

Intelligence artificielle en traitement de l'image: K-Means & DBSCAN

Pierre Minier, Merwan Muller, Baptiste Roulliaux

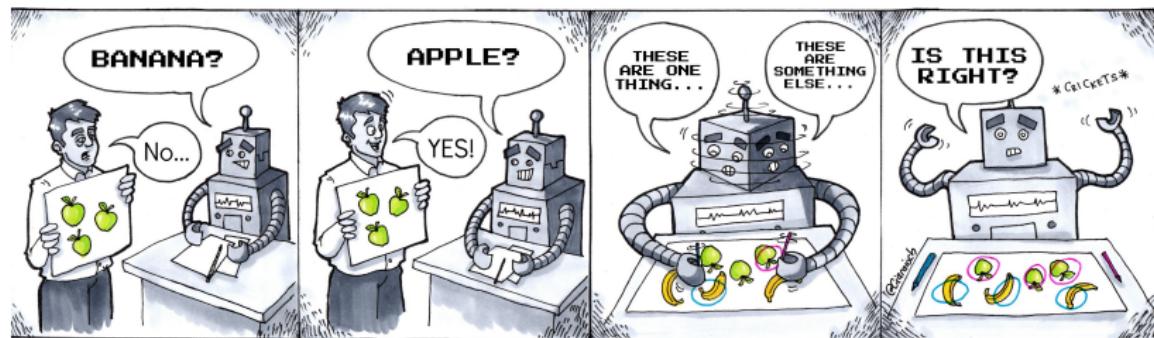
Février 2022

Introduction générale

Quelques définitions

Définition: Intelligence Artificielle

L'ensemble des théories et des techniques mises en oeuvre en vue de réaliser des machines capables de simuler l'intelligence humaine.



Supervised Learning

Unsupervised Learning

Introduction générale

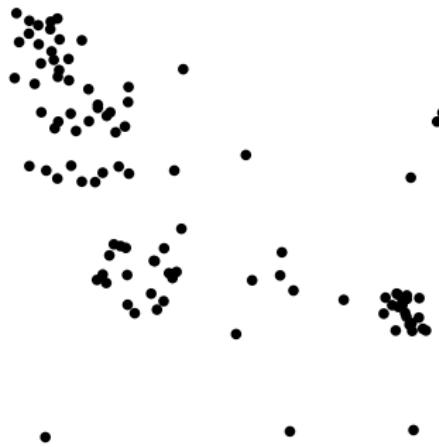
Plan de la présentation



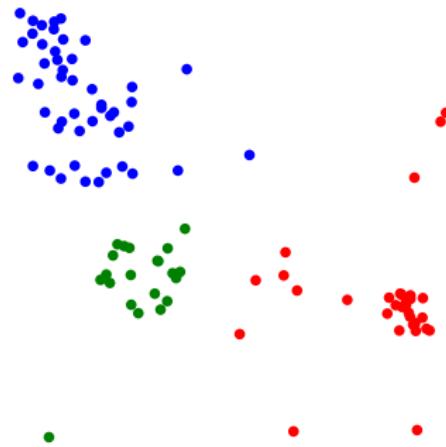
Sommaire

- ① Description de l'algorithme K-Means
- ② Améliorations de K-Means
- ③ Description de l'algorithme DBSCAN
- ④ Implémentation de DBSCAN
- ⑤ Comparaison des deux méthodes

Objectifs de K-Means



(a) Entrée



(b) Sortie

Fonctionnement basique de l'algorithme

Initialisation

K points tirés au hasard constituent les K clusters.

Fonctionnement basique de l'algorithme

Initialisation

K points tirés au hasard constituent les K clusters.

Affectation d'un point

Le cluster choisi minimise la distance entre son centroïde et le point.

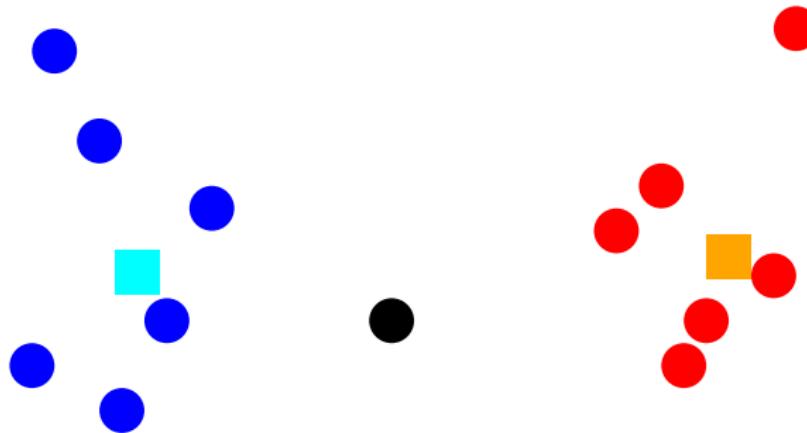
Fonctionnement basique de l'algorithme

Initialisation

K points tirés au hasard constituent les K clusters.

Affectation d'un point

Le cluster choisi minimise la distance entre son centroïde et le point.



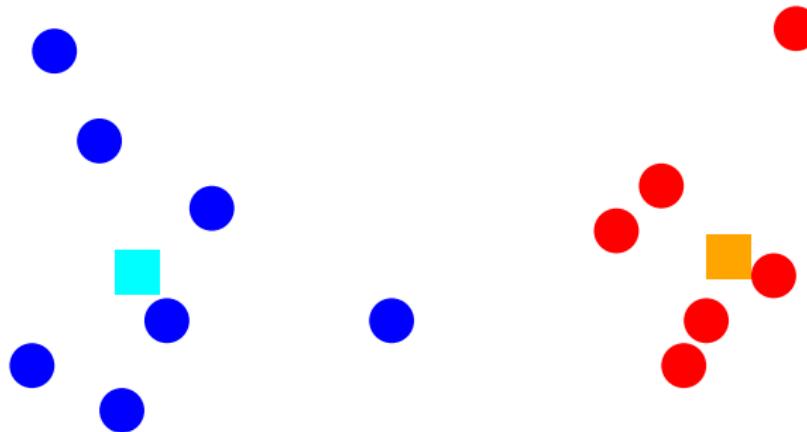
Fonctionnement basique de l'algorithme

Initialisation

K points tirés au hasard constituent les K clusters.

Affectation d'un point

Le cluster choisi minimise la distance entre son centroïde et le point.



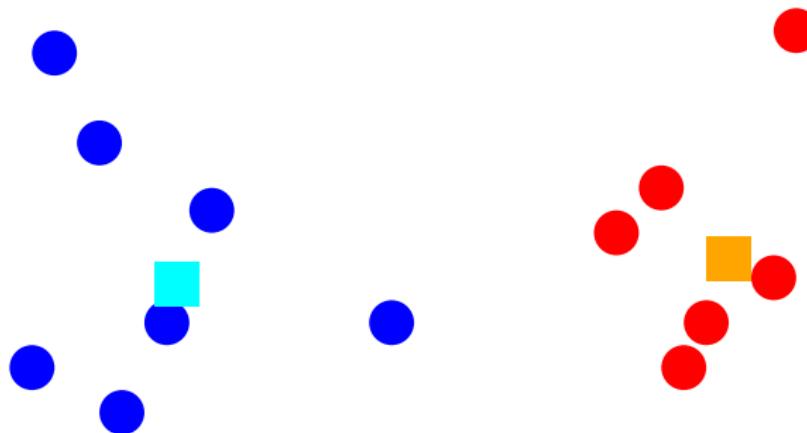
Fonctionnement basique de l'algorithme

Initialisation

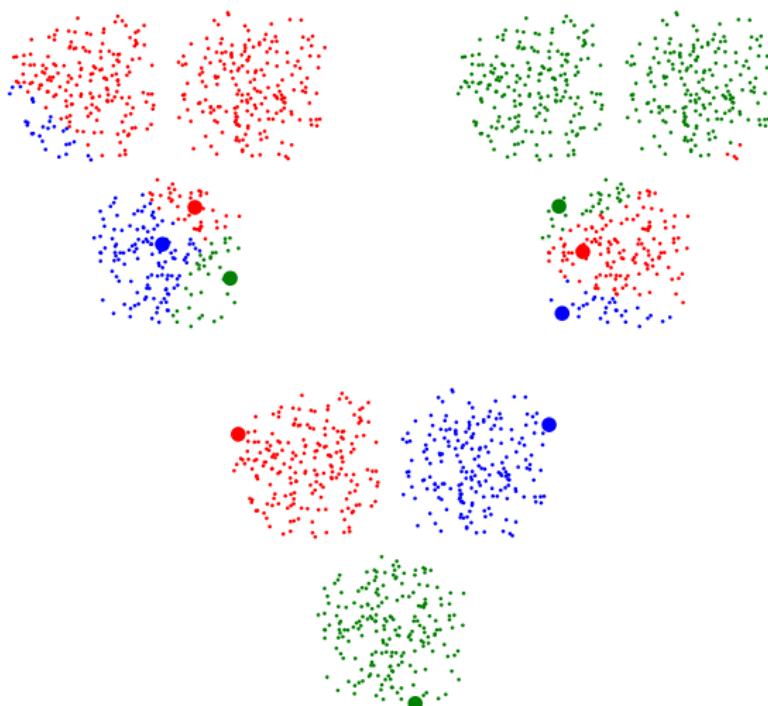
K points tirés au hasard constituent les K clusters.

Affectation d'un point

Le cluster choisi minimise la distance entre son centroïde et le point.



Initialisation aléatoire: problèmes



Améliorations

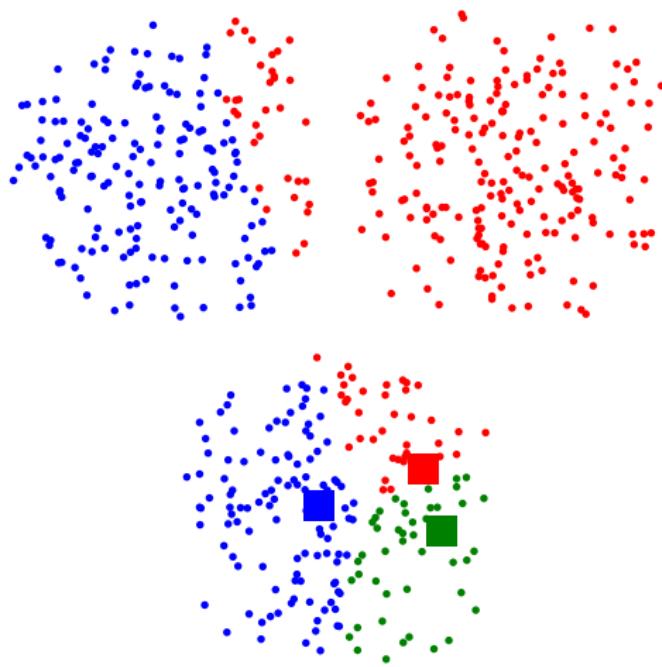
Répétition de l'algorithme

Convergence des centroïdes

- 1: Sélectionner K points au hasard pour former les K centroïdes.
- 2: **Faire**
- 3: Affecter les points aux clusters les plus proches.
- 4: Calculer les nouveaux centroïdes.
- 5: **Tant que** les centroïdes changent.
- 6: **Retourner** les derniers clusters formés

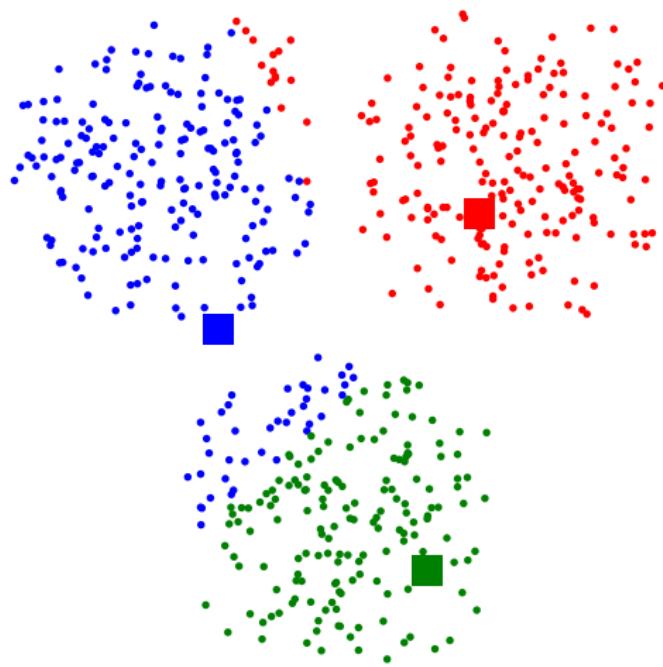
Améliorations

Répétition de l'algorithme: exemple



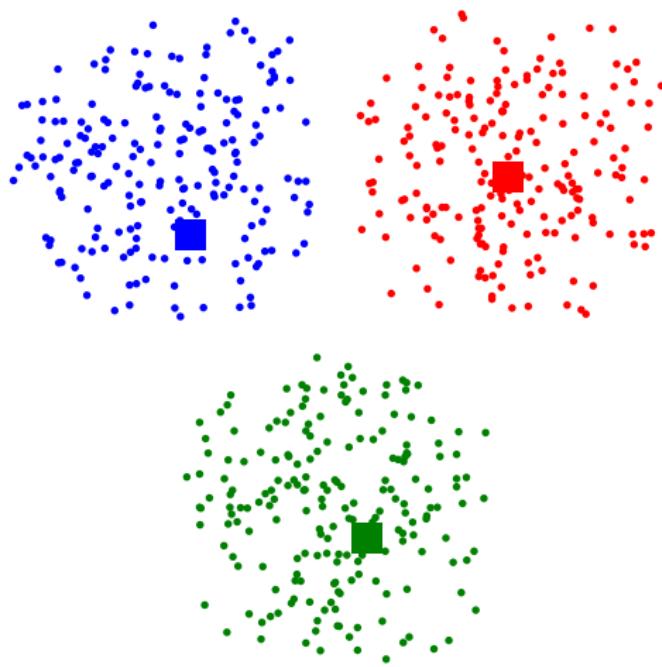
Améliorations

Répétition de l'algorithme: exemple



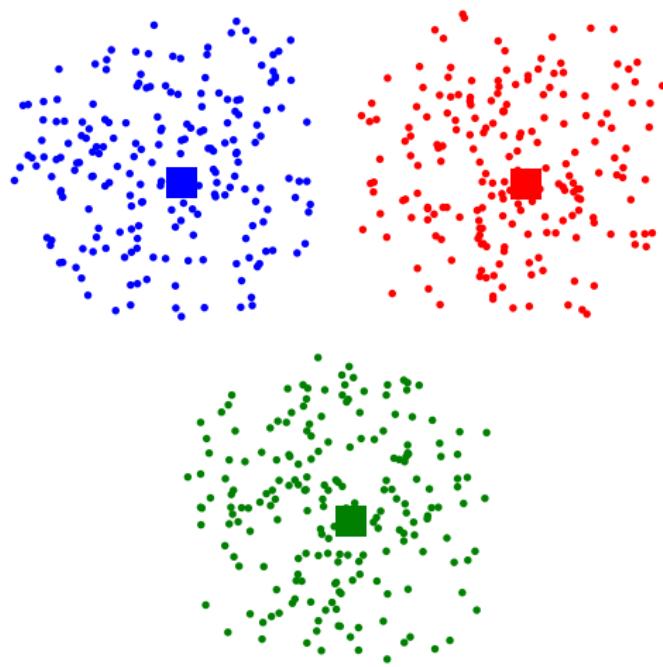
Améliorations

Répétition de l'algorithme: exemple



Améliorations

Répétition de l'algorithme: exemple



Améliorations

Initialisation déterministe

K-Means++

Initialiser les centroïdes avec les K points les plus éloignés.

Améliorations

Initialisation déterministe

K-Means++

Initialiser les centroïdes avec les K points les plus éloignés.

K centroïdes éloignés

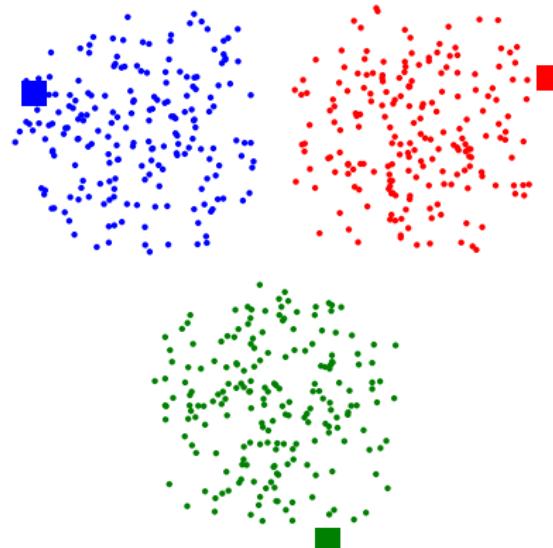
- 1: Sélectionner le premier centroïde au hasard.
- 2: **Pour** k allant de 2 à K
 - 3: Enregistrer le point maximisant la somme des distances avec les $(k-1)$ centroïdes connus.
- 4: **Fin pour**
- 5: Remplacer le premier centroïde par le point maximisant la somme des distances avec les autres centroïdes.
- 6: **Retourner** les centroïdes

Améliorations

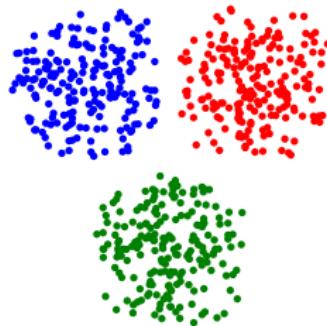
Initialisation déterministe

K-Means++

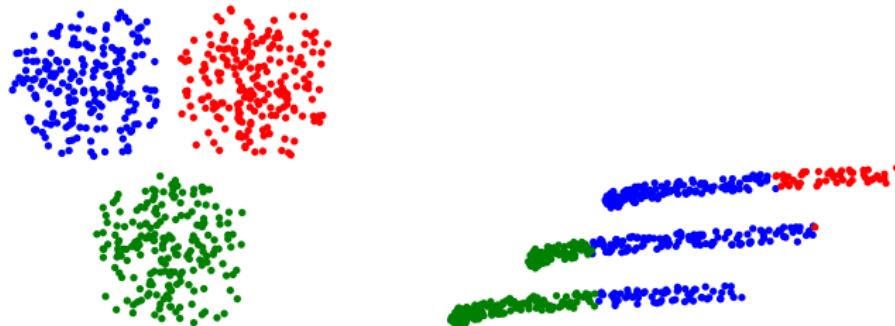
Initialiser les centroïdes avec les K points les plus éloignés.



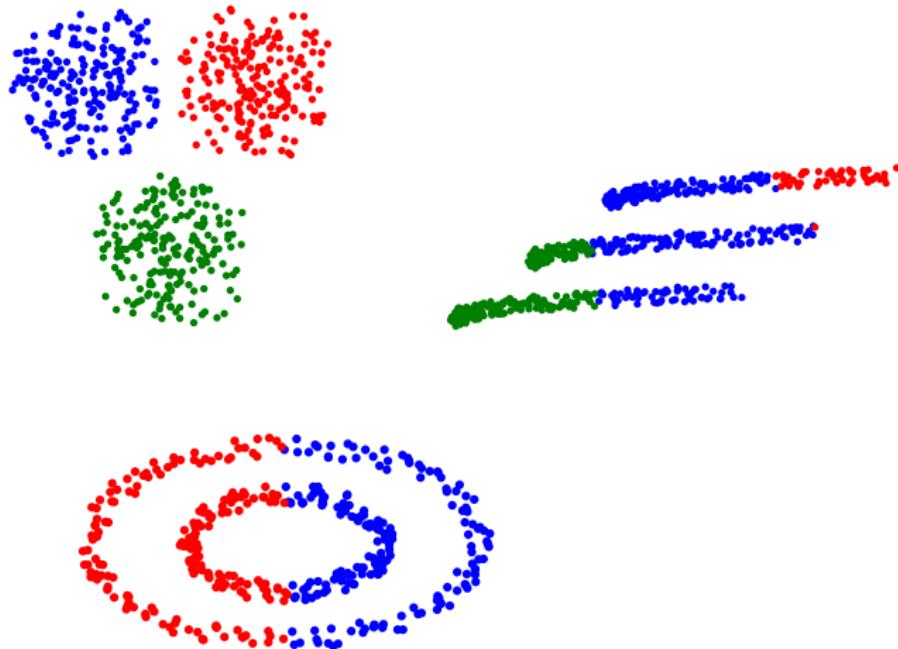
Résultats



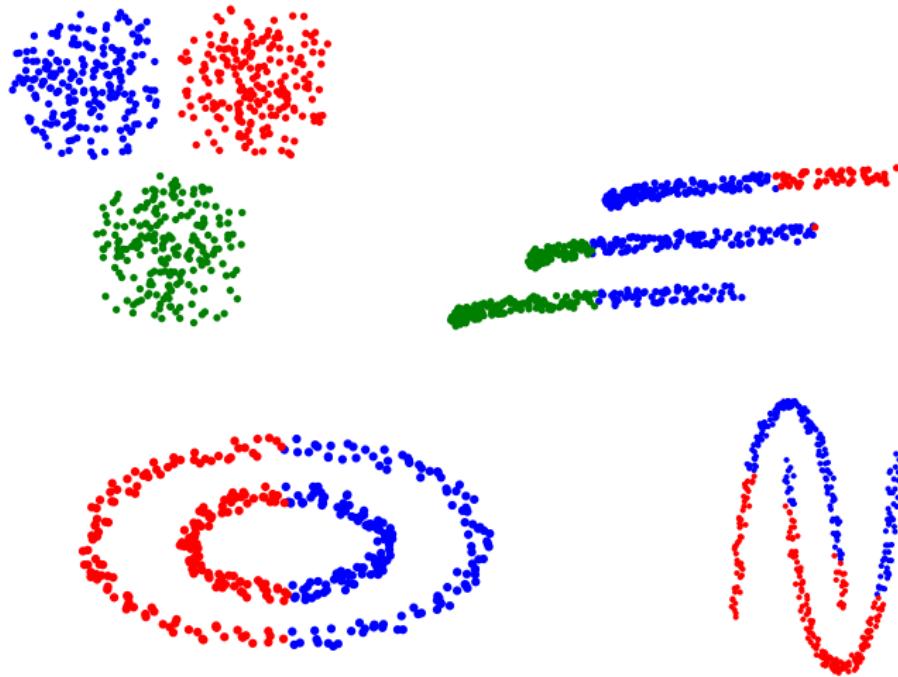
Résultats



Résultats



Résultats



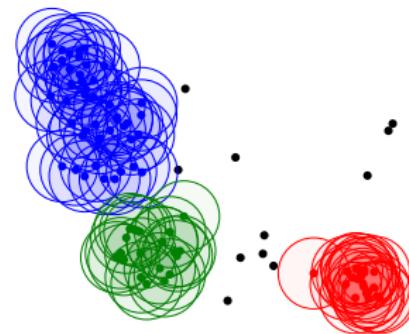
Objectifs

DBSCAN

DBSCAN est un algorithme proposé en 1996. Il repose sur un étude locale des points, et écarte le bruit des mesures.



(v) Entrée



(w) Sortie

Principes

Entrées de l'algorithme



Initialisation

- ① Les points à traiter, tous étiquetés comme "non visités".
- ② $min_p \in \mathbb{N}^*$: le nombre minimal de points d'un cluster.
- ③ $\epsilon \in \mathbb{R}_+^*$: le rayon seuil de voisinage.

Principes

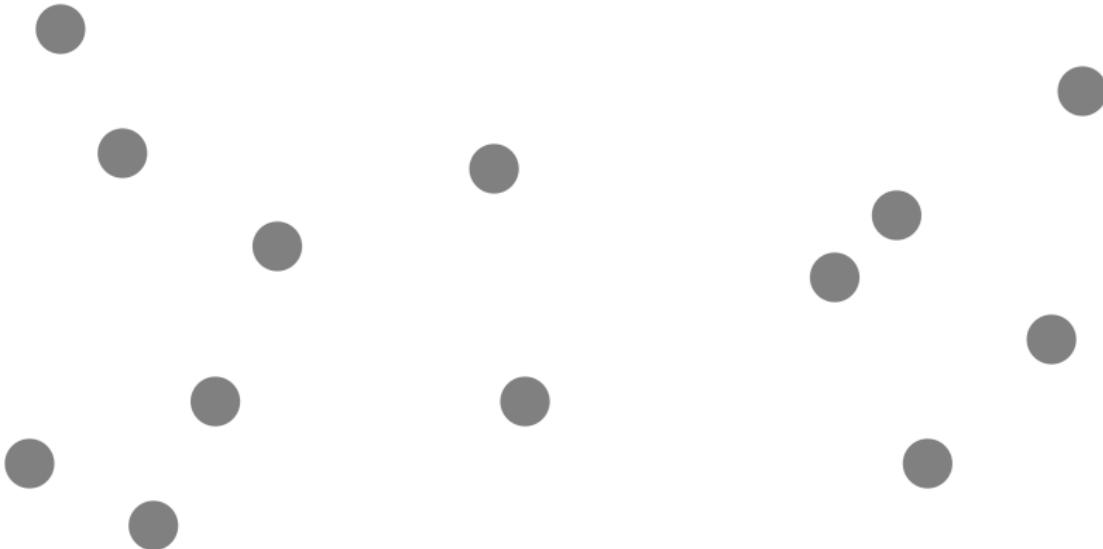
Détection d'un cluster

Détection d'un nouveau cluster

- 1: Sélectionner un point P "non visité"
- 2: **Si** $\mathcal{B}(P, \epsilon)$ contient au moins min_p points
- 3: Passer P en "visité"
- 4: Enregistrer un nouveau cluster \mathcal{C}
- 5: Affecter les points à l'intérieur de $\mathcal{B}(P, \epsilon)$ à \mathcal{C}
- 6: Diffuser \mathcal{C}
- 7: **Sinon**
- 8: Passer P en "bruit"
- 9: **Fin Si**

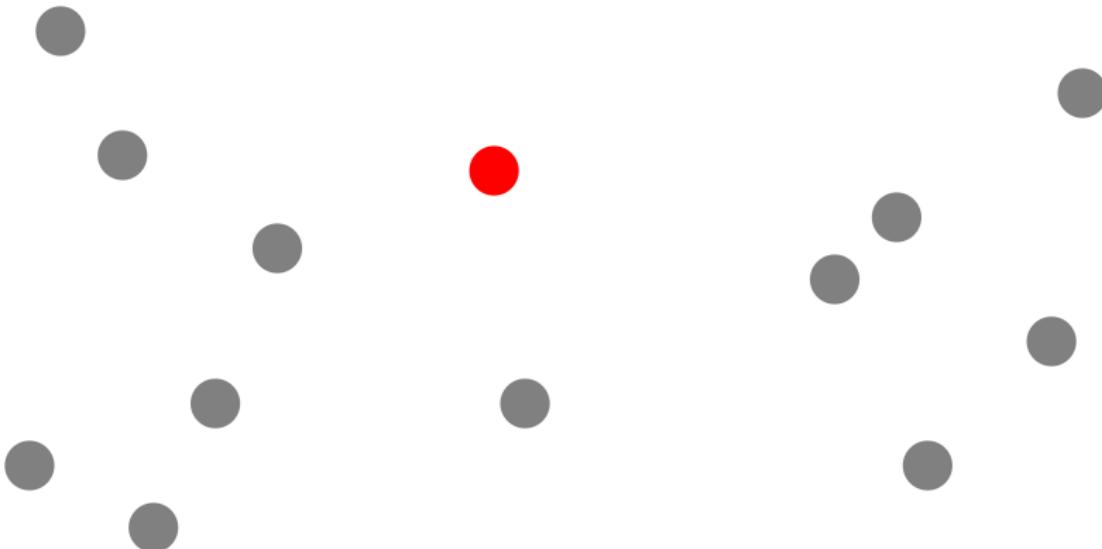
Principe

Détection d'un cluster: exemple



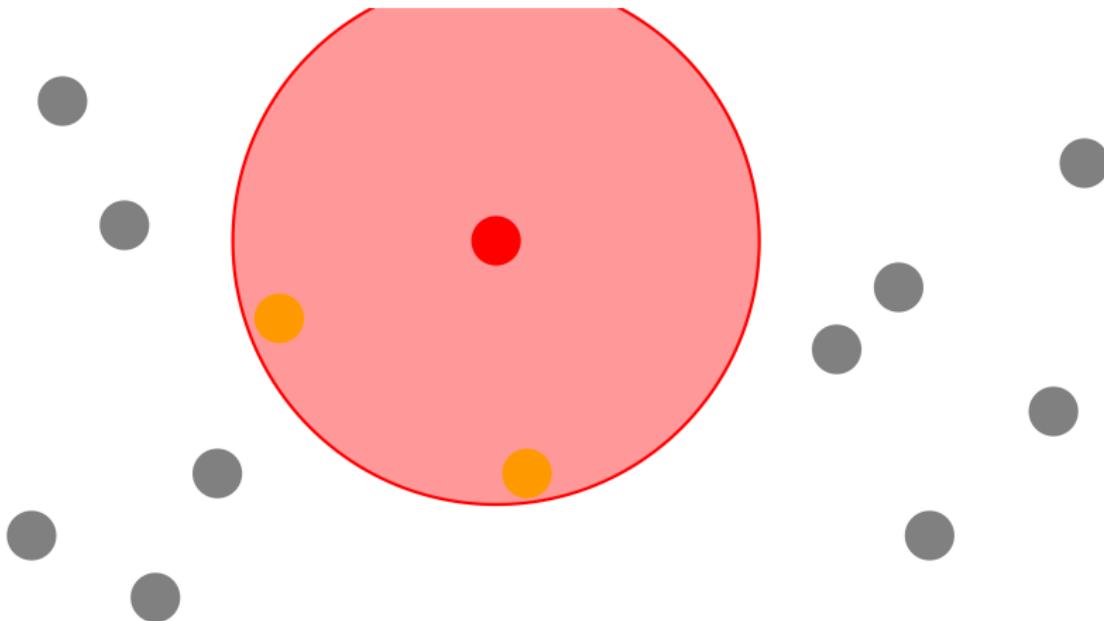
Principe

Détection d'un cluster: exemple



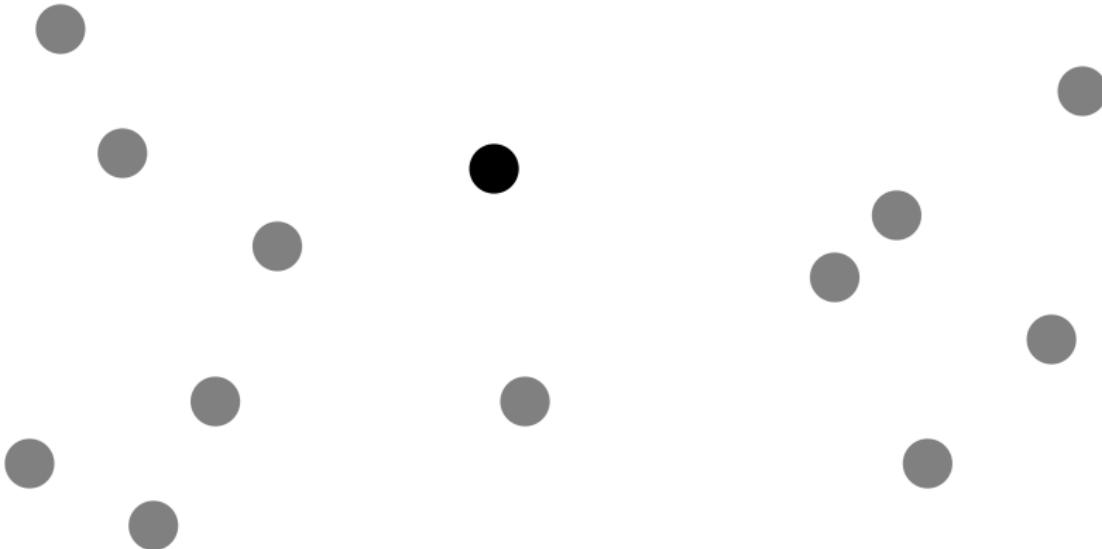
Principe

Détection d'un cluster: exemple



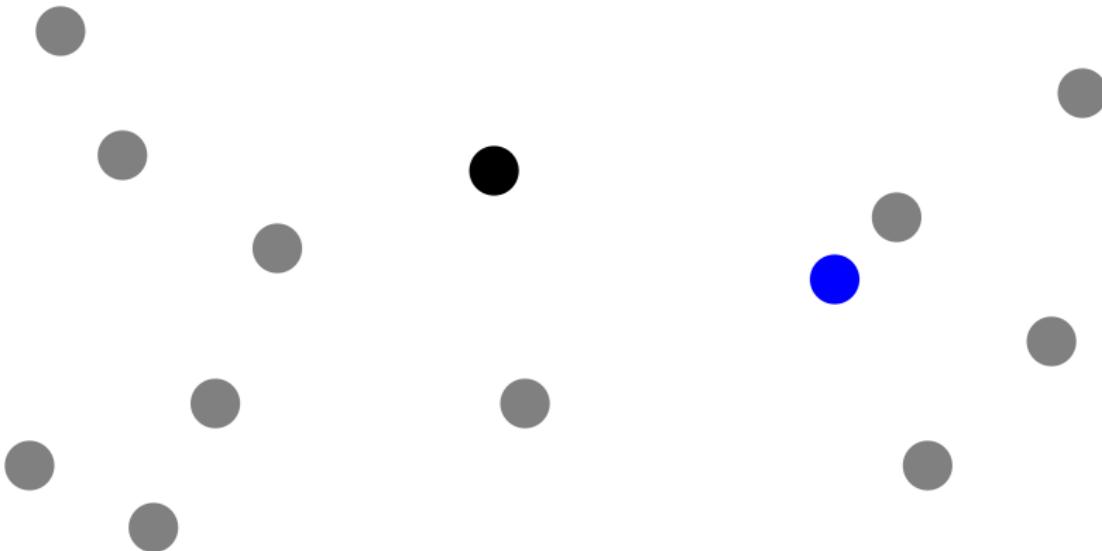
Principe

Détection d'un cluster: exemple



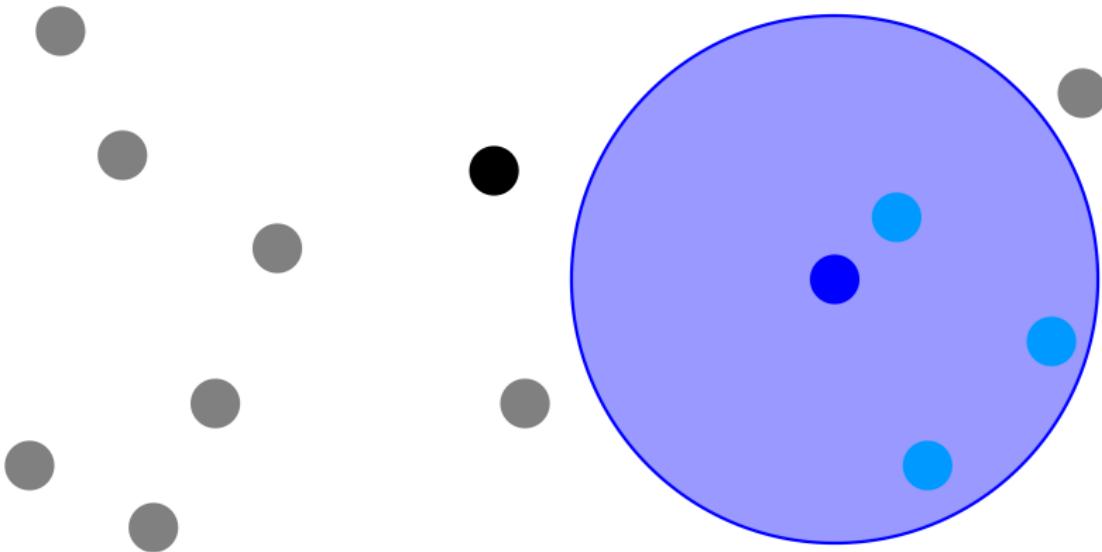
Principe

Détection d'un cluster: exemple



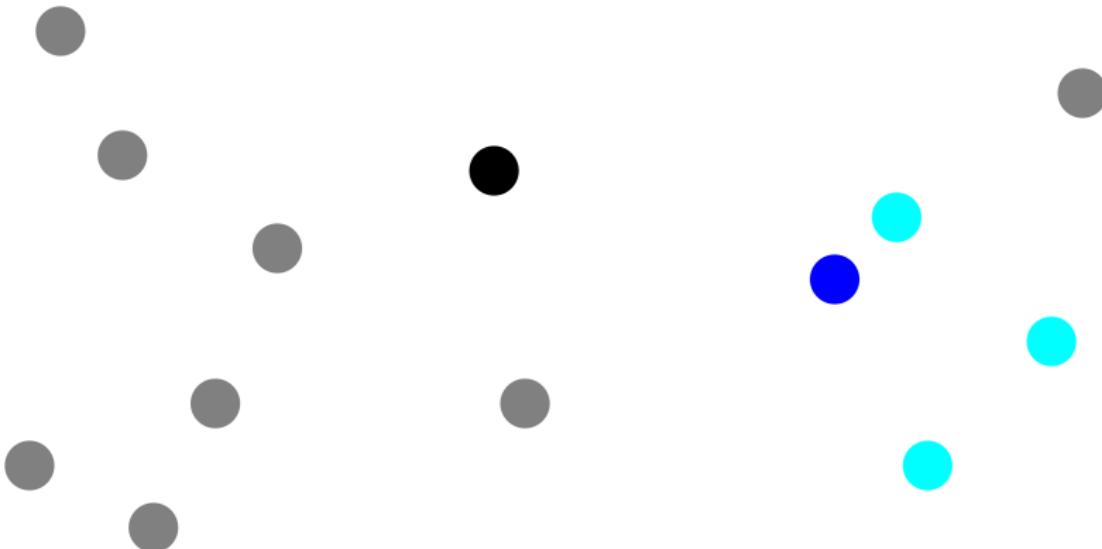
Principe

Détection d'un cluster: exemple



Principe

Détection d'un cluster: exemple



Principes

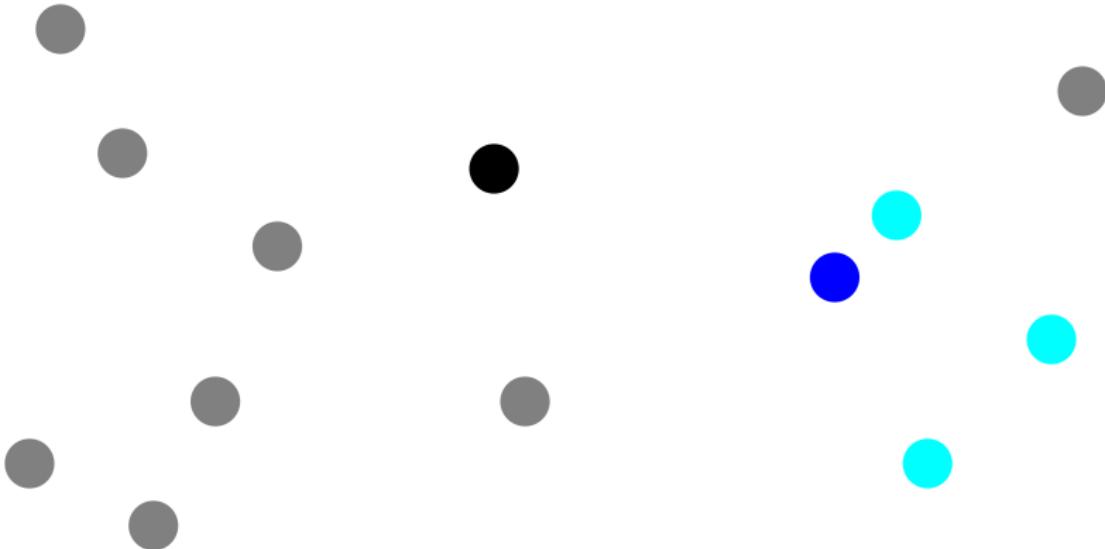
Diffusion d'un cluster

Diffusion d'un nouveau cluster

- 1: **Tant que** tous les points de \mathcal{C} n'ont pas été visités
- 2: Sélectionner un point P "non visité"
- 3: Passer P en "visté"
- 4: Affecter les points "non visité" ou "bruit" de $\mathcal{B}(P, \epsilon)$ à \mathcal{C}
- 5: **Fin Tant que**

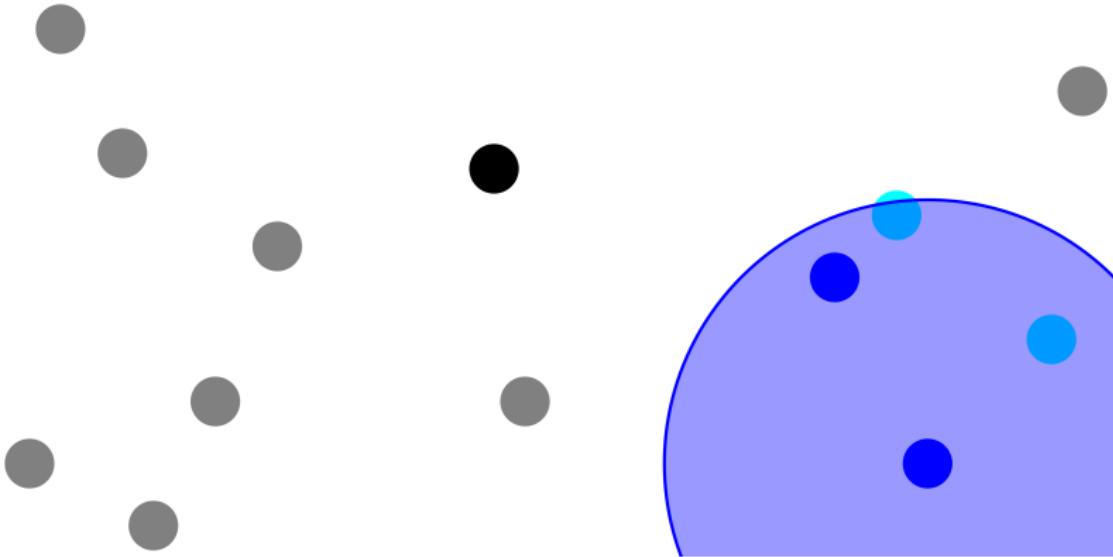
Principe

Diffusion d'un cluster: exemple



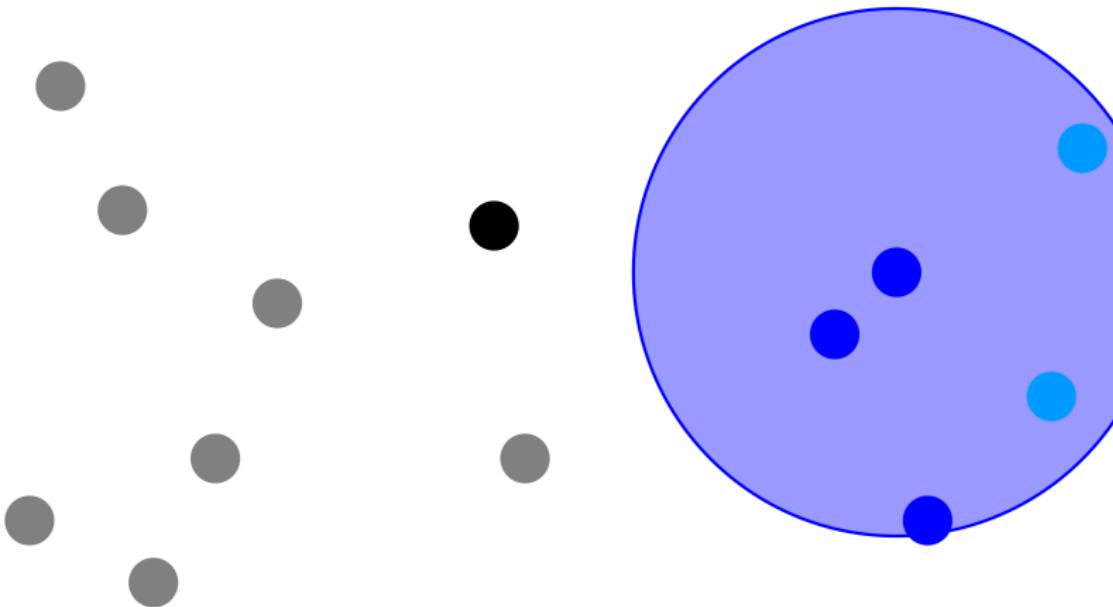
Principe

Diffusion d'un cluster: exemple



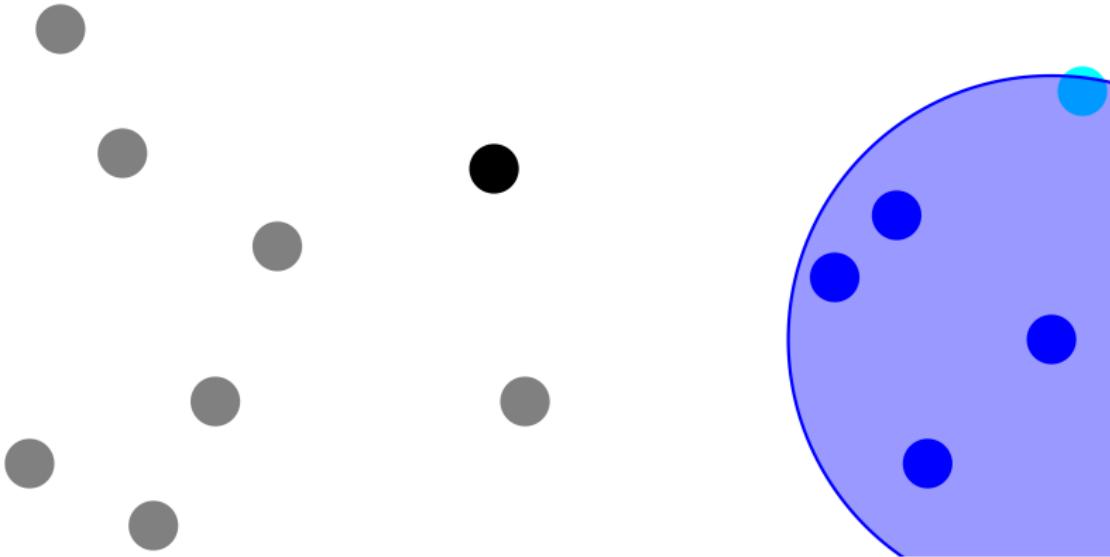
Principe

Diffusion d'un cluster: exemple



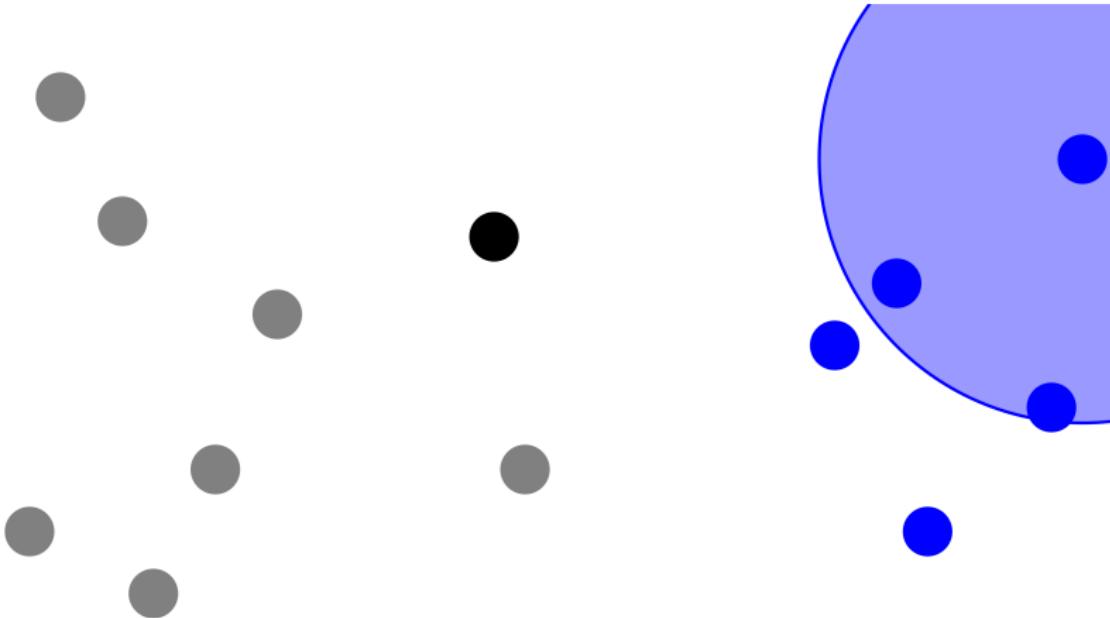
Principe

Diffusion d'un cluster: exemple



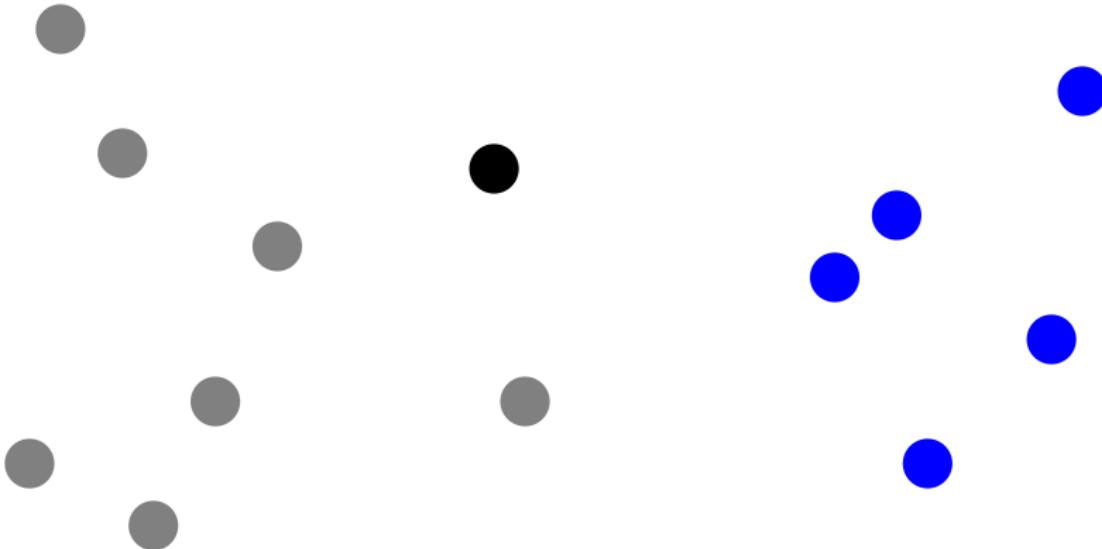
Principe

Diffusion d'un cluster: exemple



Principe

Diffusion d'un cluster: exemple



Principes

Terminaison et complexité

Terminaison de DBSCAN

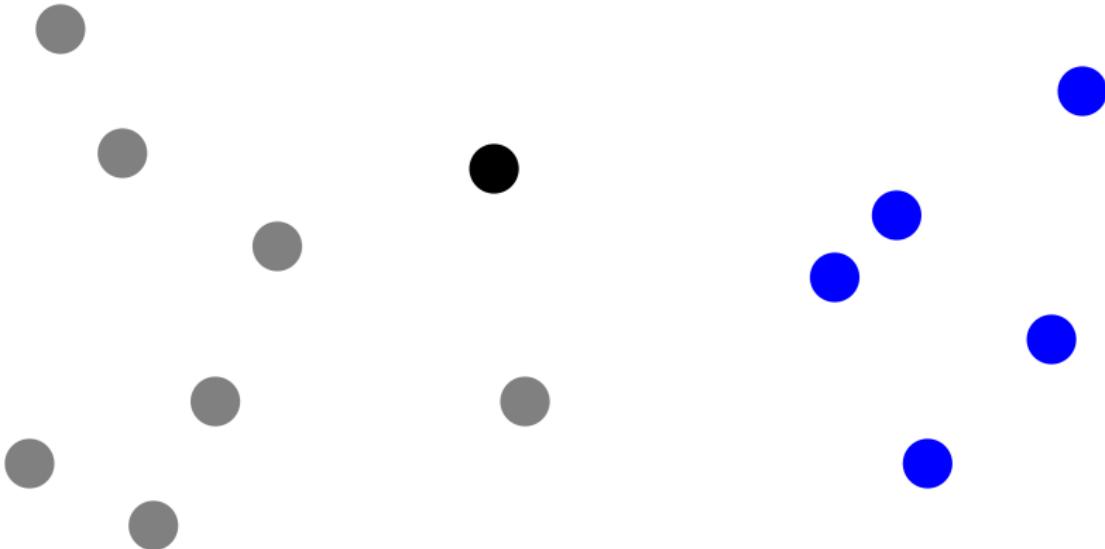
L'algorithme se termine lorsque tous les points sont étiquetés comme "visité" ou "bruit".

Complexité de DBSCAN

En notant N le nombre de points, la complexité est en $O(N^2)$.

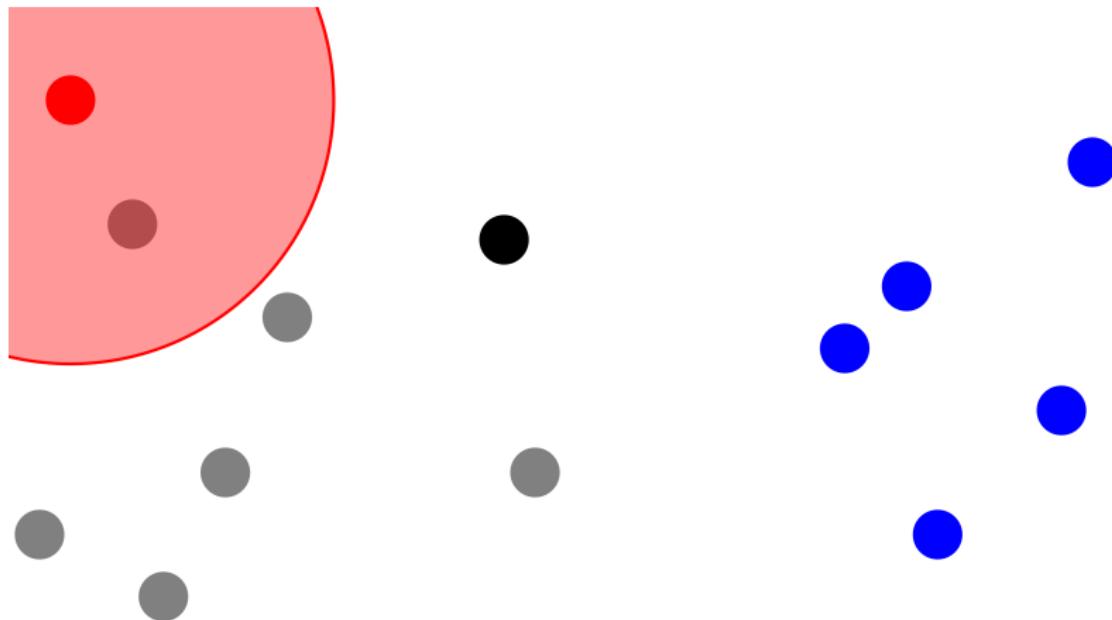
Principe

Terminaison d'un cluster: exemple



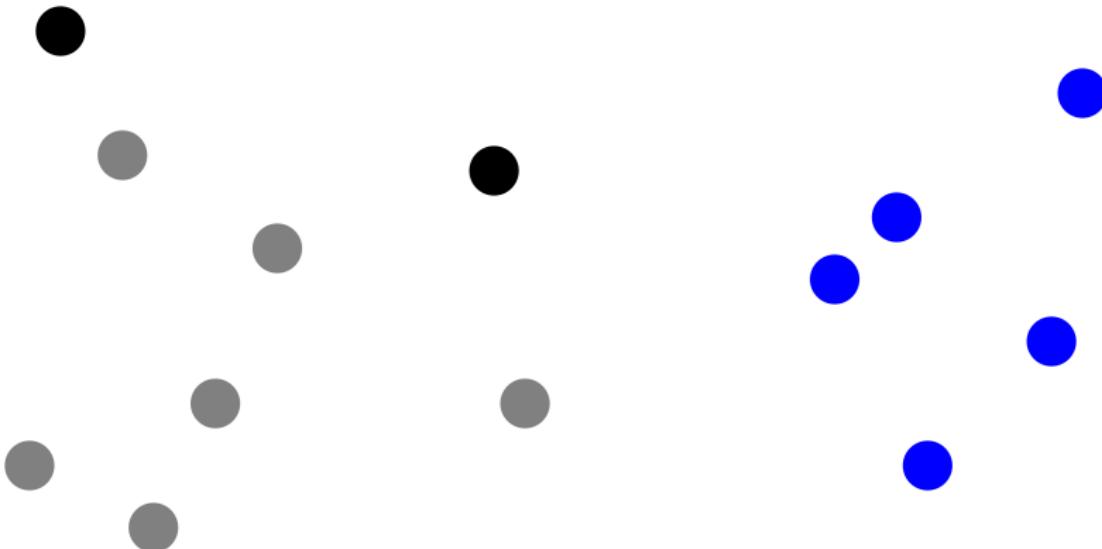
Principe

Terminaison d'un cluster: exemple



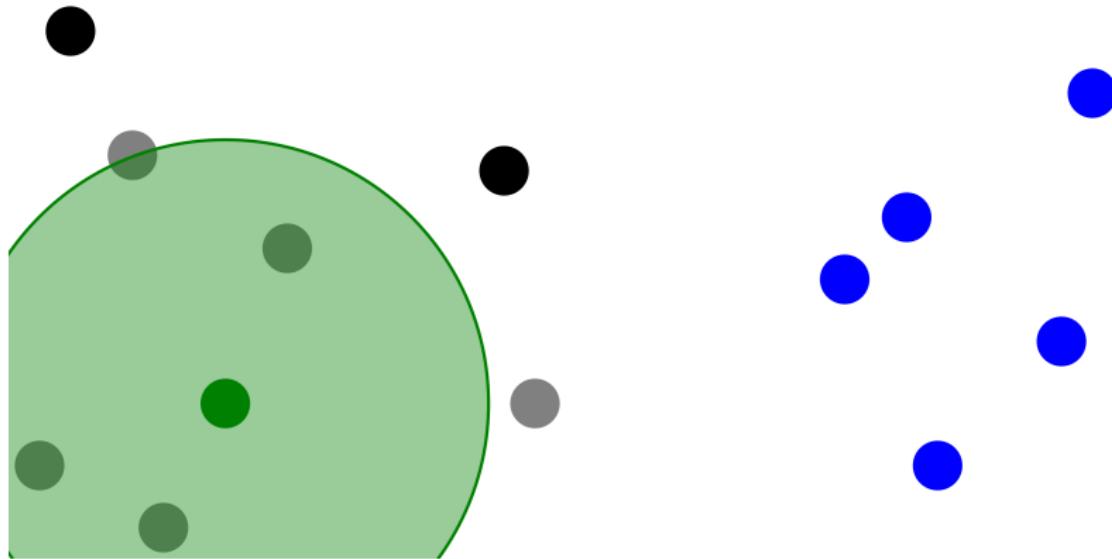
Principe

Terminaison d'un cluster: exemple



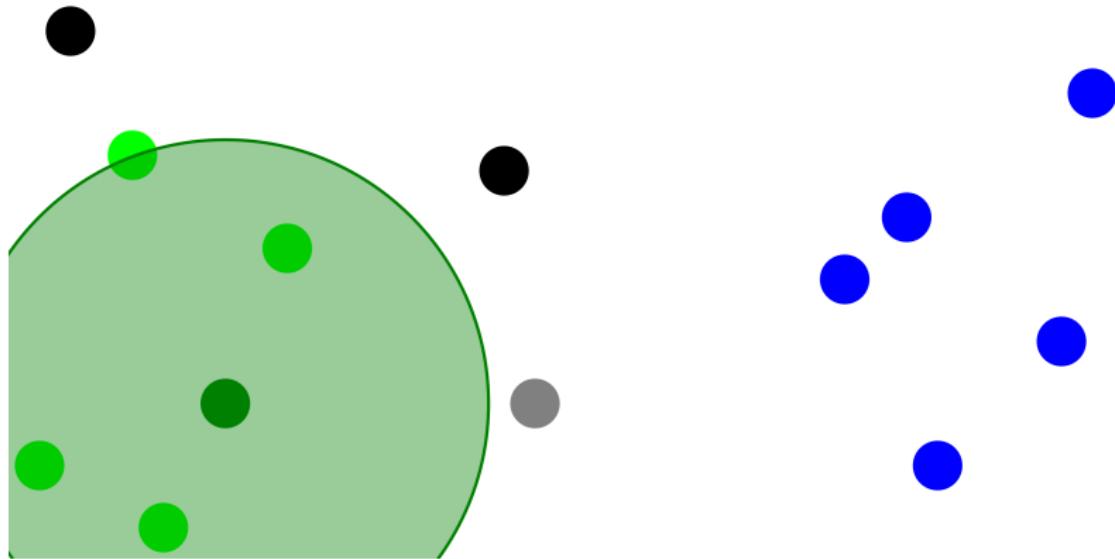
Principe

Terminaison d'un cluster: exemple



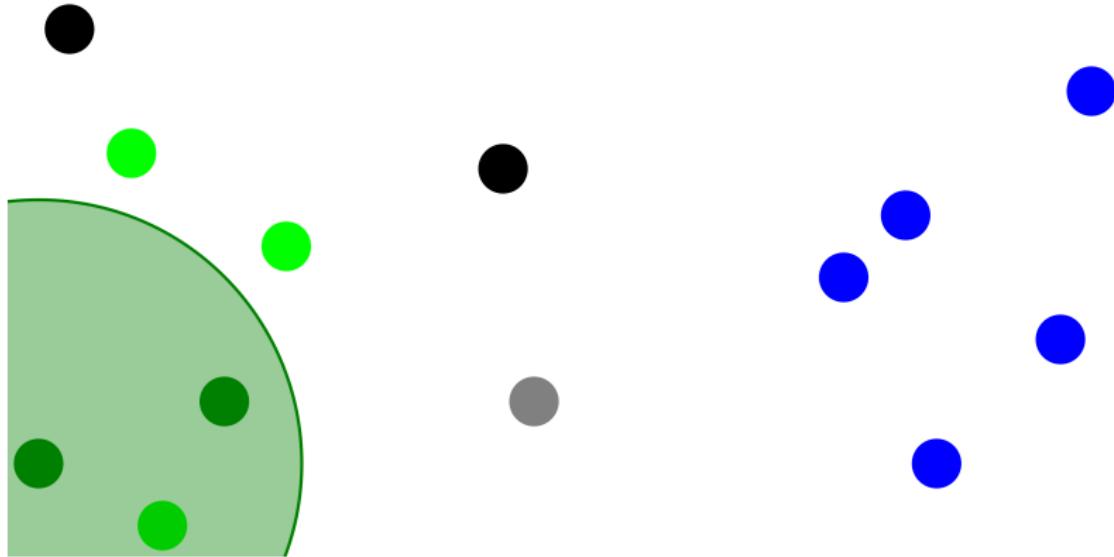
Principe

Terminaison d'un cluster: exemple



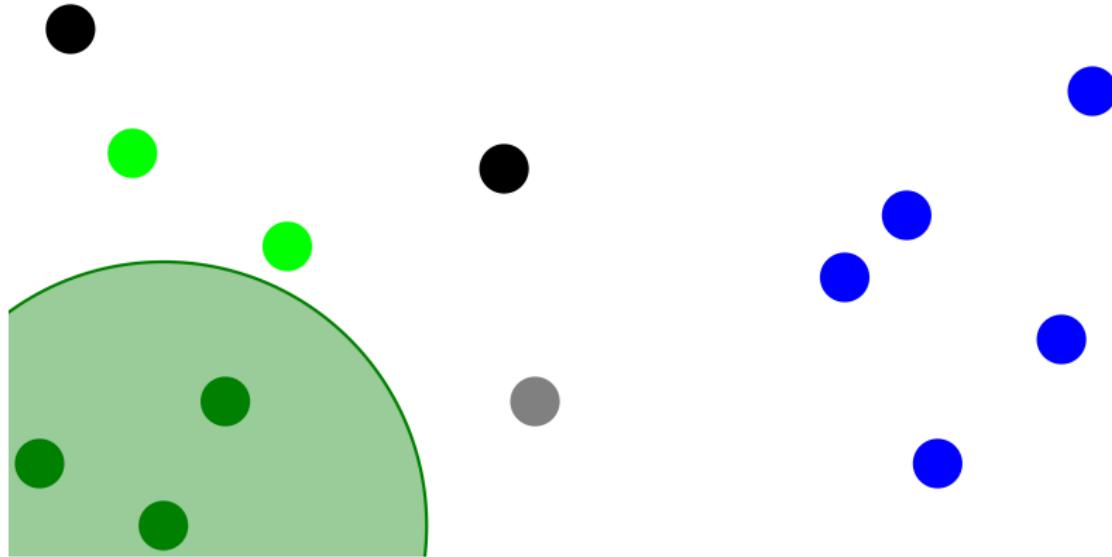
Principe

Terminaison d'un cluster: exemple



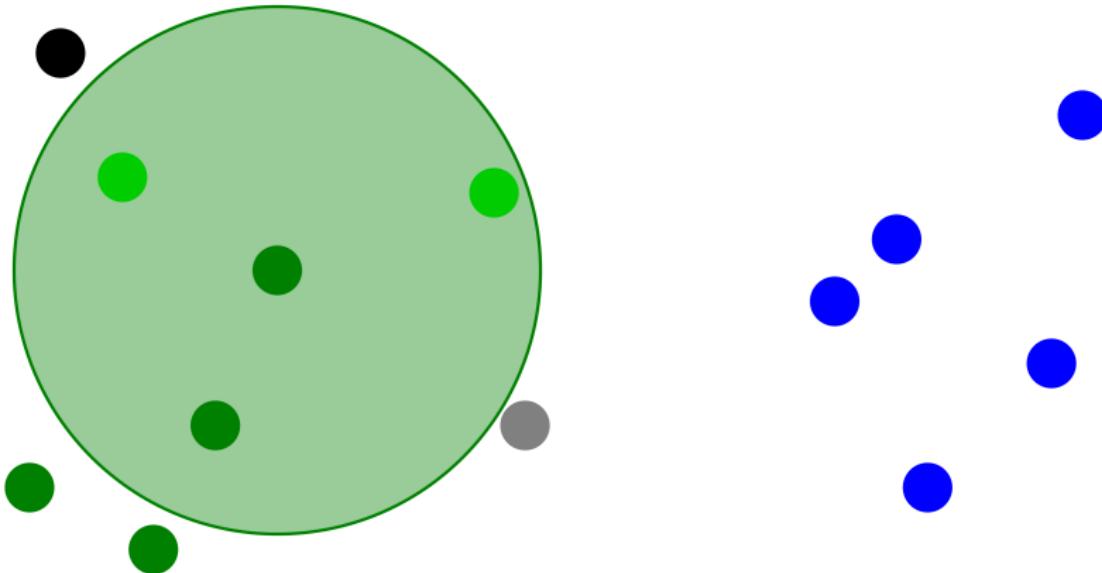
Principe

Terminaison d'un cluster: exemple



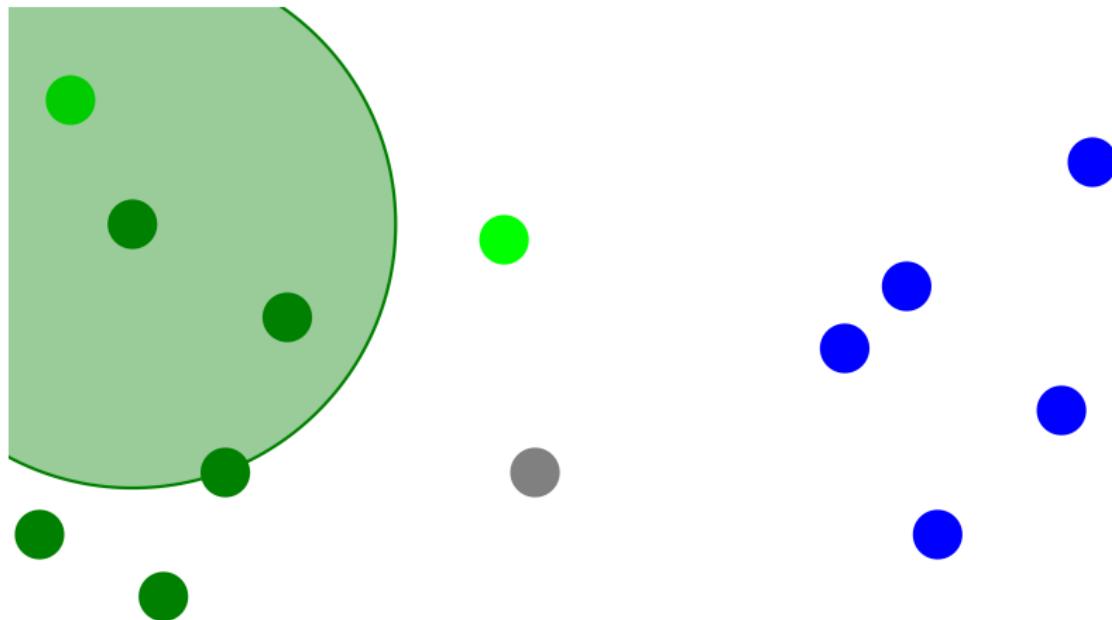
Principe

Terminaison d'un cluster: exemple



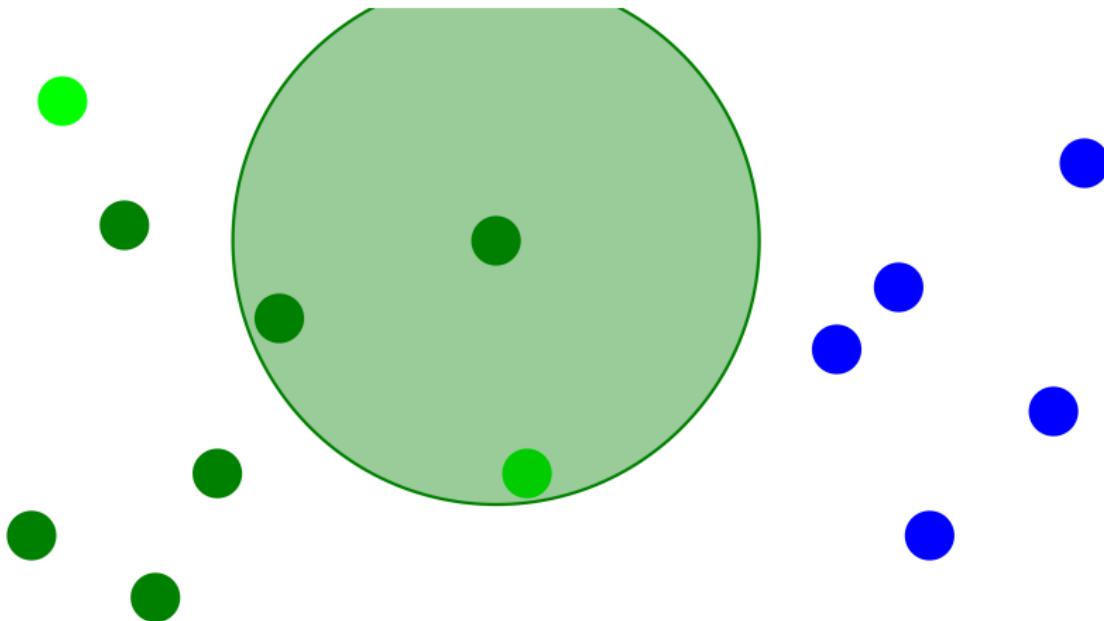
Principe

Terminaison d'un cluster: exemple



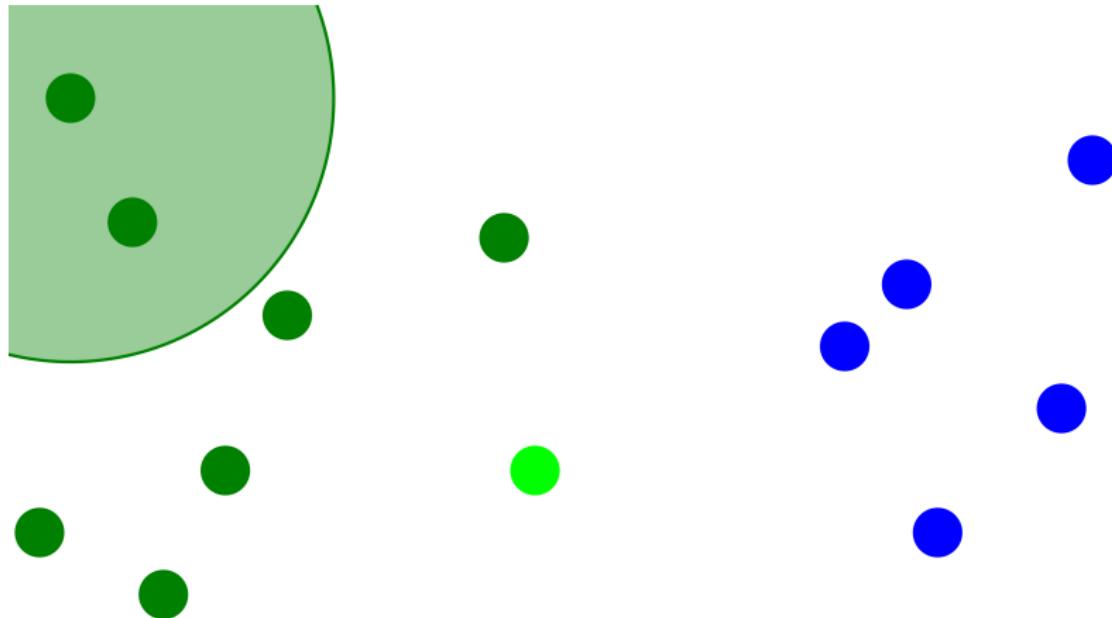
Principe

Terminaison d'un cluster: exemple



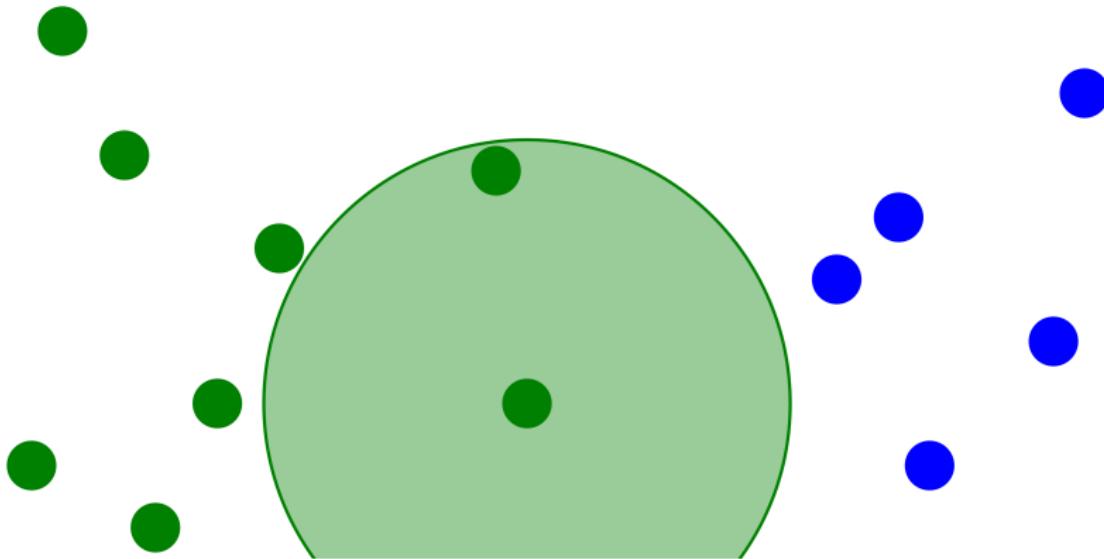
Principe

Terminaison d'un cluster: exemple



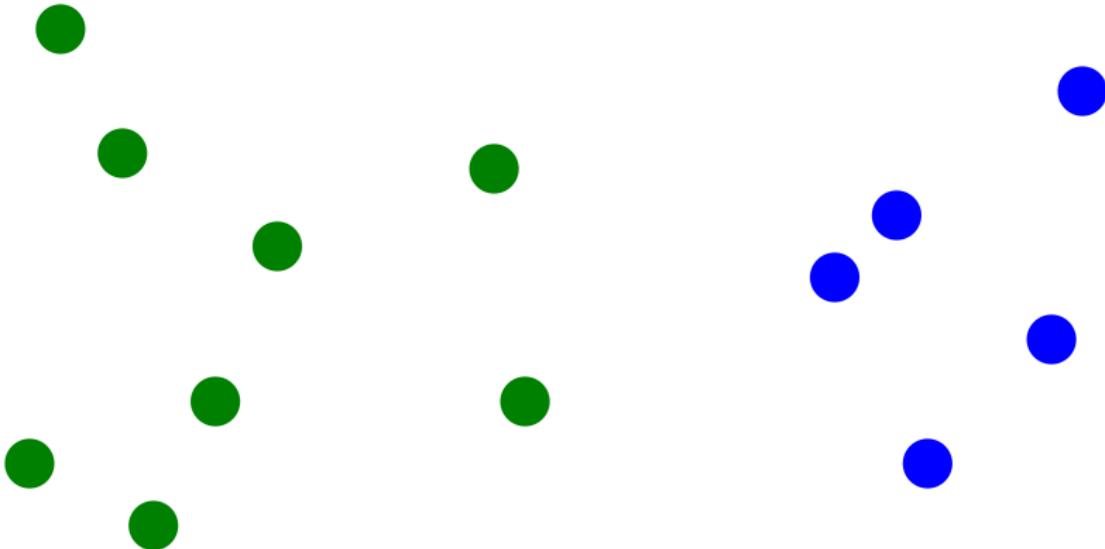
Principe

Terminaison d'un cluster: exemple



Principe

Terminaison d'un cluster: exemple



Implémentation

Avec des listes chaînées

Stocker les informations

Les points et les clusters sont des structures, manipulées par des pointeurs.

Implémentation

Avec des listes chaînées

Stocker les informations

Les points et les clusters sont des structures, manipulées par des pointeurs.

Manipuler les informations

- Tableau: accès direct mais la taille est fixe.
- Liste chaînée: taille variable mais l'accès est indirect.

Implémentation

Avec des listes chaînées

Stocker les informations

Les points et les clusters sont des structures, manipulées par des pointeurs.

Manipuler les informations

- Tableau: accès direct mais la taille est fixe.
- Liste chaînée: taille variable mais l'accès est indirect.

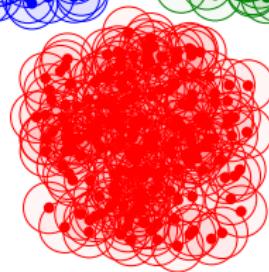
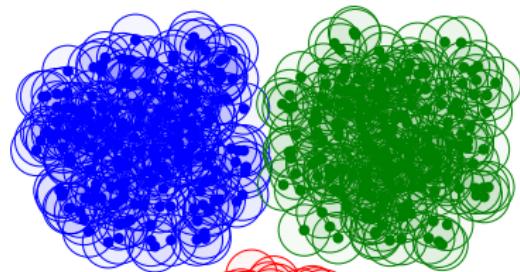
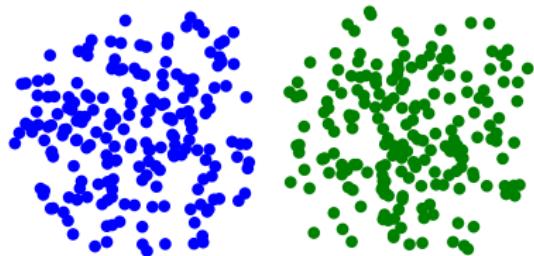
Choix pour DBSCAN

Parcourt régulier et linéaire les données.

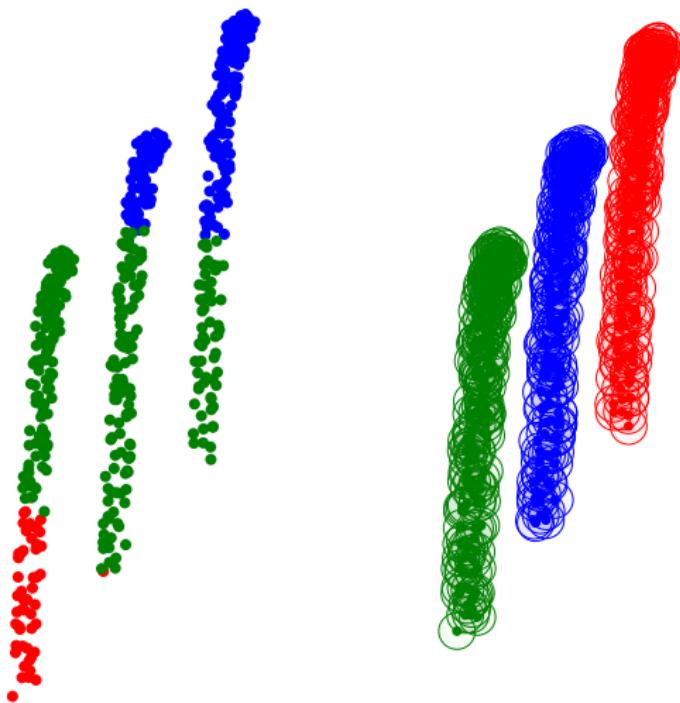
Choix: listes chaînées.

Bonus: réduction dynamique du nombre de points à étudier.

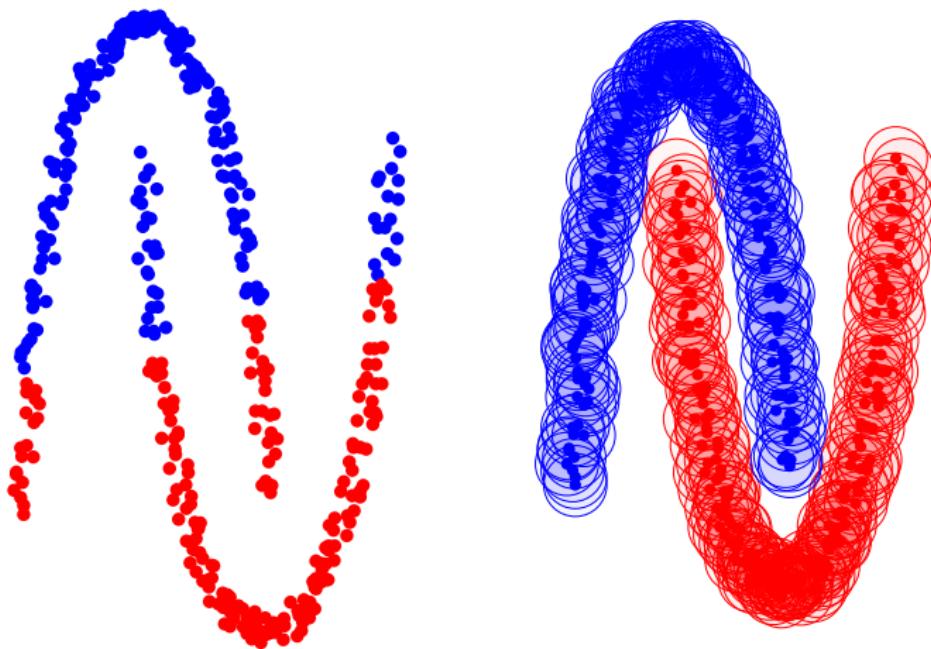
Résultats



Résultats



Résultats



Résultats

