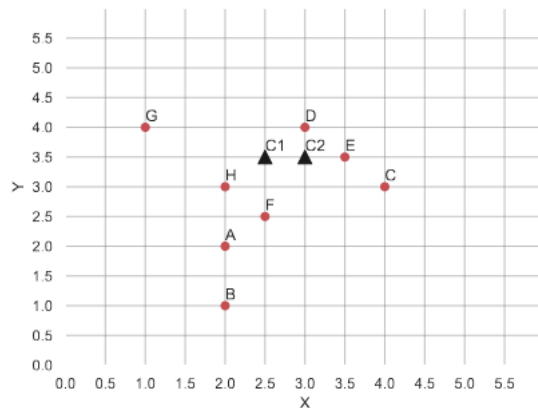


Classifique como verdadeira ou falsas as seguintes afirmações:

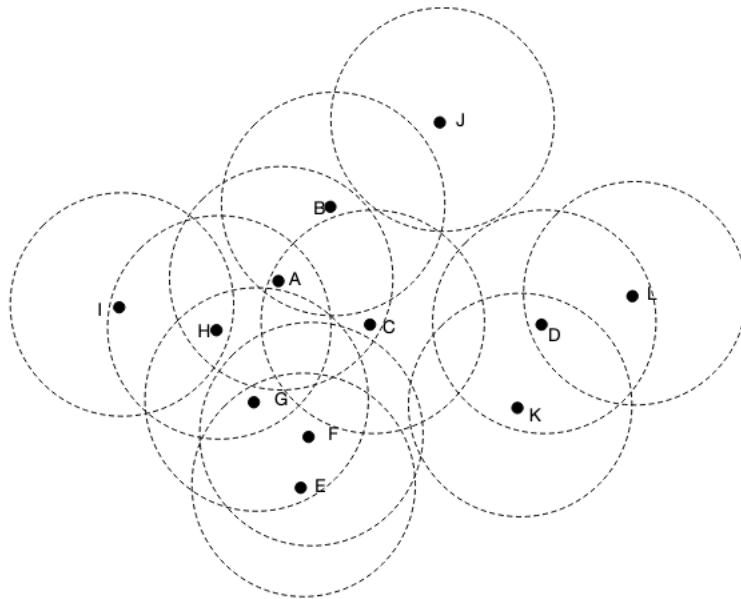
Afirmação	Verd.	Falso
O centroide C desloca-se para cima e direita		
O centroide B desloca-se para o ponto intermédio entre os dois pontos mais próximos de si.		
O centroide B acaba por morrer devido a ter poucos pontos e no final ficamos com 2 clusters.		
No k-means não há outliers.		

13. Considere a distribuição de dados (pontos) da figura abaixo. Aplique de forme rigorosa o algoritmo K-Means até estabilizar, considerando $K=2$ com os centros iniciais C1 e C2, e que estão representados pelos triângulos na figura. Deverá indicar as coordenadas finais de cada um dos centros bem como a que cluster é que cada um dos pontos irá pertencer. Deverá considerar

como métrica a distância Euclideana.

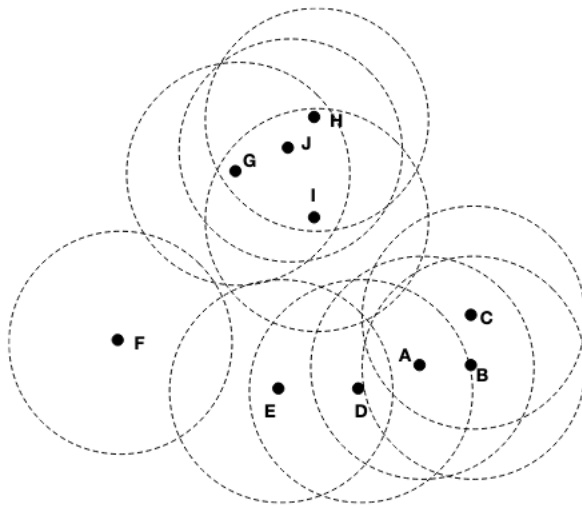


14. Considere a distribuição de dados (pontos) da figura, um ϵ psilon equivalente ao raio da circunferência, um número mínimo de pontos = 3. Considere que um ponto conta para a sua própria vizinhança. Aplique o algoritmo DBSCAN a partir do ponto inicial A de forma a identificar o(s) cluster(s) existente(s) e os pontos alcançáveis (reachable) e centrais (core) do(s) cluster(s) bem como os "outliers"



15. Considere a distribuição de dados (pontos) da figura, um ϵ psilon equivalente ao raio da circunferência, um número mínimo de pontos = 3. Considere que um ponto conta para a sua própria vizinhança. Aplique o algoritmo DBSCAN a partir do ponto inicial A de forma a identificar o(s) cluster(s) existente(s) e os pontos alcançáveis (reachable) e centrais (core) do(s) cluster(s) bem como

os ``outliers''



16. Considere a distribuição de dados (pontos) da figura abaixo. Aplique de forma rigorosa o algoritmo K-Means até estabilizar, considerando $K=2$ com os centros iniciais $C1$ e $C2$, e que estão representados pelos triângulos na figura. Deverá indicar as coordenadas finais de cada um dos centros bem como a que cluster é que cada um dos pontos irá pertencer. Deverá considerar como métrica a distância Euclideana.

