

المحتوى

1. مراجعة خوارزميات تعلم الآلة

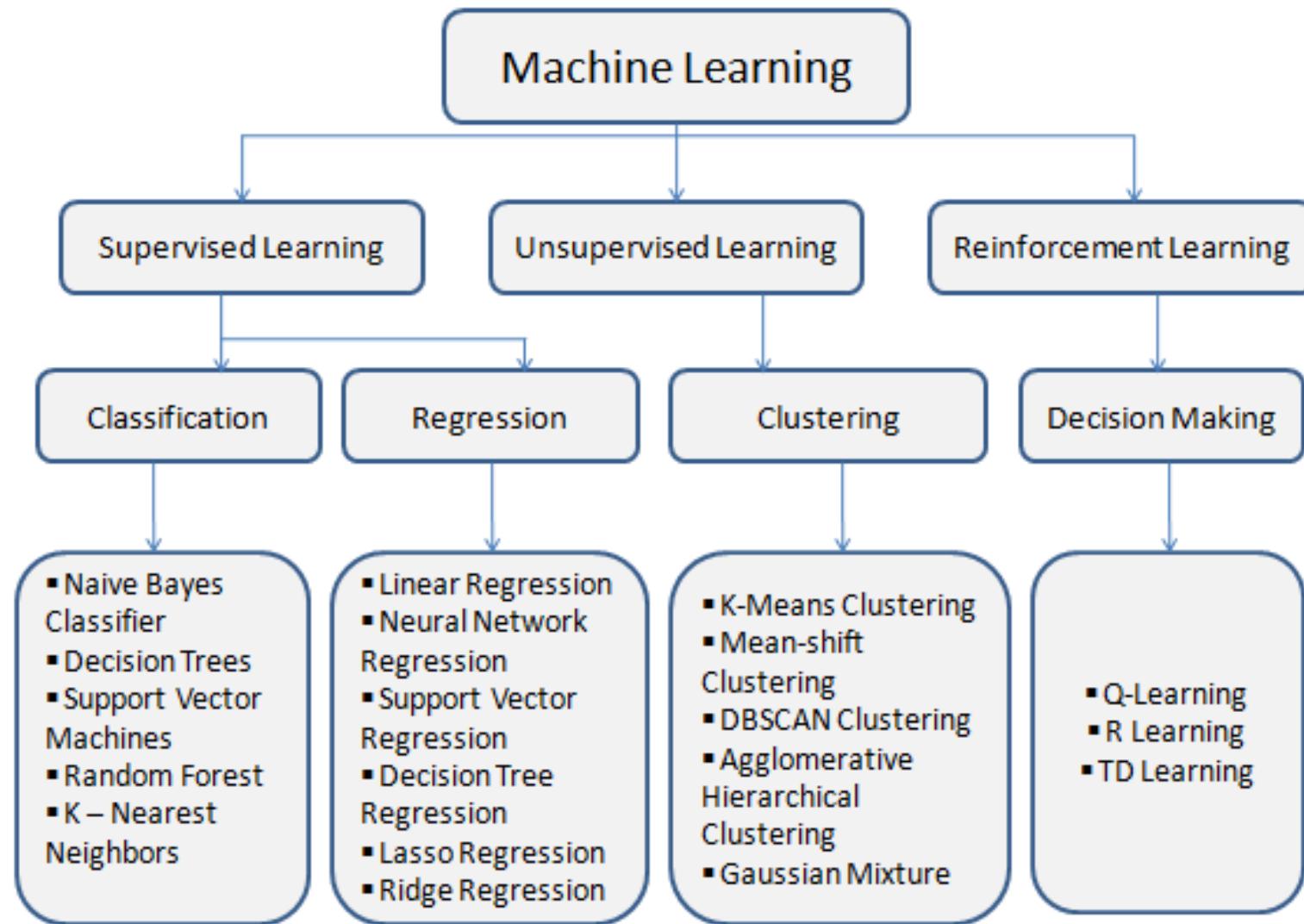
2. مراجعة خوارزمية K-Means Clustering

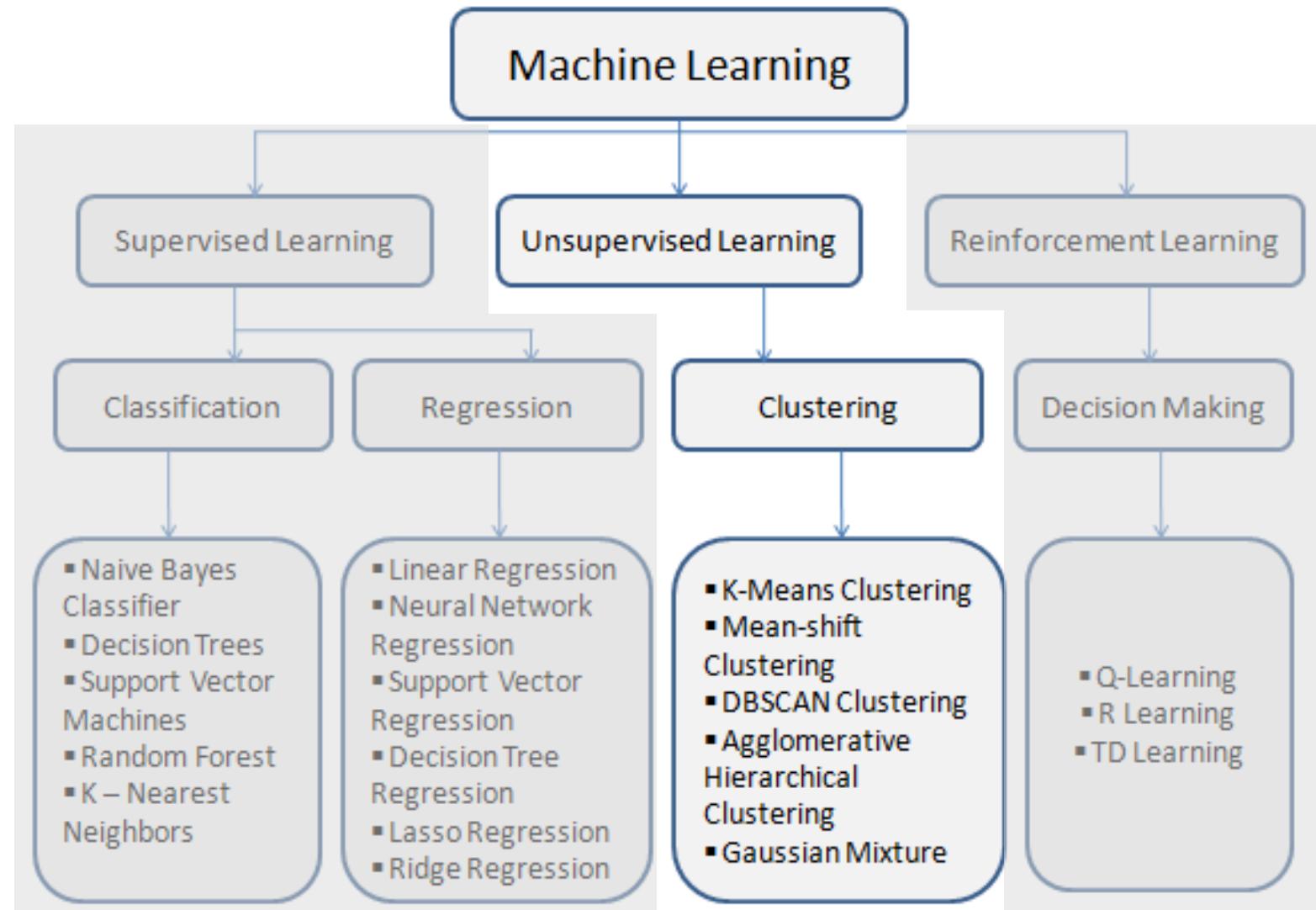
3. ليش خوارزمية DBScan موجودة؟

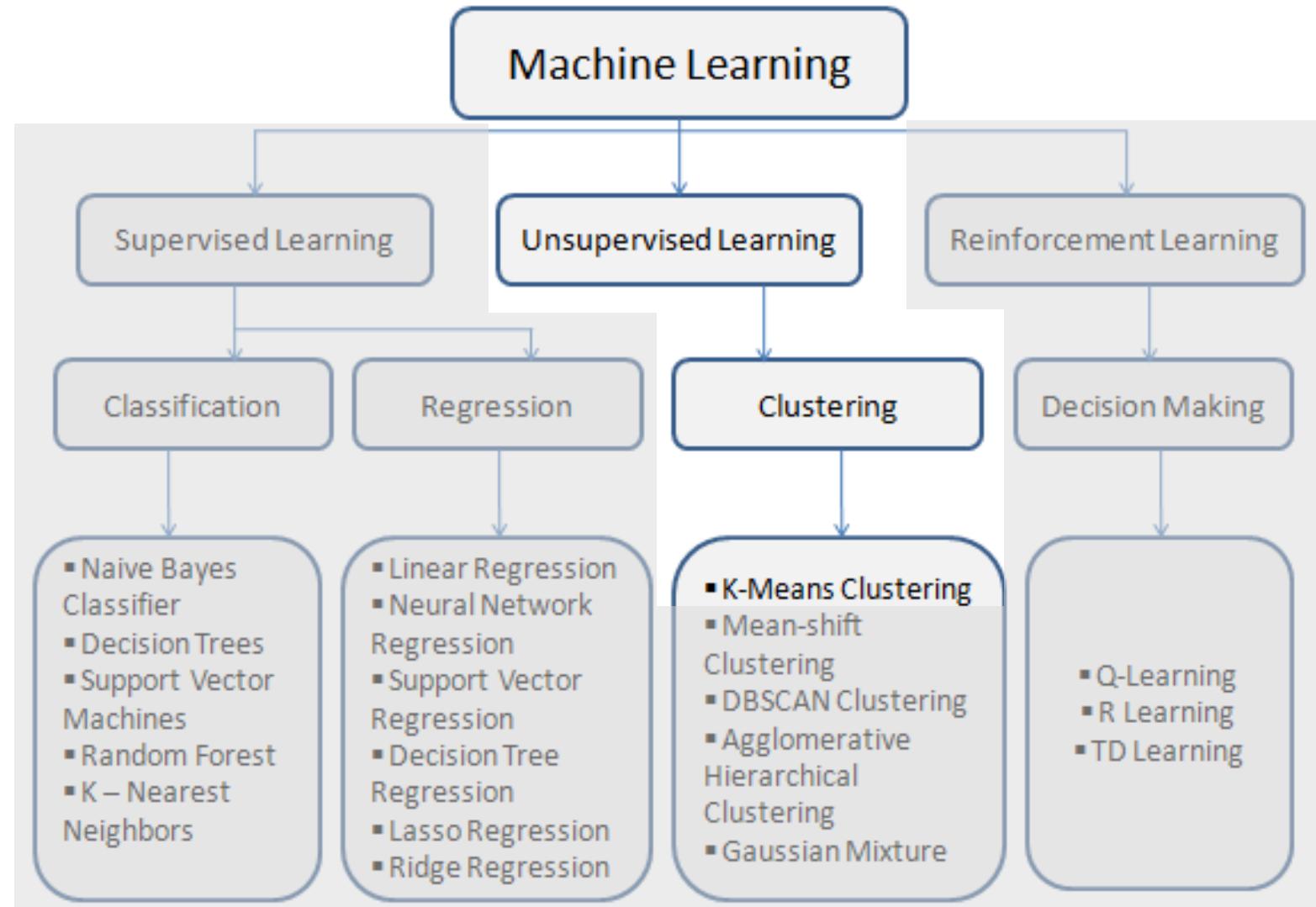
4. طريقة عمل خوارزمية DBScan

5. طريقة عمل خوارزمية DBScan على Python

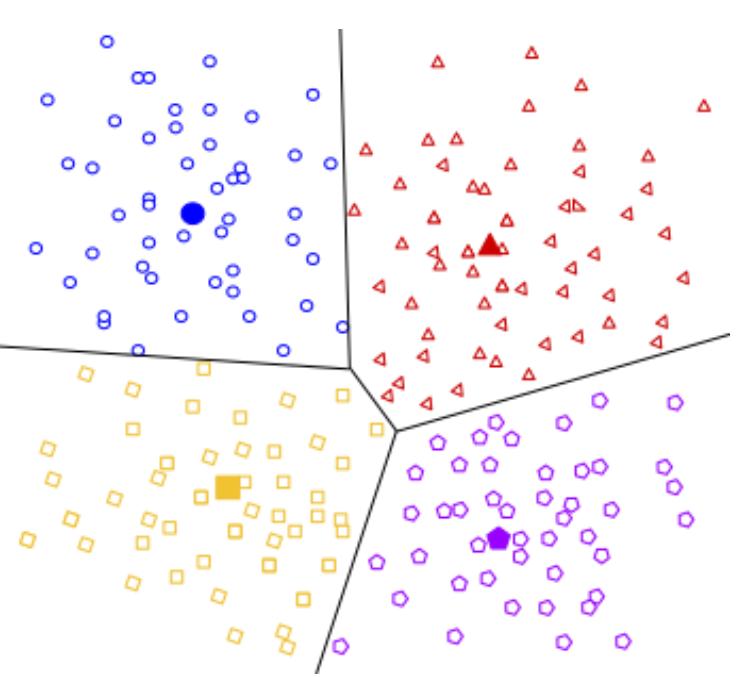
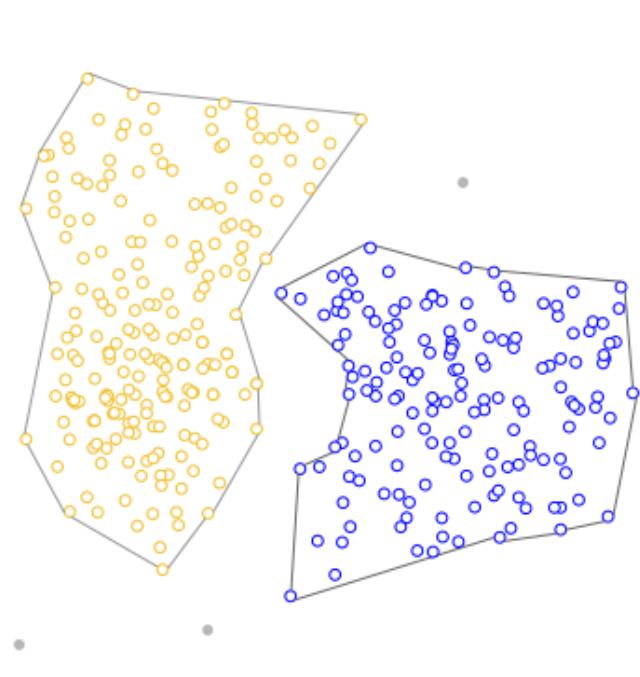
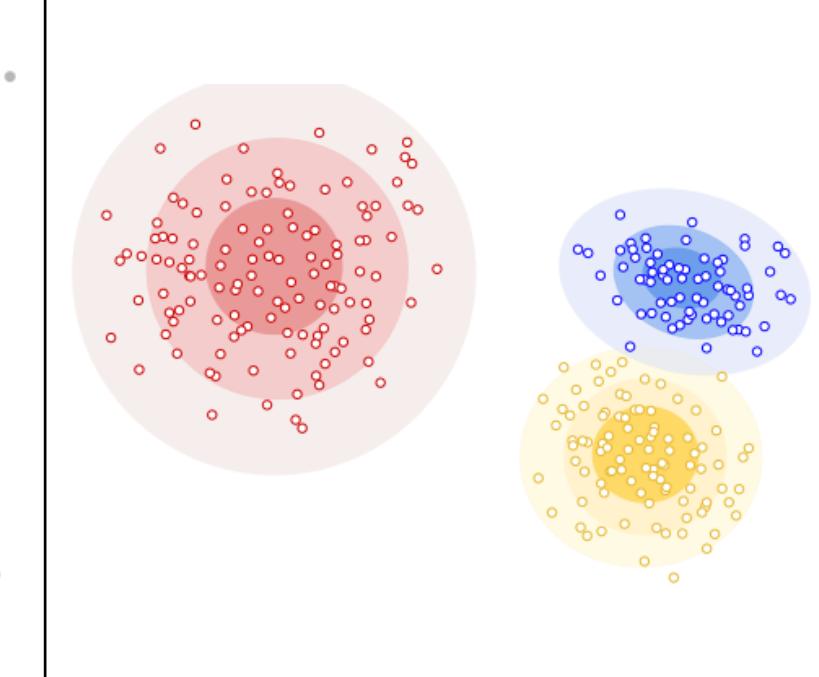
5. اختبار بسيط







أنواع الـ Clustering

Centroid-based Clustering	Density-based Clustering	Distribution-based Clustering
		

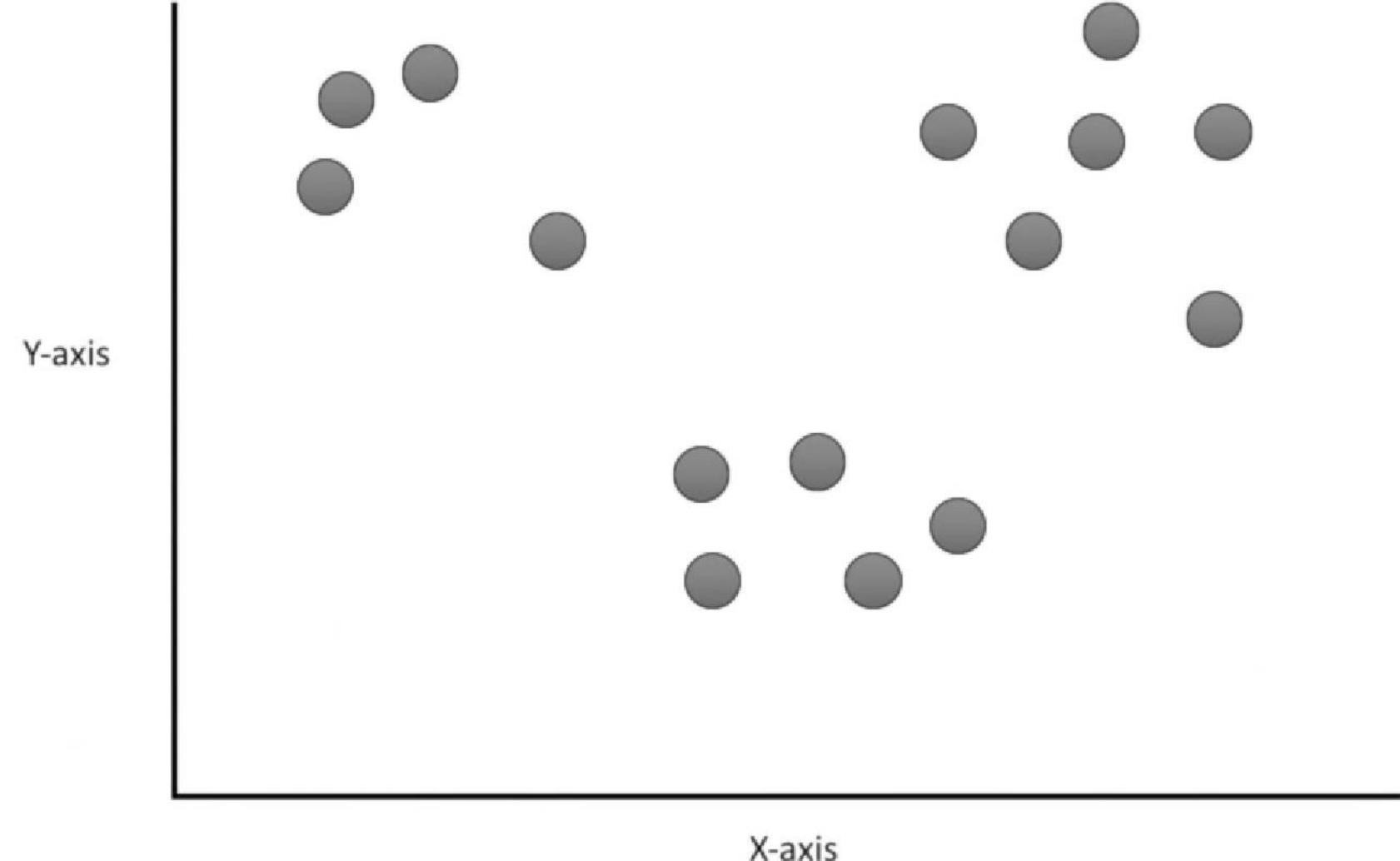


مراجعة خوارزمية K-Means Clustering

[Centroid-based Clustering]

خطوات العمل :

1. تحديد عدد (K). Clusters
2. اختيار قيم عشوائية بعدد K .
3. خطوات نكررها:
 - (1) نمثي على كل القيم ونشوف وش أقرب نقطة Centroid لها.
 - (2) حساب ال Mean لكل Cluster
4. نوقف خطوات التكرار إذا ما فيه قيمة جديدة.

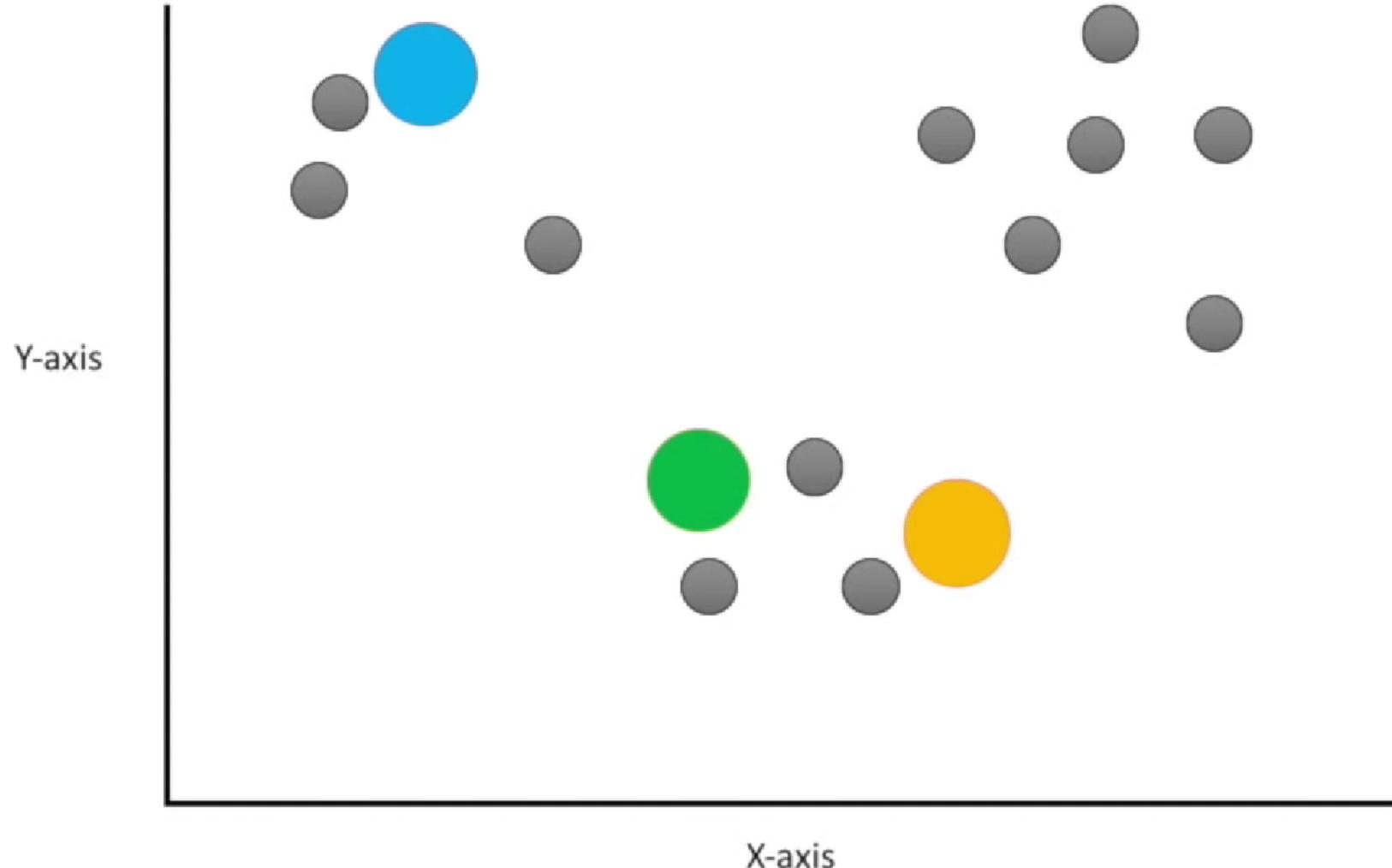




مراجعة خوارزمية K-Means Clustering

[Centroid-based Clustering]

نحدد $K = 3$ ، ونختار 3 قيم عشوائية

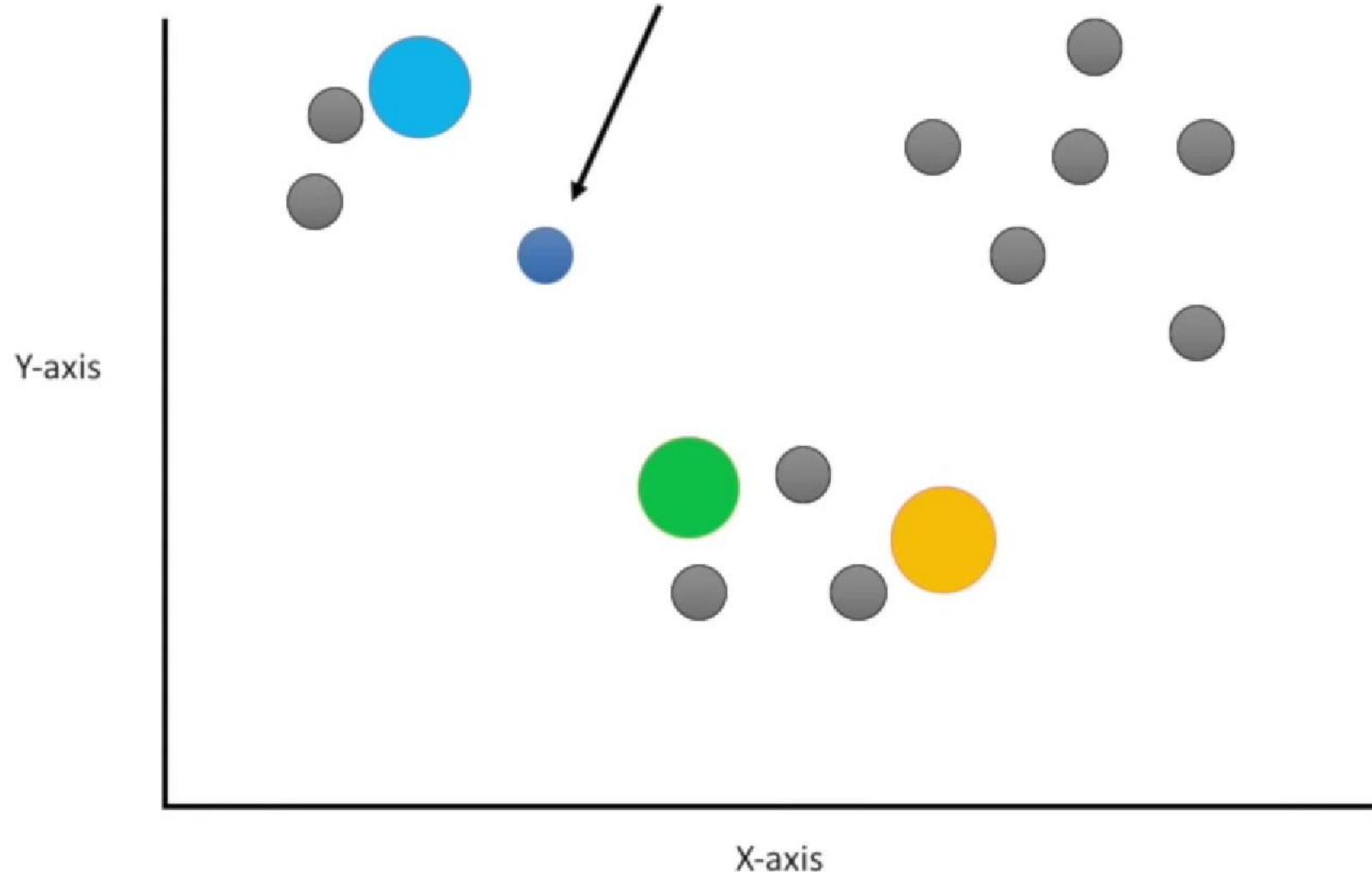




مراجعة خوارزمية K-Means Clustering

[Centroid-based Clustering]

نمشي على كل القيم ونحدد اقرب Centroid لها ونضيفه الى مجموعته

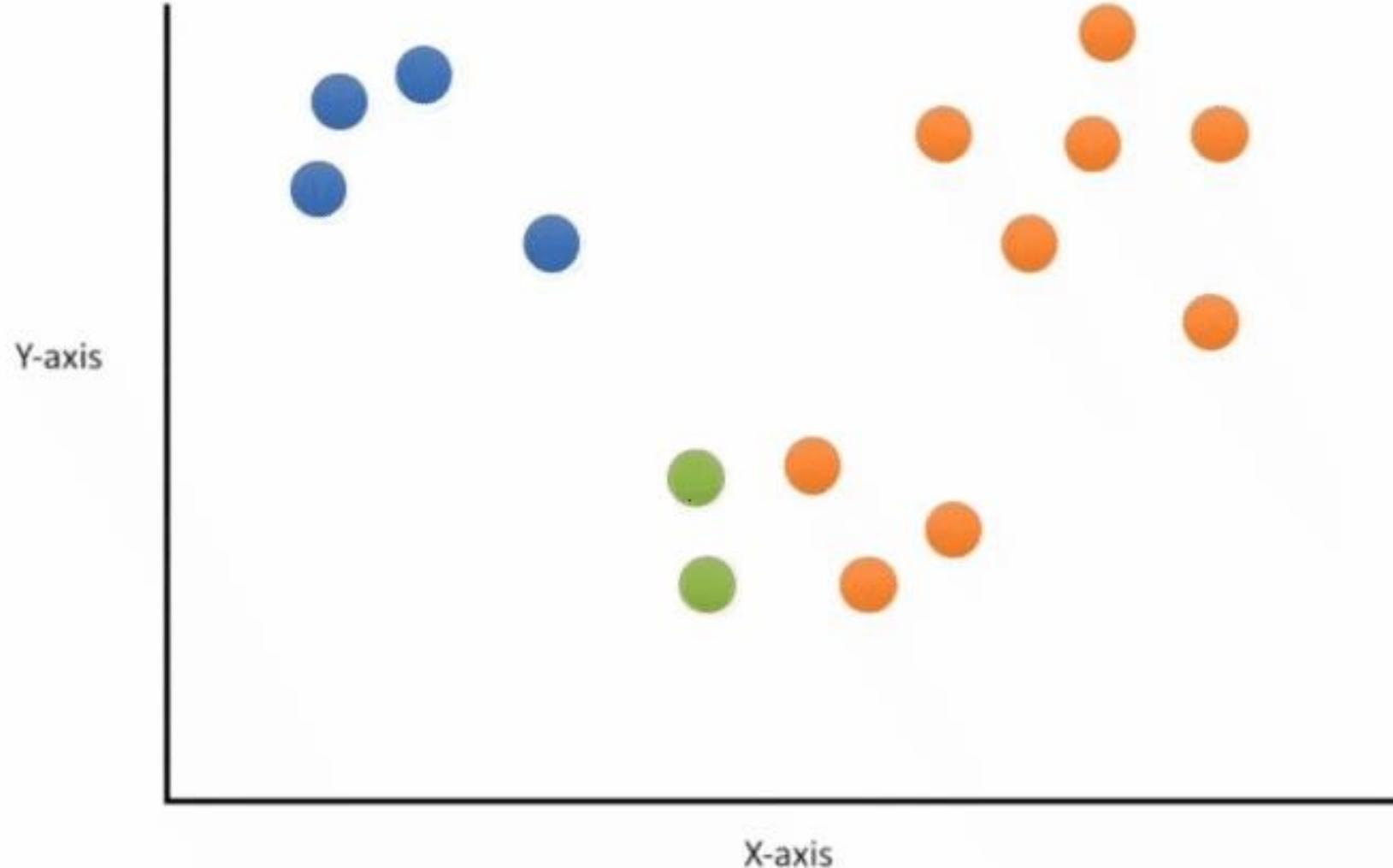




مراجعة خوارزمية K-Means Clustering

[Centroid-based Clustering]

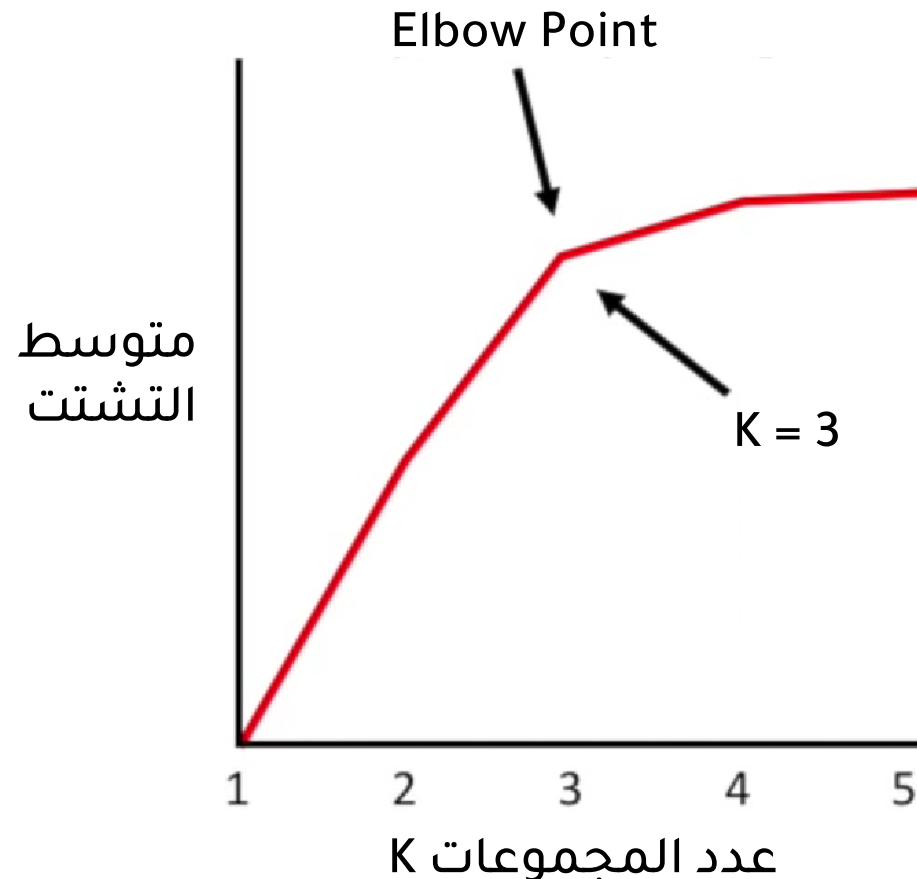
نحسب ال Mean ونكرر الخطوة السابقة ونوقف لما يكون مافيه تغييرات جديدة

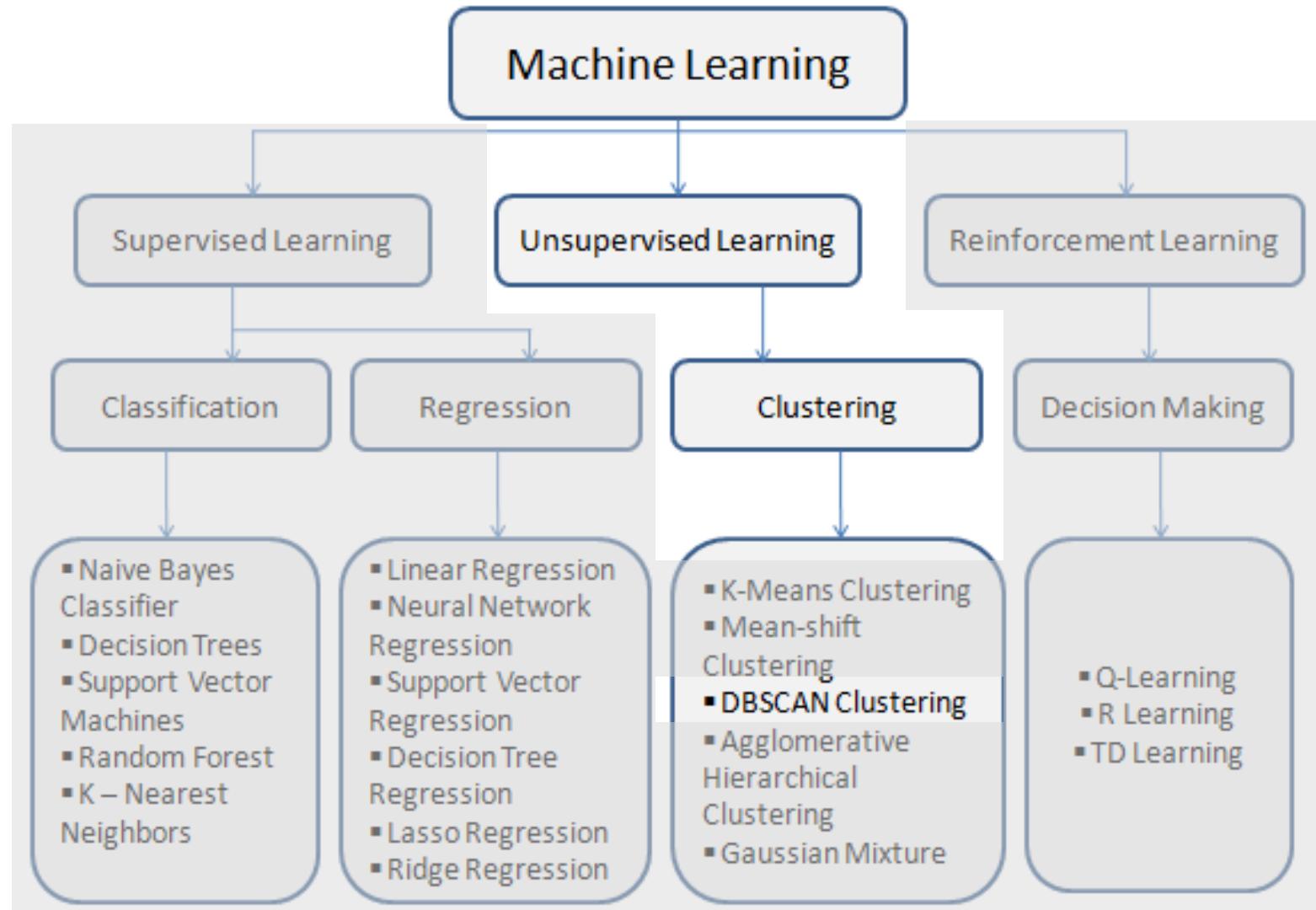




طريقة أفضل لاختيار قيمة الـ K

“Elbow Plot”



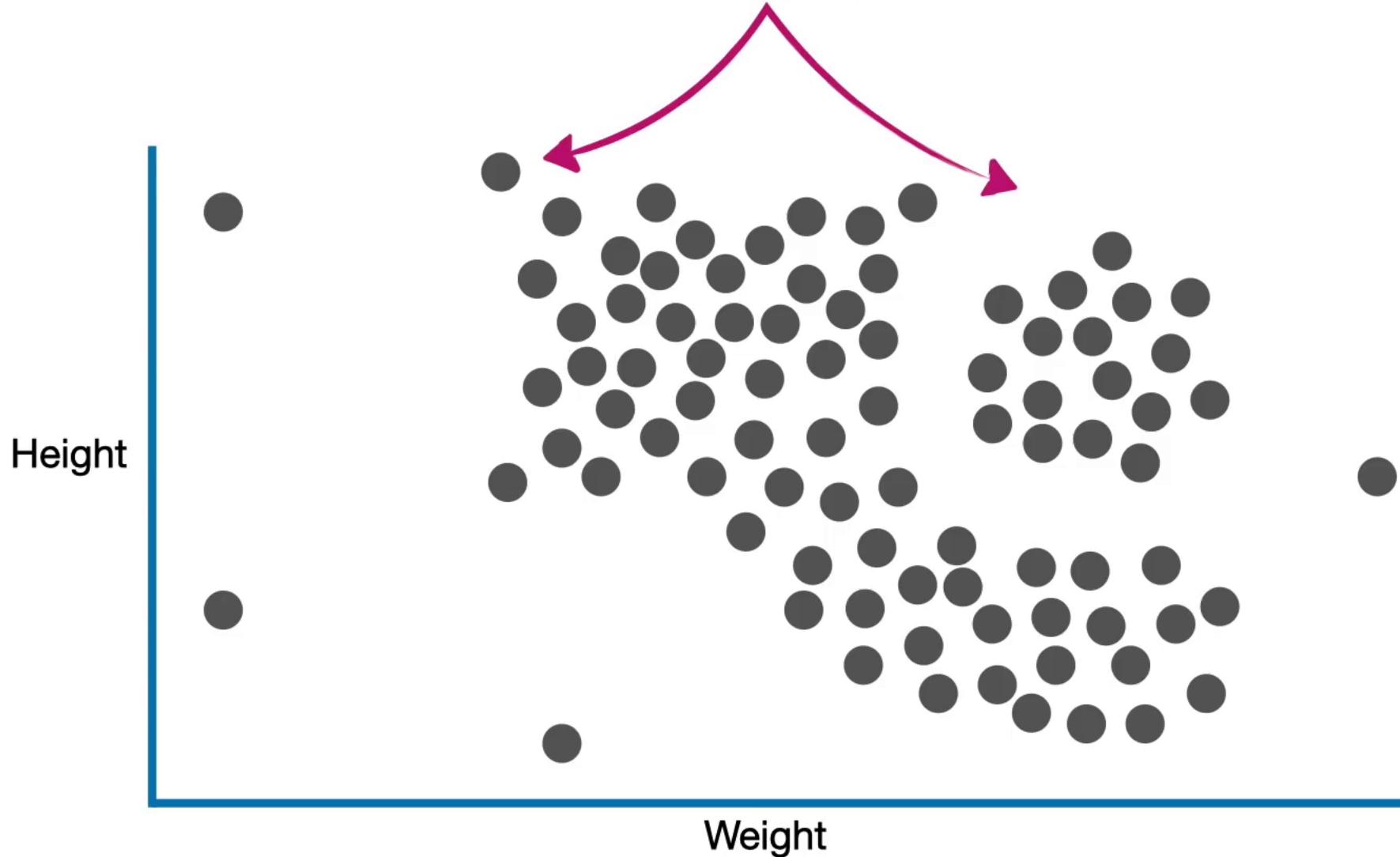


جمعنا بيانات لأشخاص بطول والوزن لكل شخص

	Weight	Height
Person 1	56	150
Person 2	62	170
Person 3	71	168
...



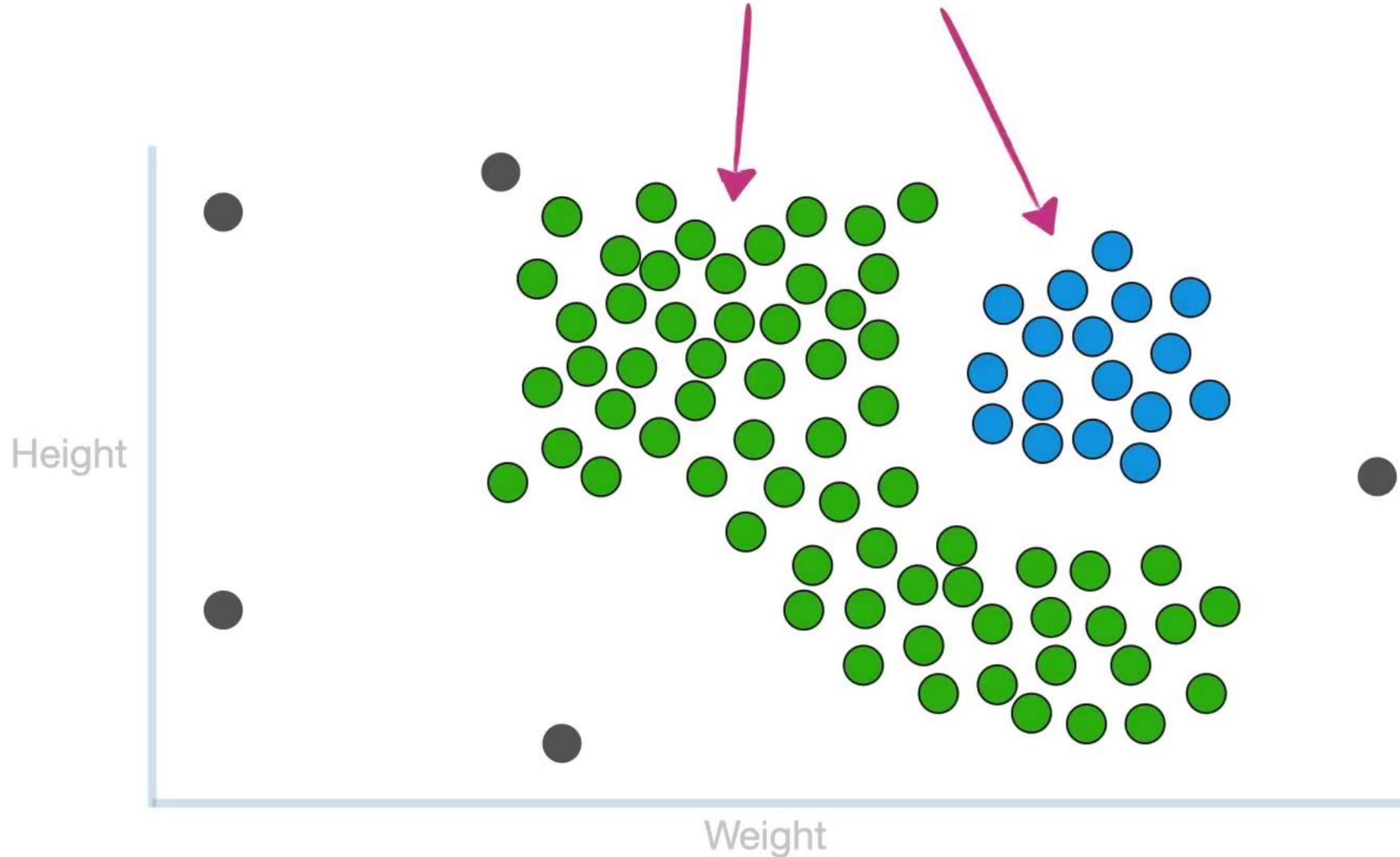
مثلنا بيانات الطول والوزن برسم
scatter plot





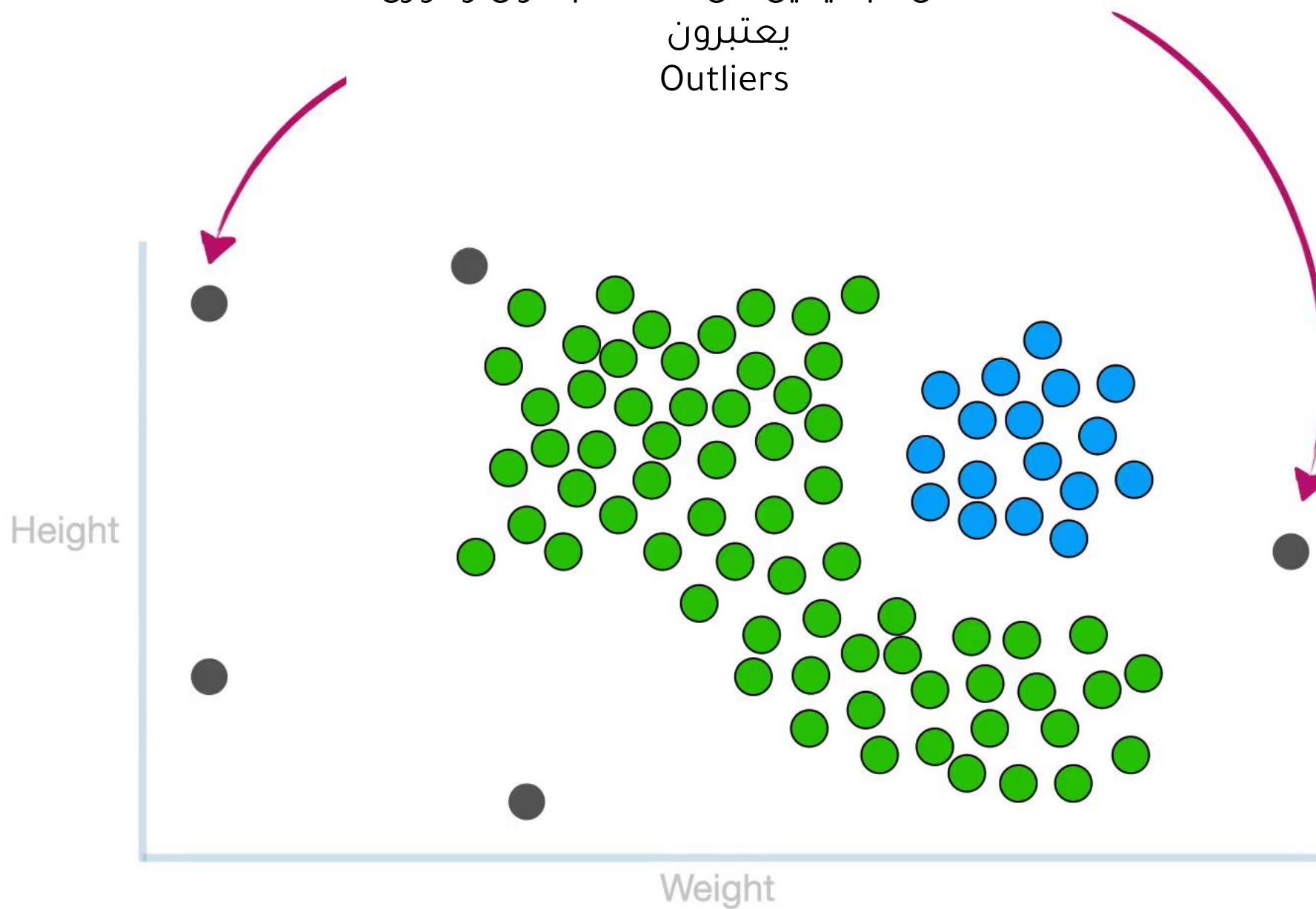
بمجرد النظر، من الواضح وجود

Two clusters



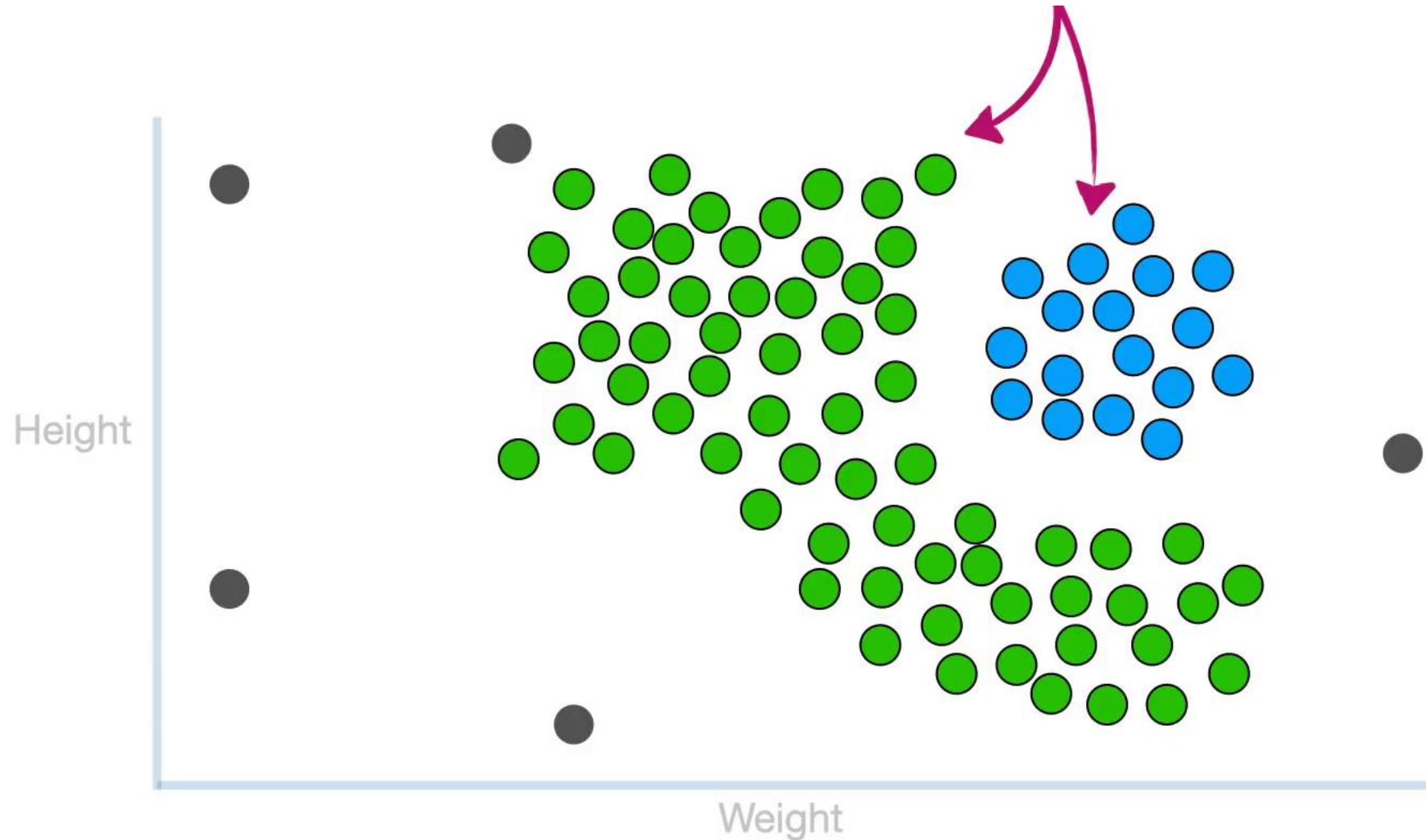
الأشخاص البعيدين عن العامة بطول وزن

يعتبرون
Outliers



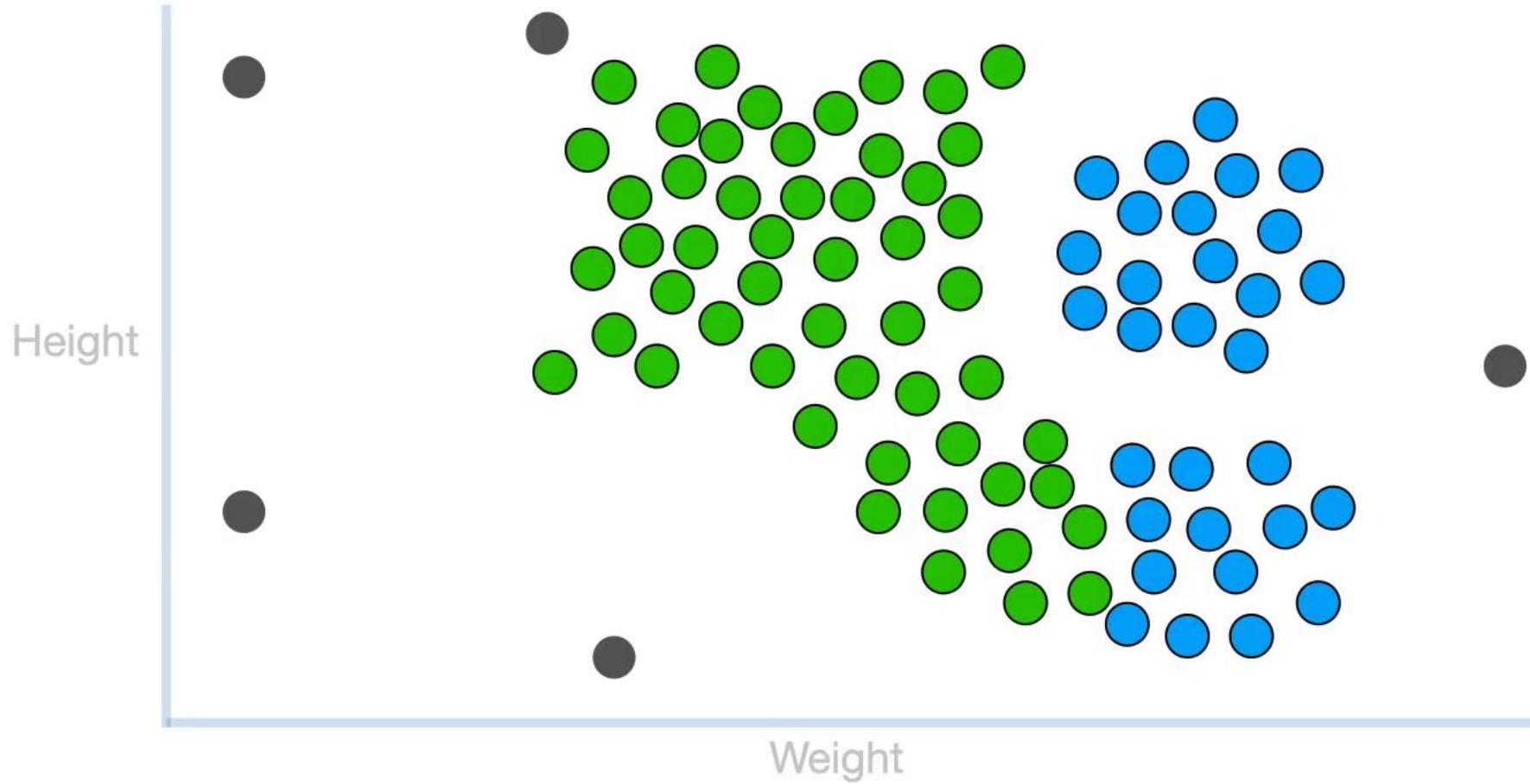


في الطرق التقليدية لـ Clustering مثل الـ K-Means راح نواجه مشكلة في التقسيم



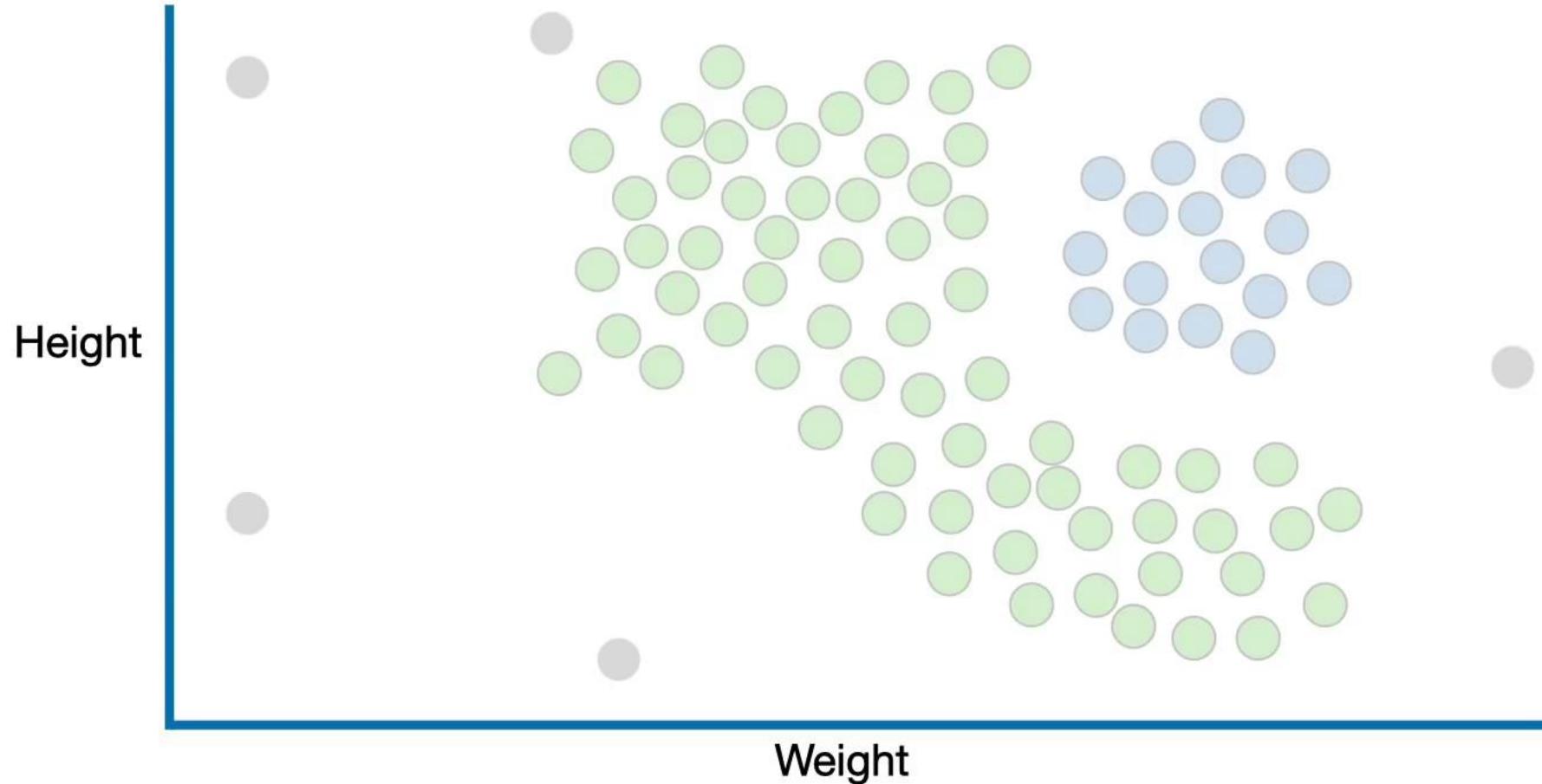


شكل ١٢ Clustering in K-Means



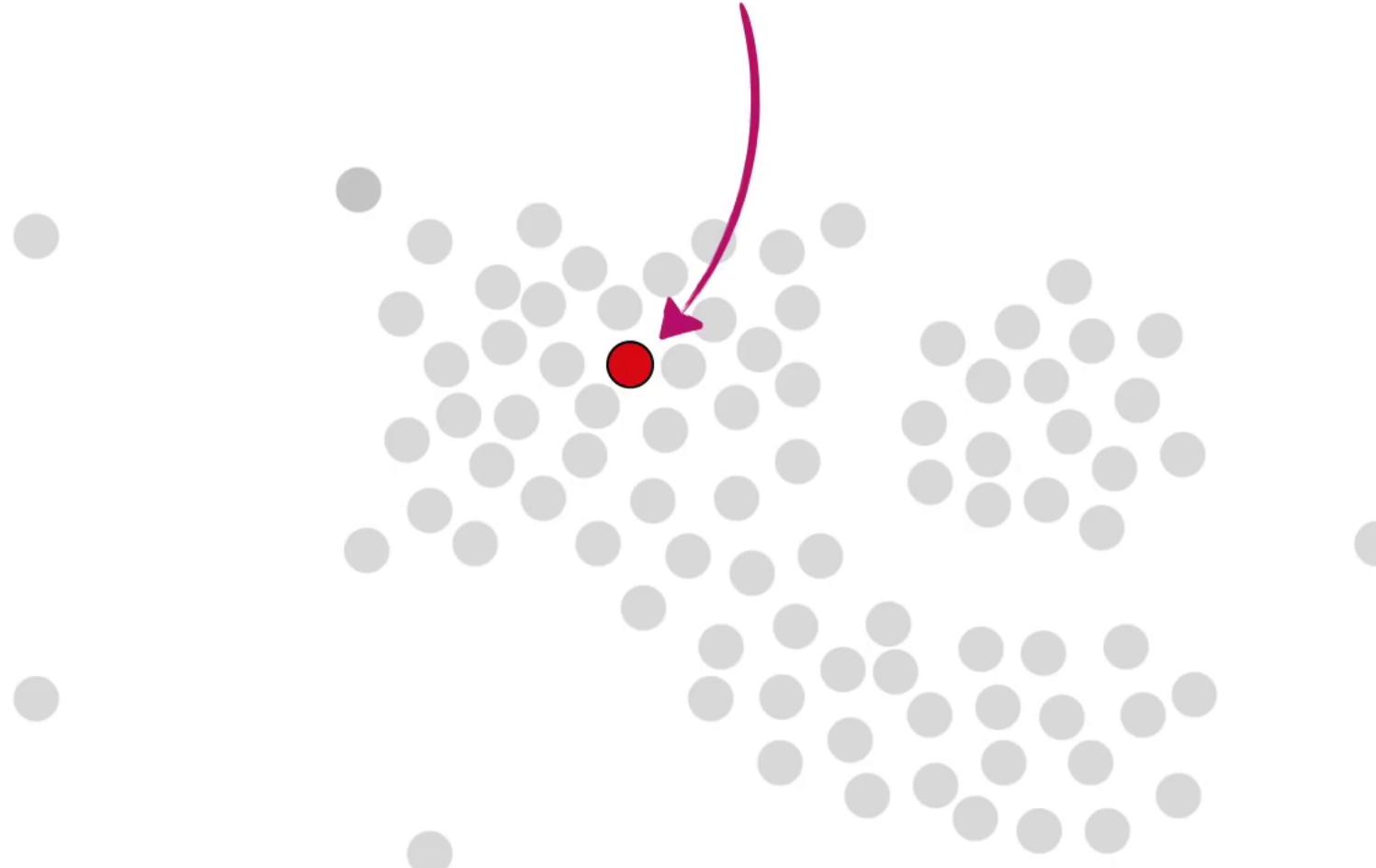


من هذه المشكلة أتت خوارزمية dbSCAN لحل
هذه المشكلة



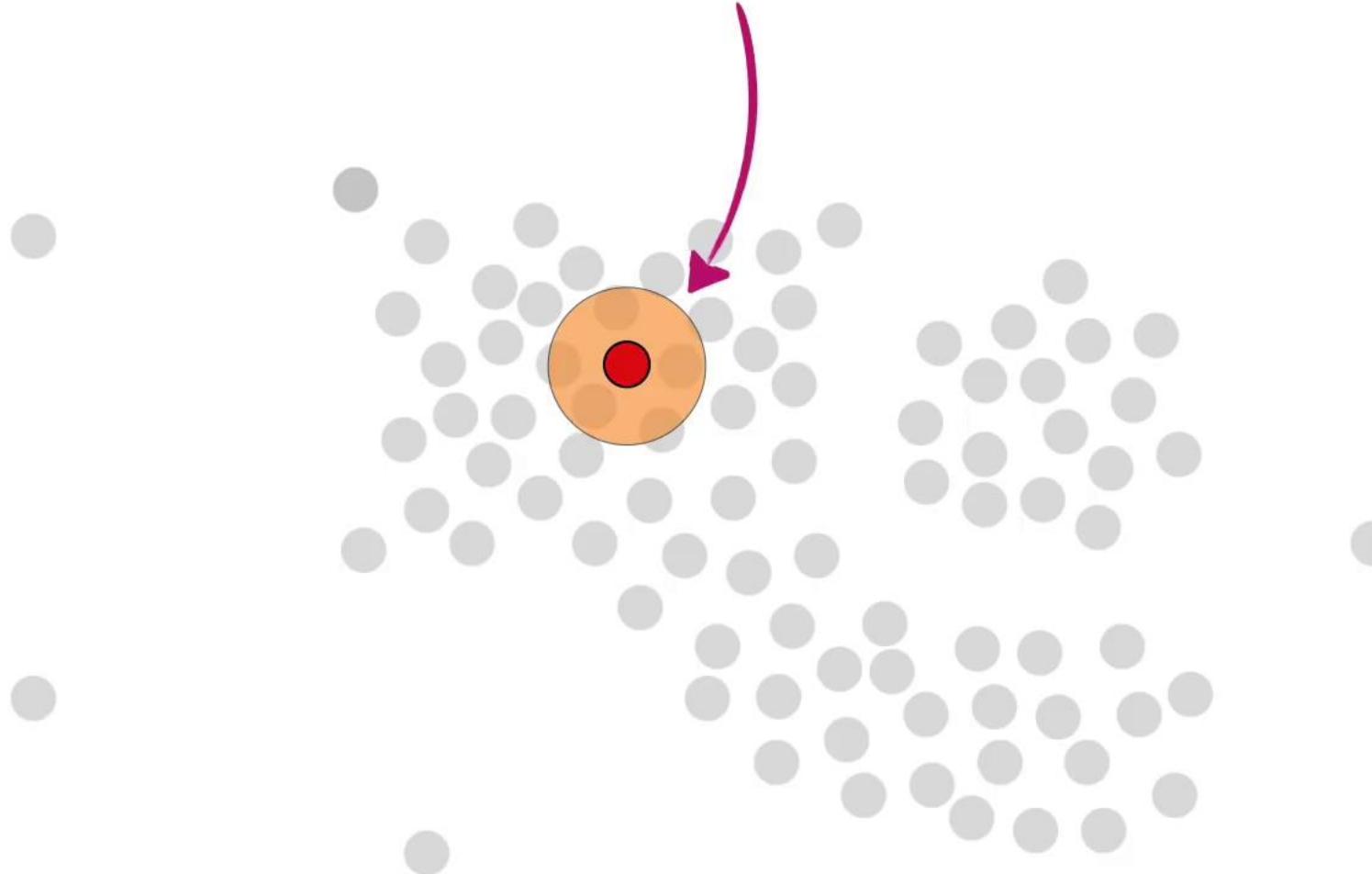


بشكل عشوائي نختار ببدأ عملية الـ
Clustering



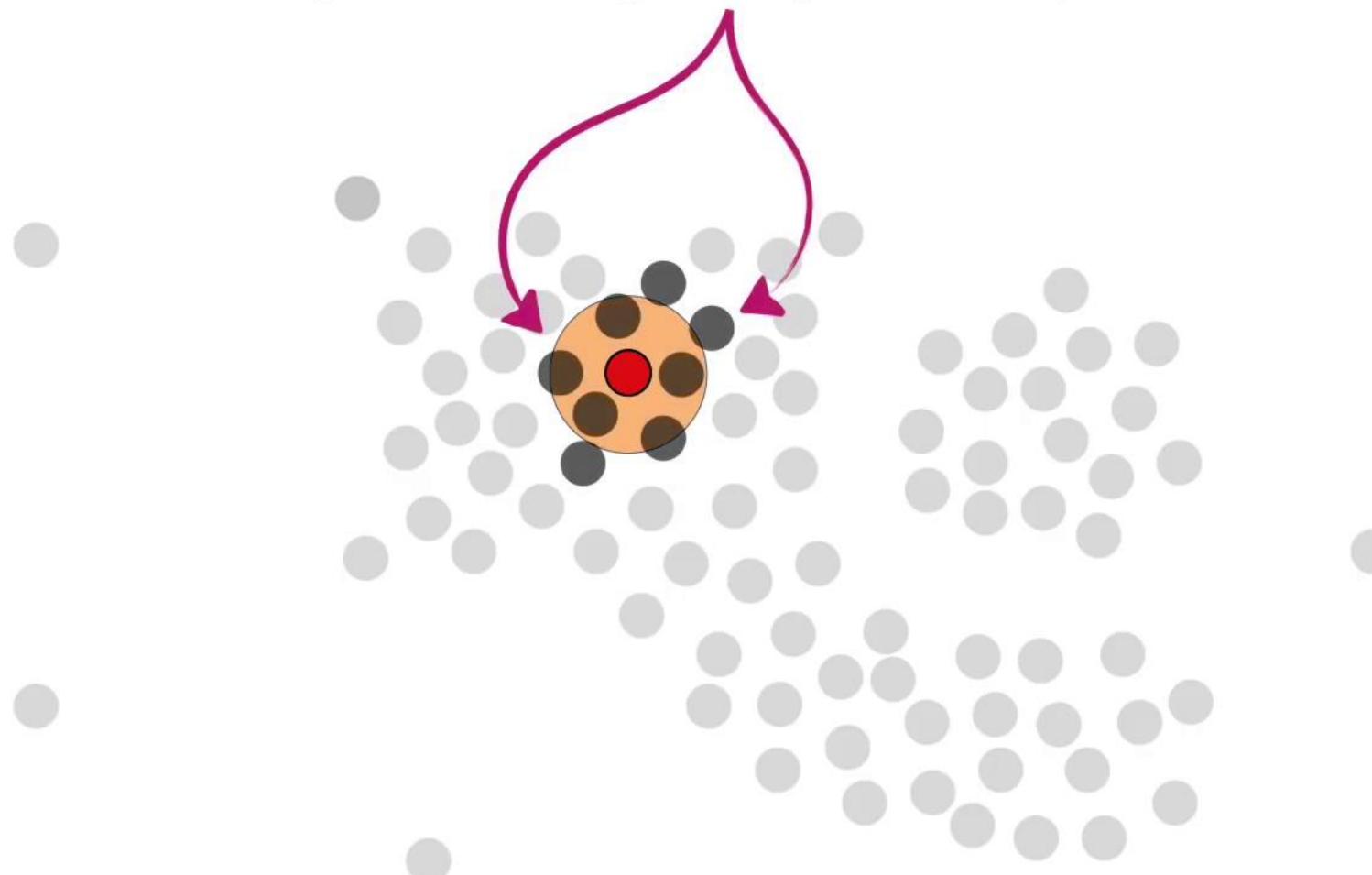


رسم دائرة حول النقطة المختارة **المشار**
لها باللون البرتقالي



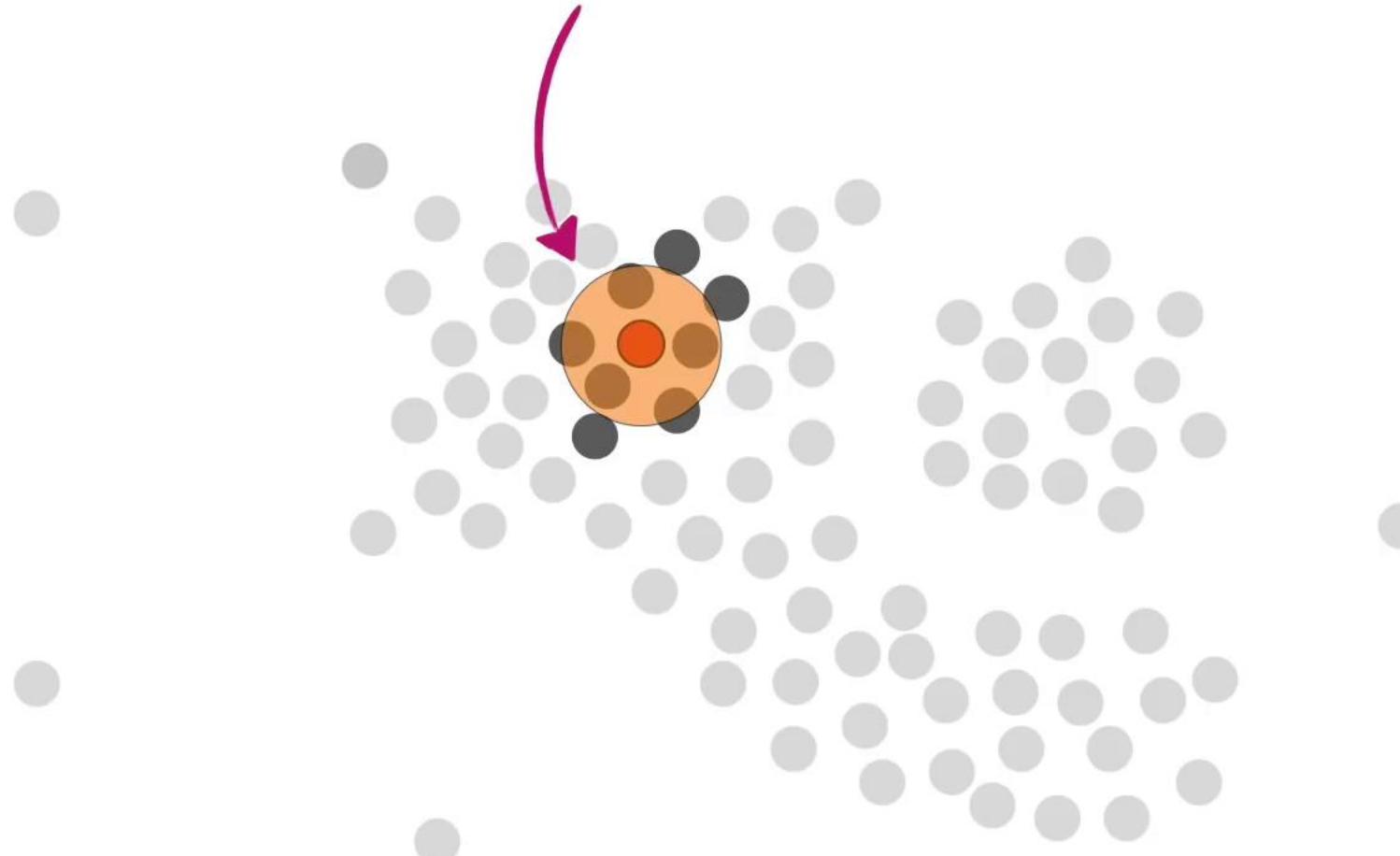


كما هو ملاحظ الدائرة البرتقالية تحيط بثمان
نقاط بشكل كامل/جزئي





ملاحظة: حجم قطر الدائرة تكون معرف من قبل المستخدم للخوارزمية الـ DBScan





الفحص البصري | Visual Inspection

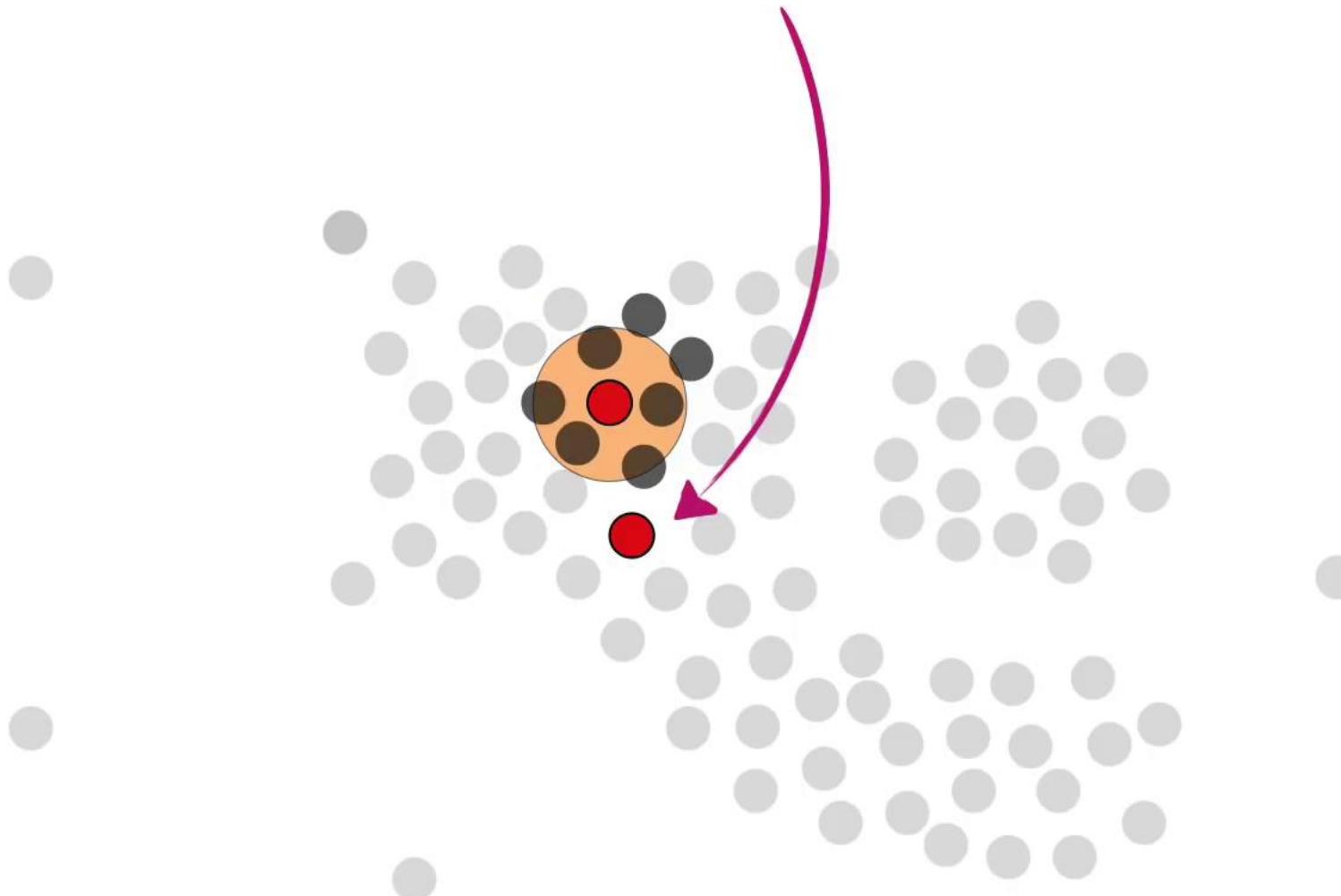
قم برسم نقاط البيانات الخاصة بك وحاول الحصول على فهم لمقاييس البيانات وكثافة النقاط. يمكن أن يمنحك ذلك فكرة تقريرية عن نطاق المسافات بين النقاط المجاورة.

توزيع المسافات | Distance Distribution

قم بحساب المسافات بين جميع نقاط البيانات وأنشئ Histogram لهذه المسافات. ابحث عن نقطة ال "knee point" في Histogram حيث تبدأ المسافات في الزيادة بسرعة. يمكن أن يمنحك ذلك فكرة عن النطاق المعقول للقطر.

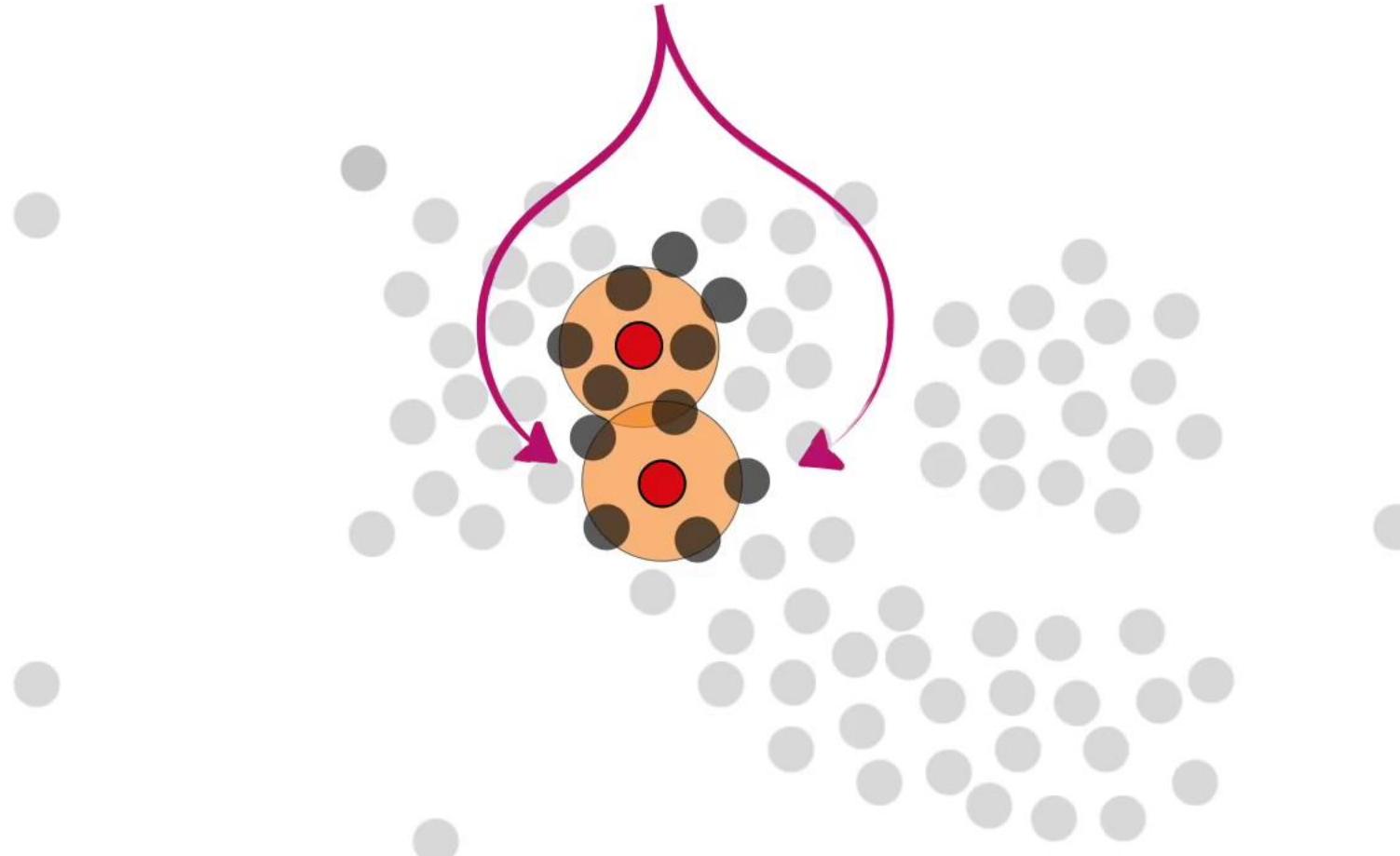


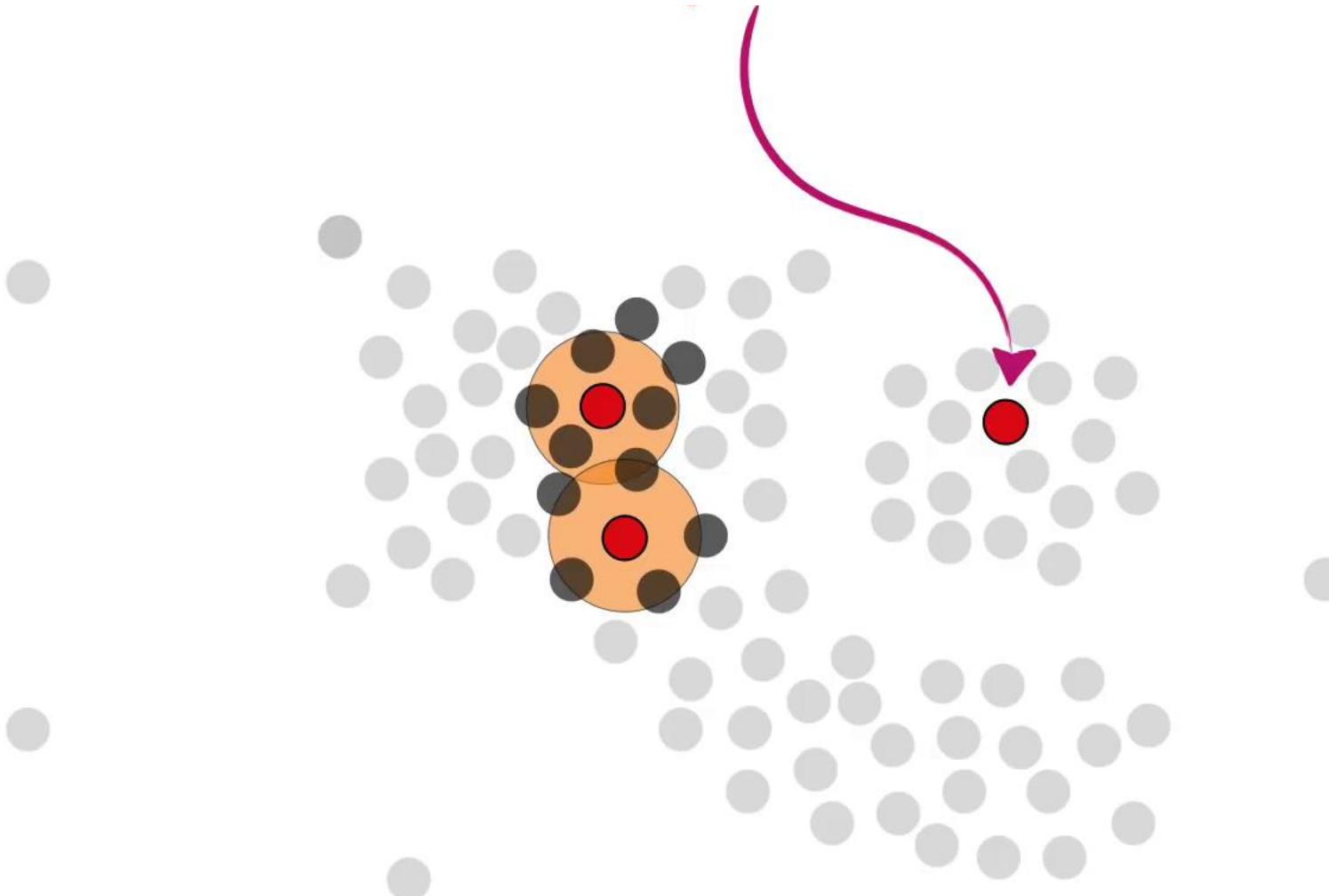
نختار نقطة مجددًا

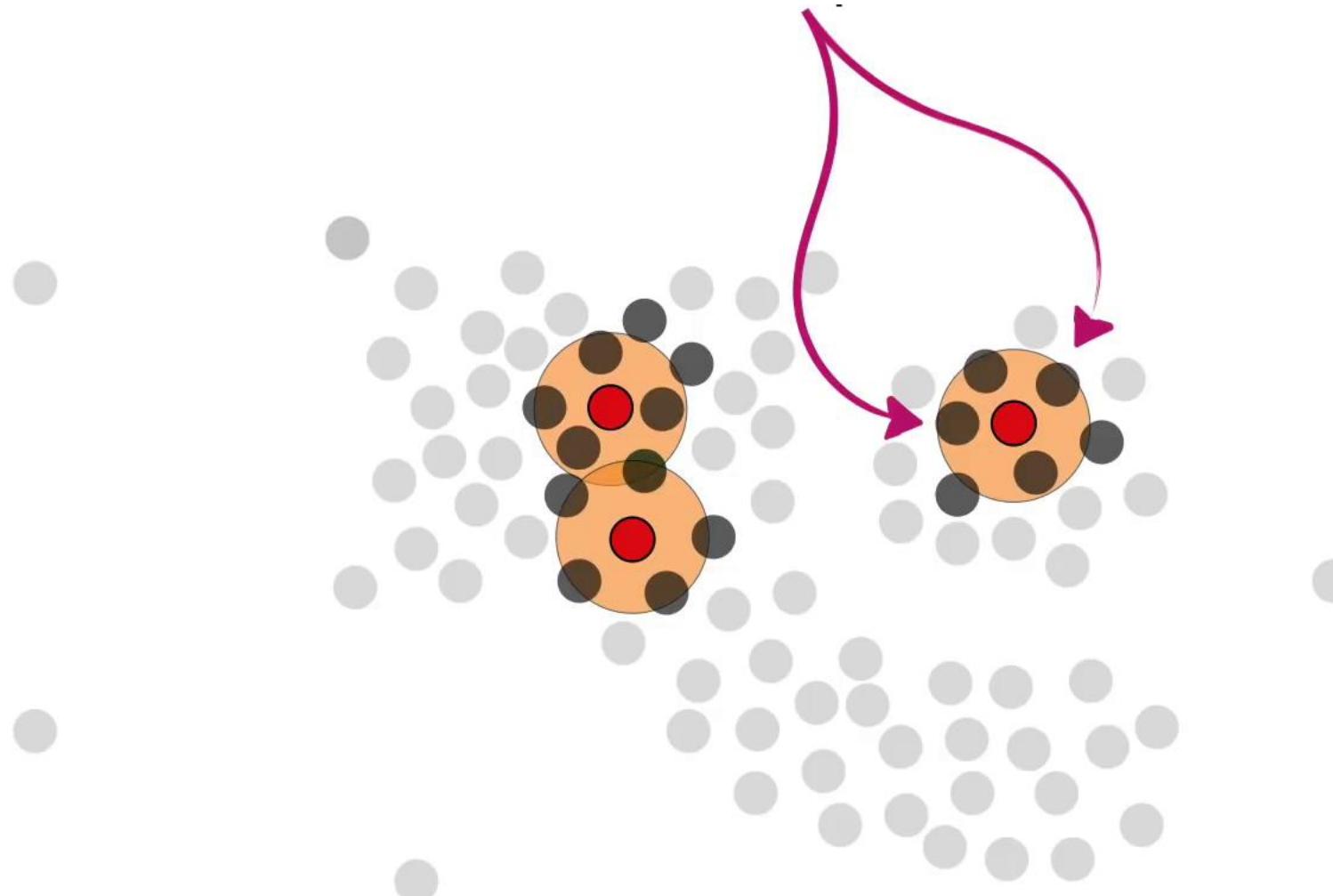


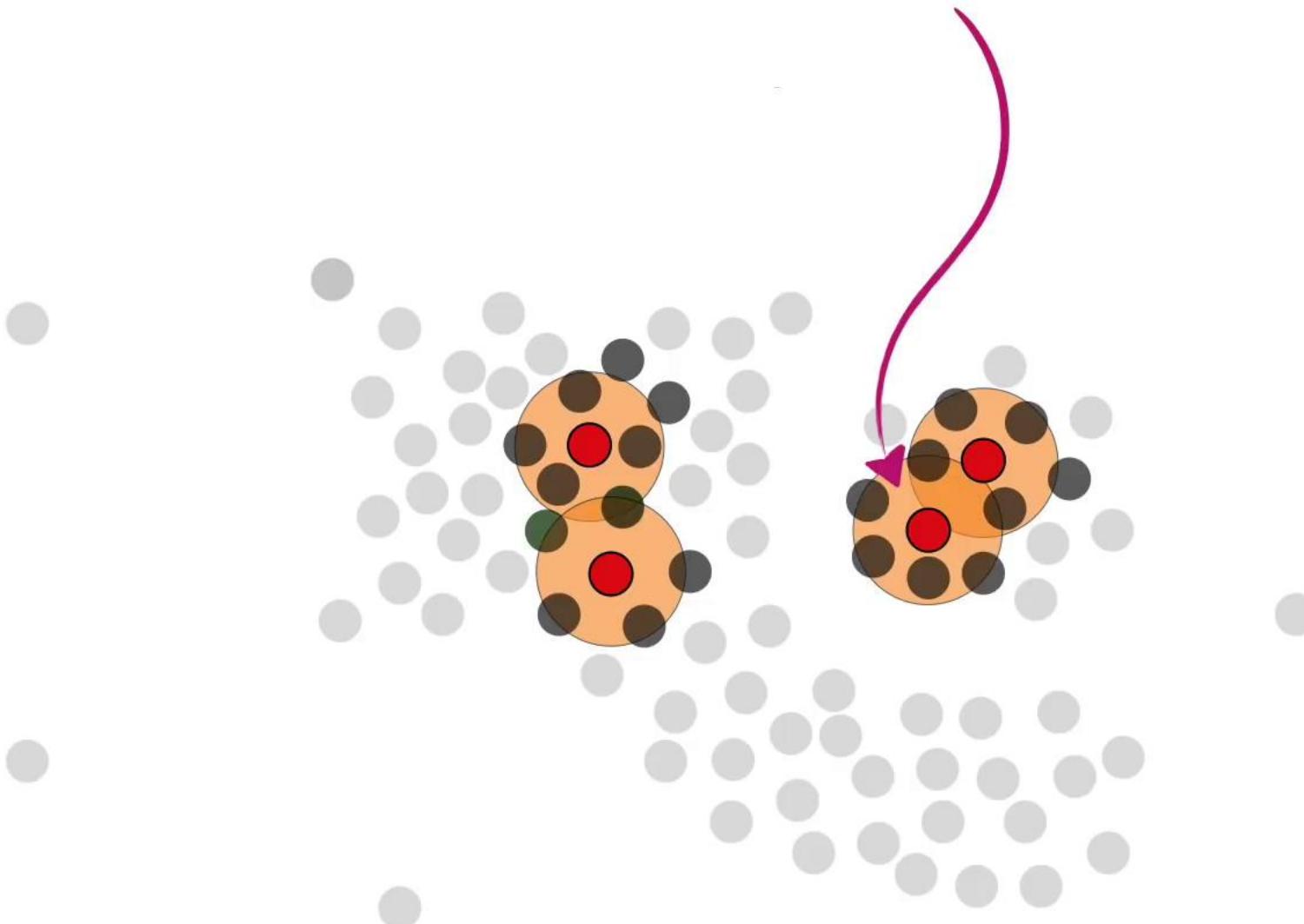
طريقة عمل خوارزمية DBScan

نلاحظ بعد رسم الدائرة البرتقالية وجود نقاط تحاط بها وأيضاً دائرة هذه النقطة تقاطع مع النقطة السابقة







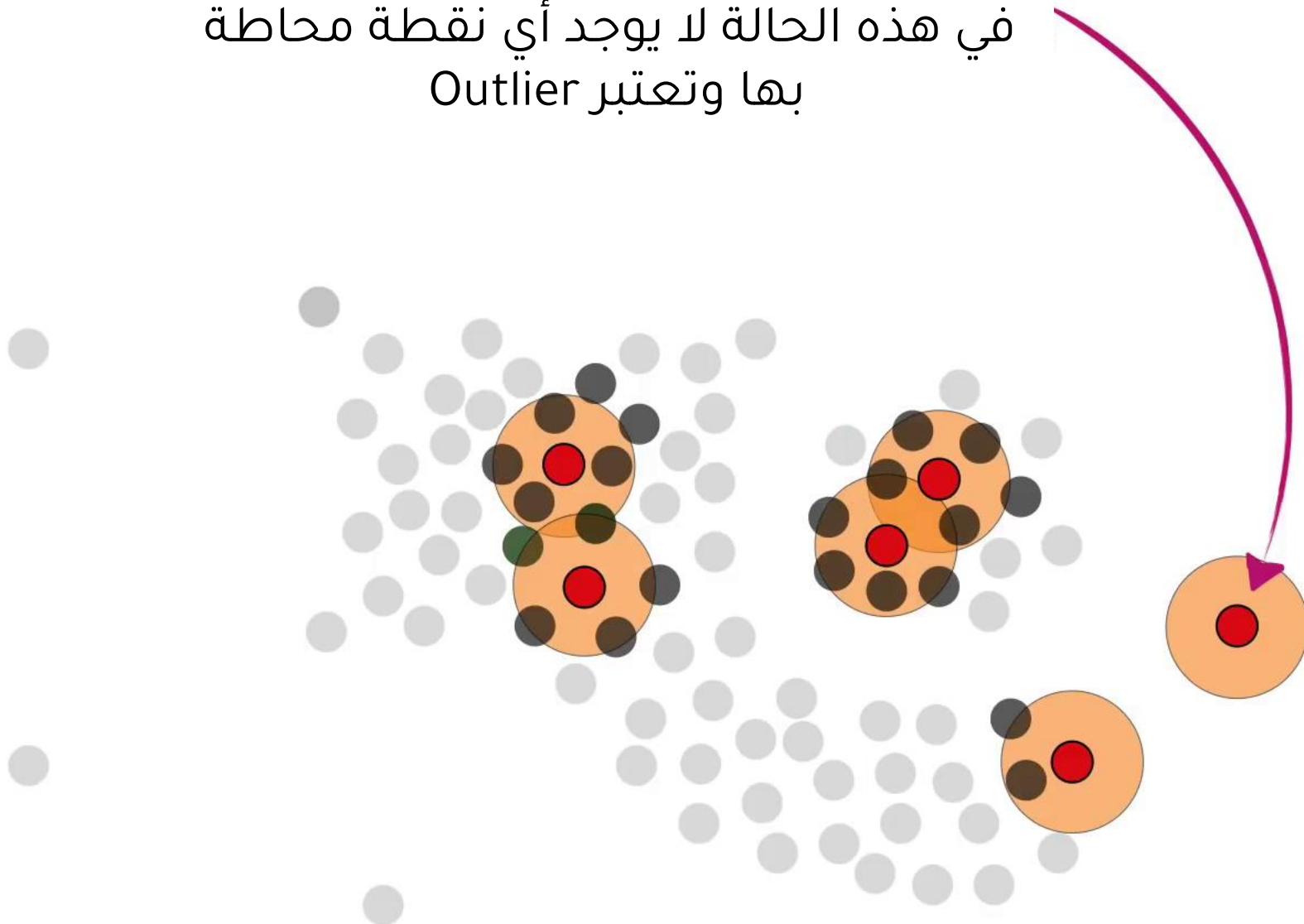


من الملاحظ هنا وجود
فقط نقطتين تحاط بهم



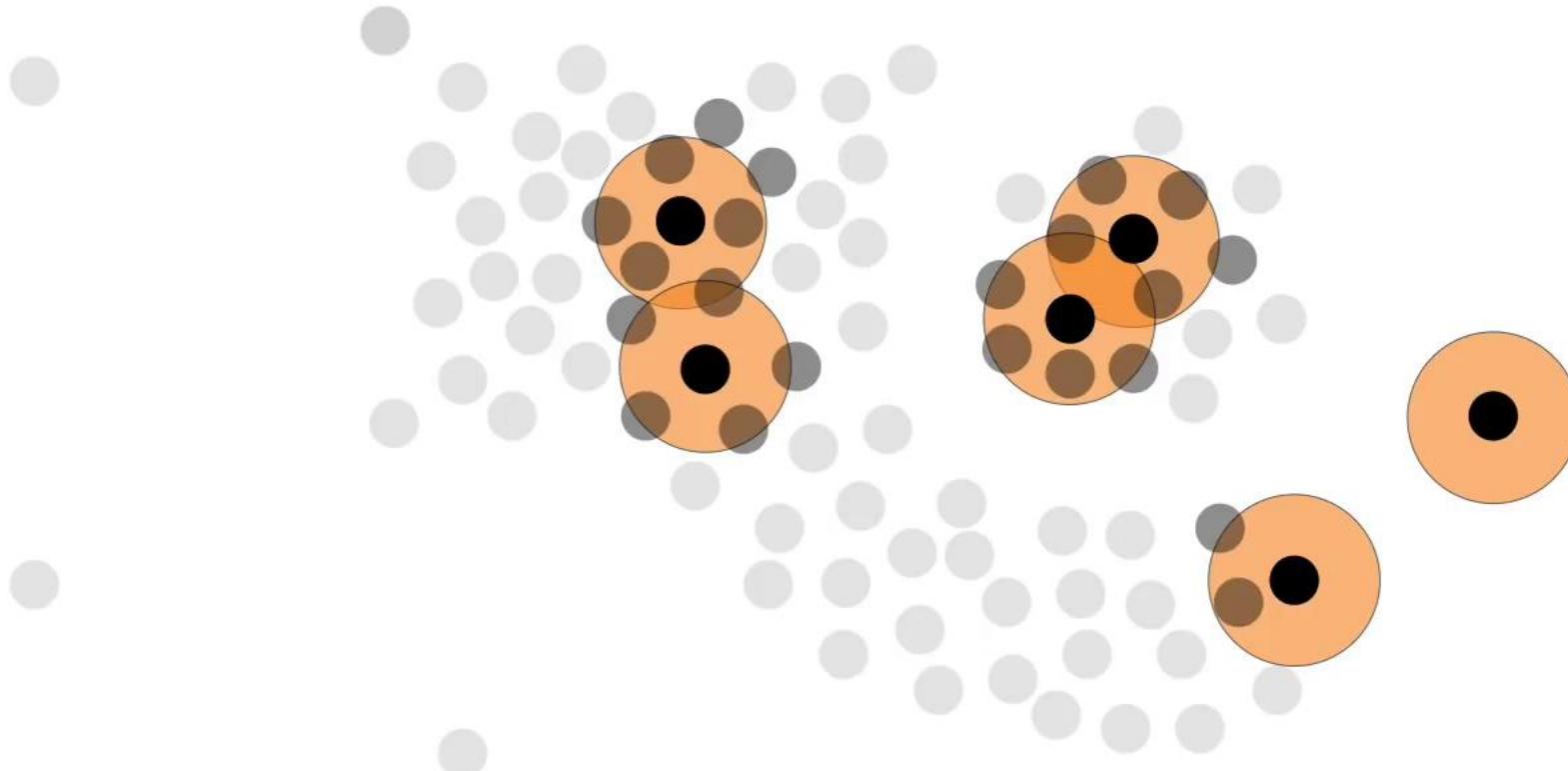


في هذه الحالة لا يوجد أي نقطة محاطة
بها وتعتبر Outlier



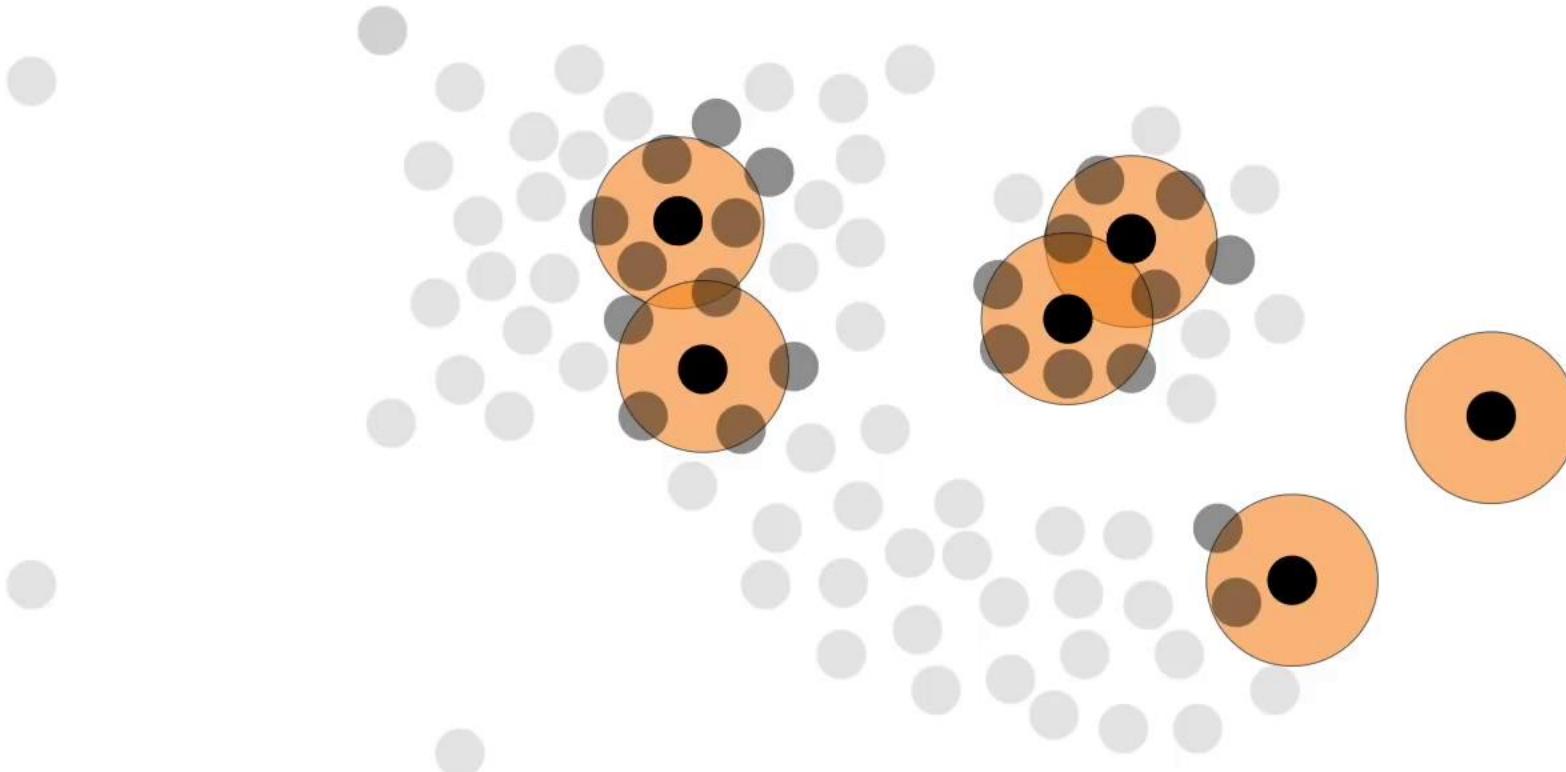


بعد ما كررنا هذه العملية لخمس نقاط وجدنا
أن غالبية النقاط يحاط بها على الأقل **4 نقاط**



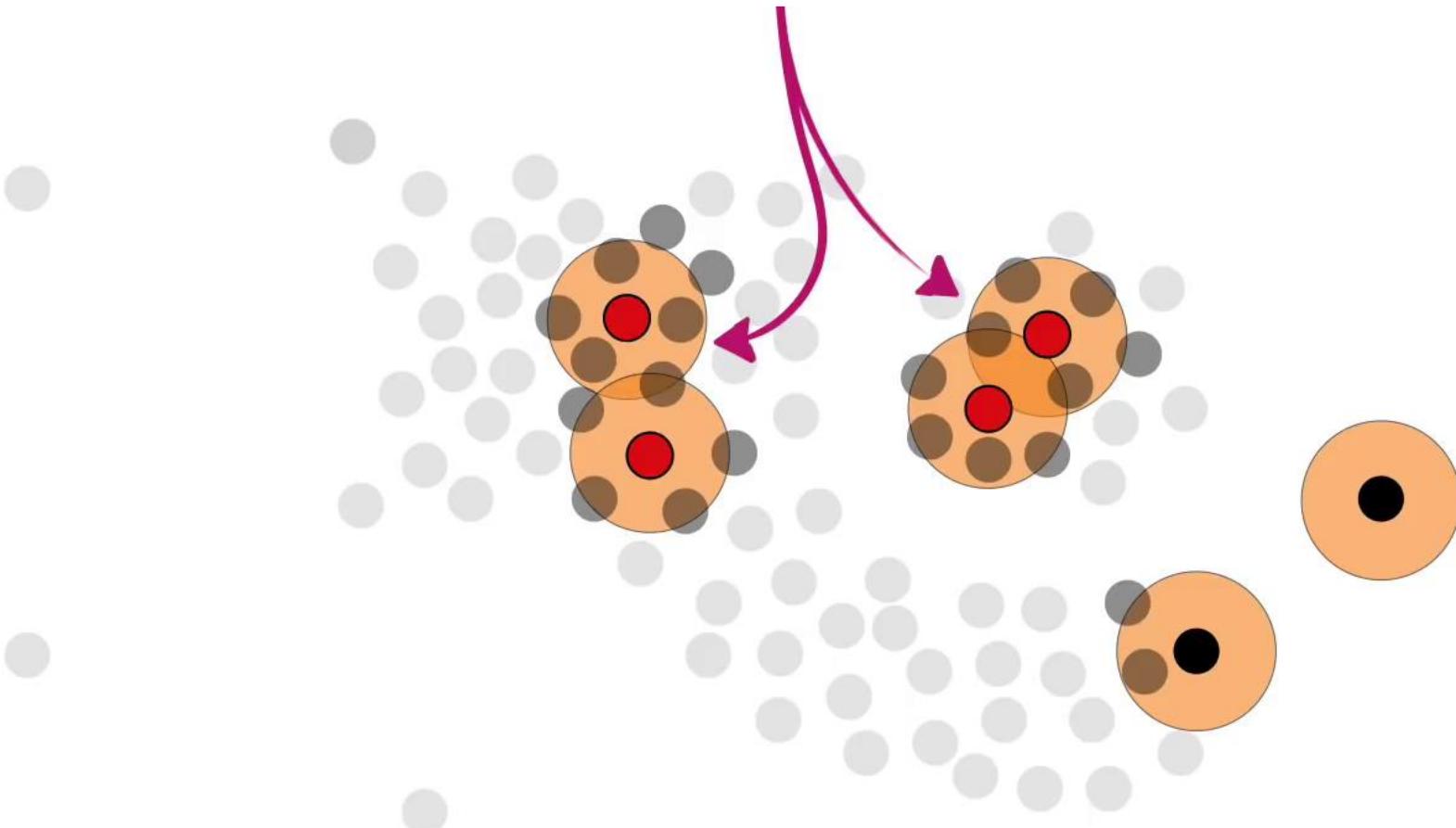


ملاحظة: عدد النقاط المحاط بها تكون معرف من قبل المستخدم للخوارزمية لا dbSCAN



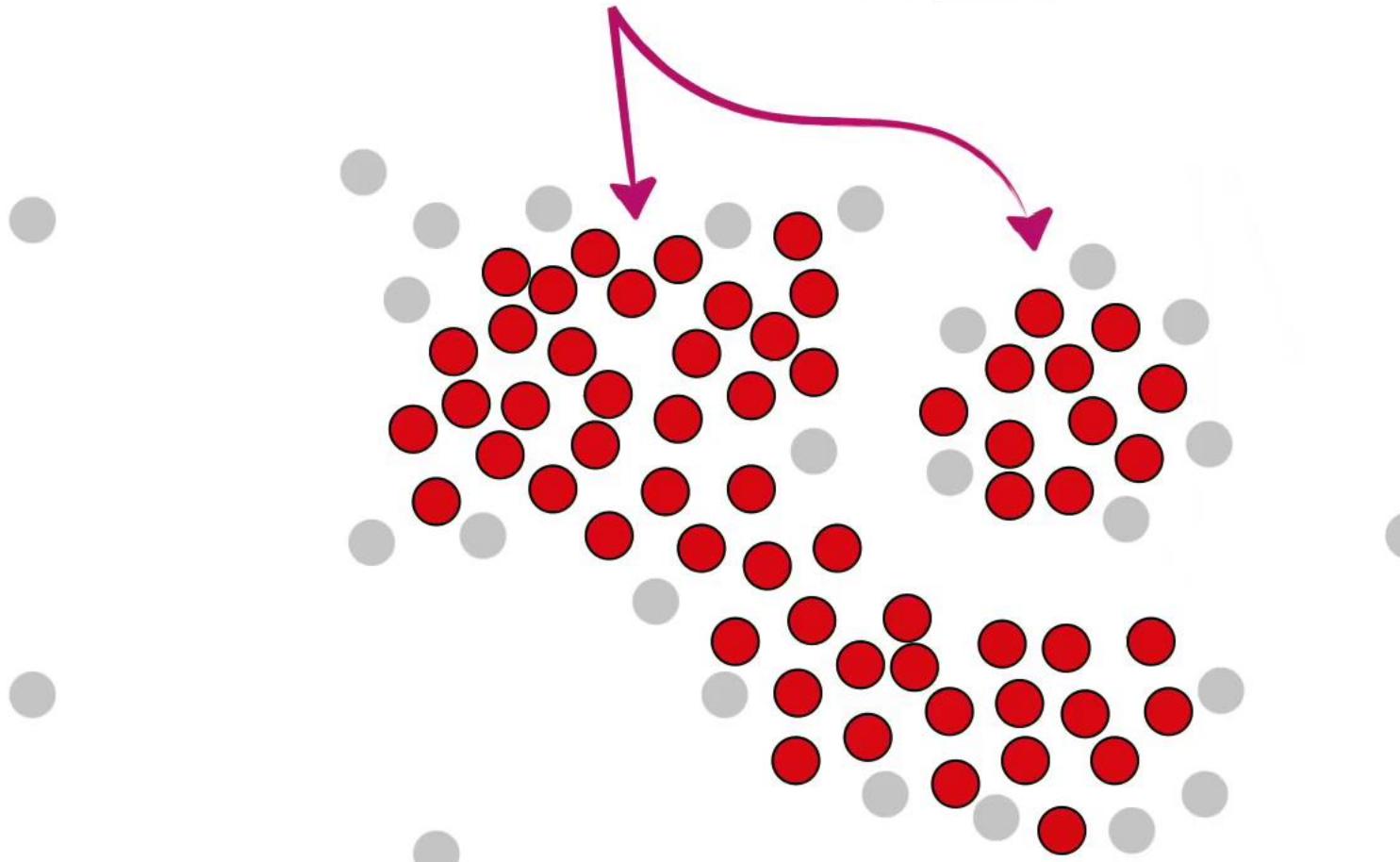


و على هذا الأساس تعتبر أي نقطة تحيط ب أربعة
نقاط على الأقل تسمى بـ Core Point





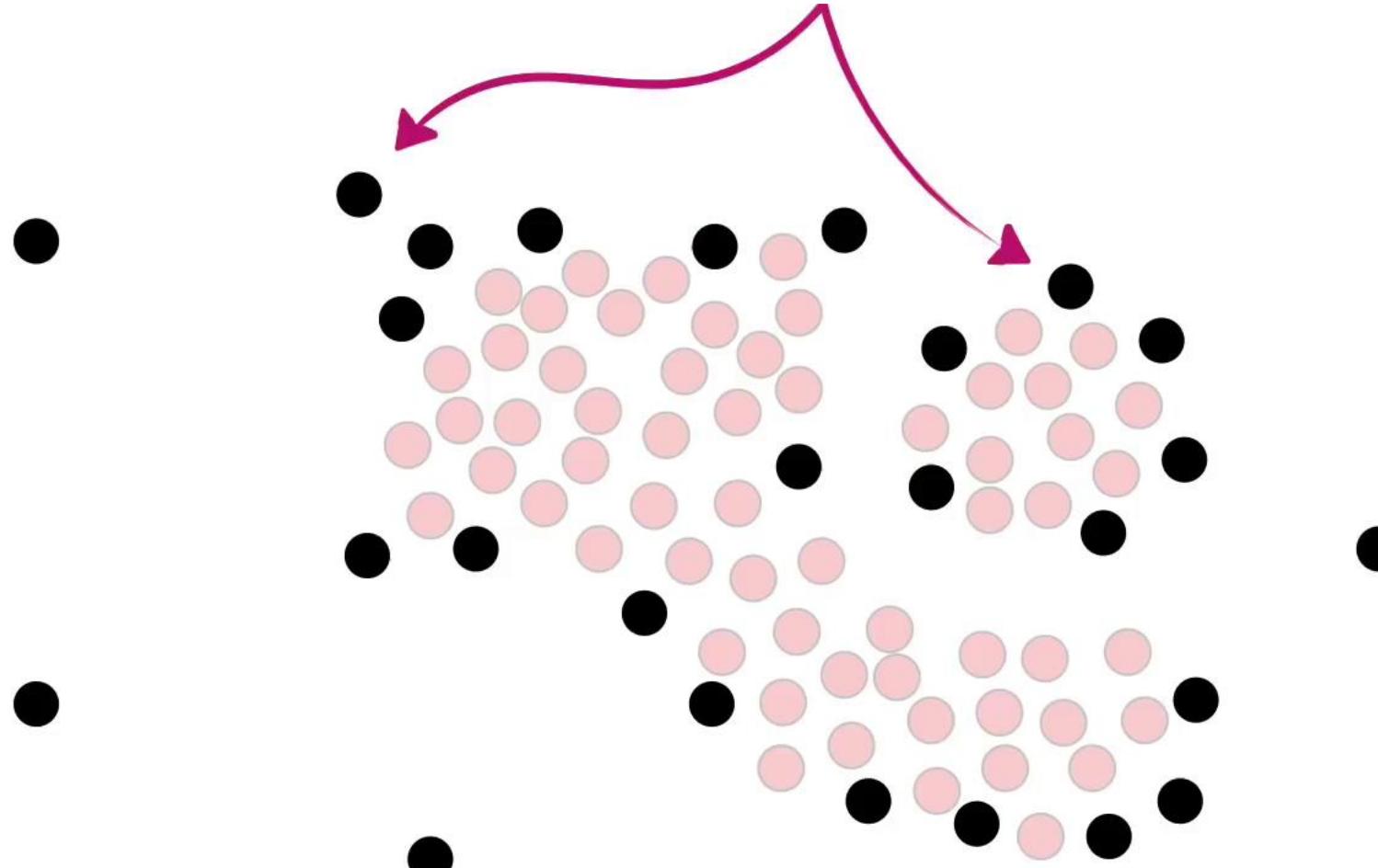
بعد تكرار العملية على كافة النقاط، النقاط
المشار لها باللون الأحمر تعتبر Core Points





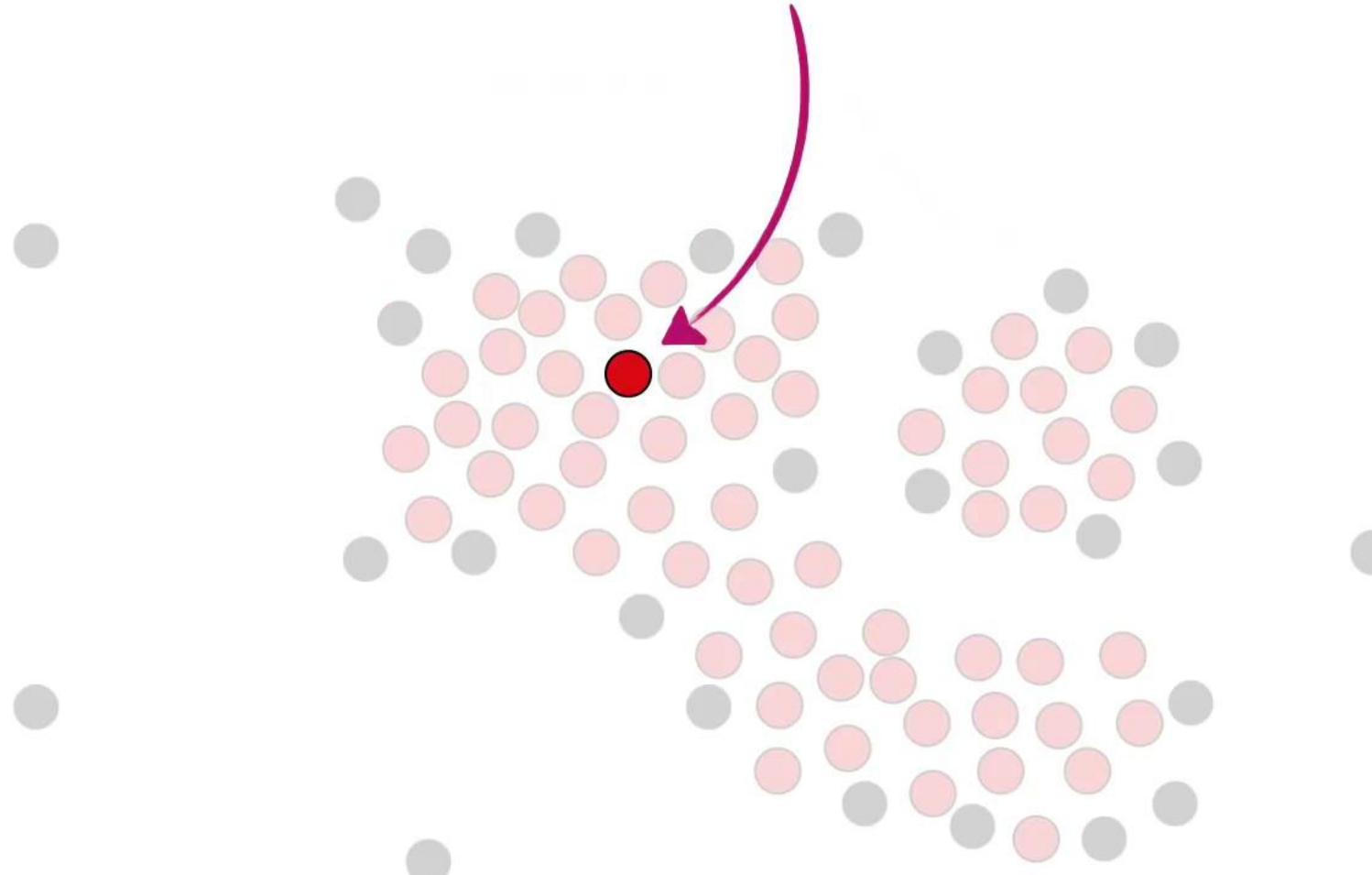
بقية النقاط التي لم تحيط بأربع نقاط على الأقل

تعتبر Non-Core Points

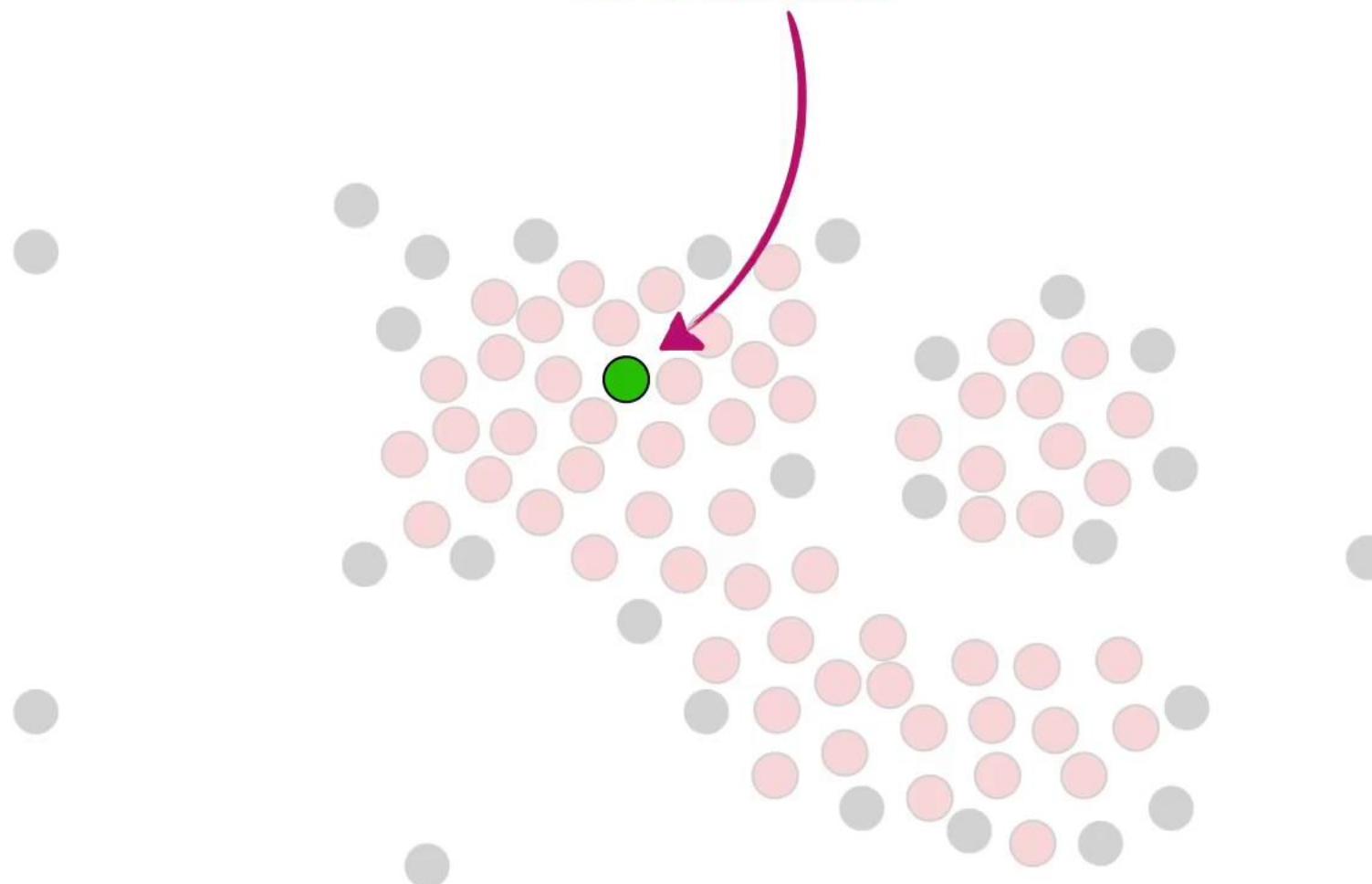




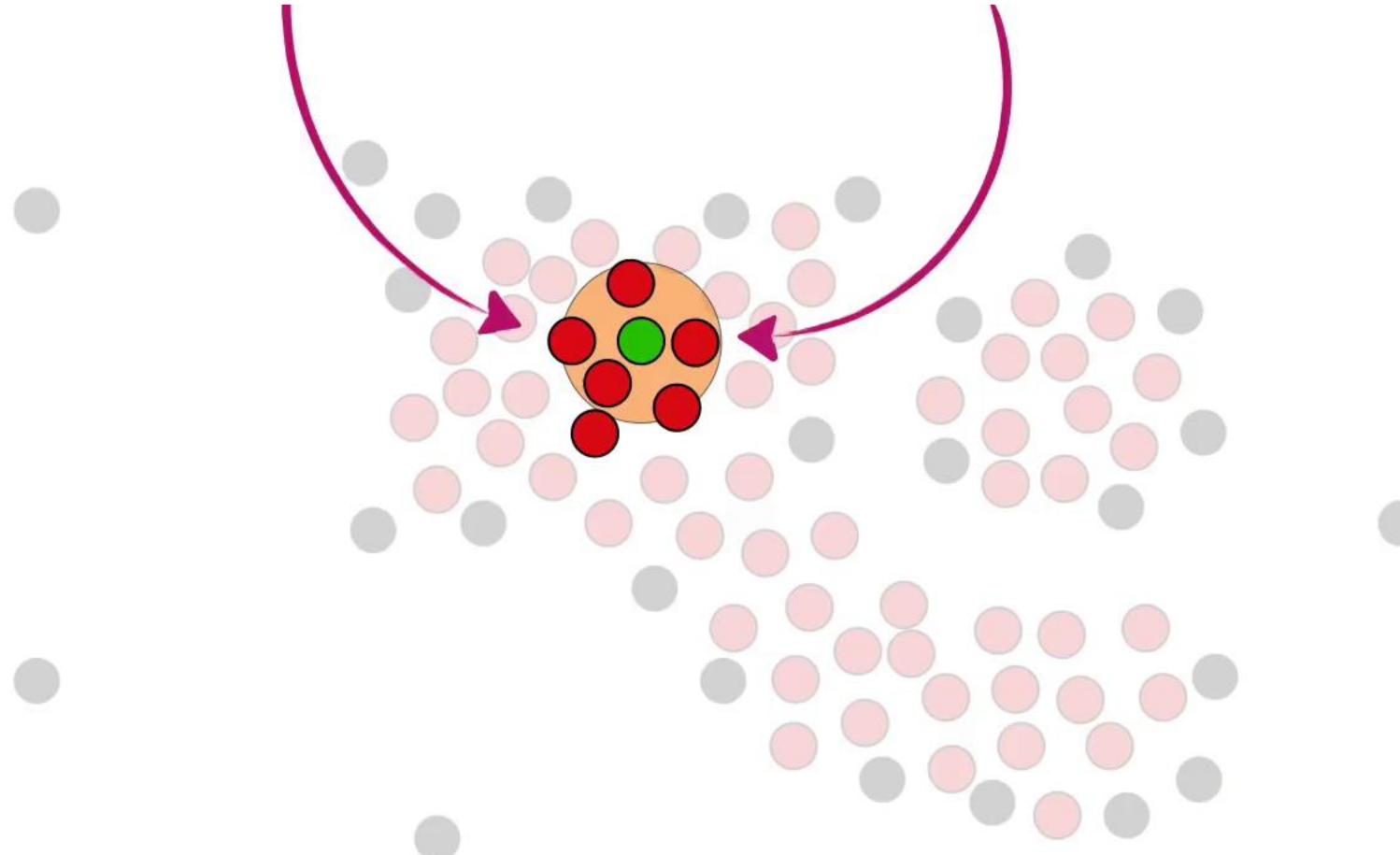
الآن نبدأ بعملية تصنیف الـ Clustering

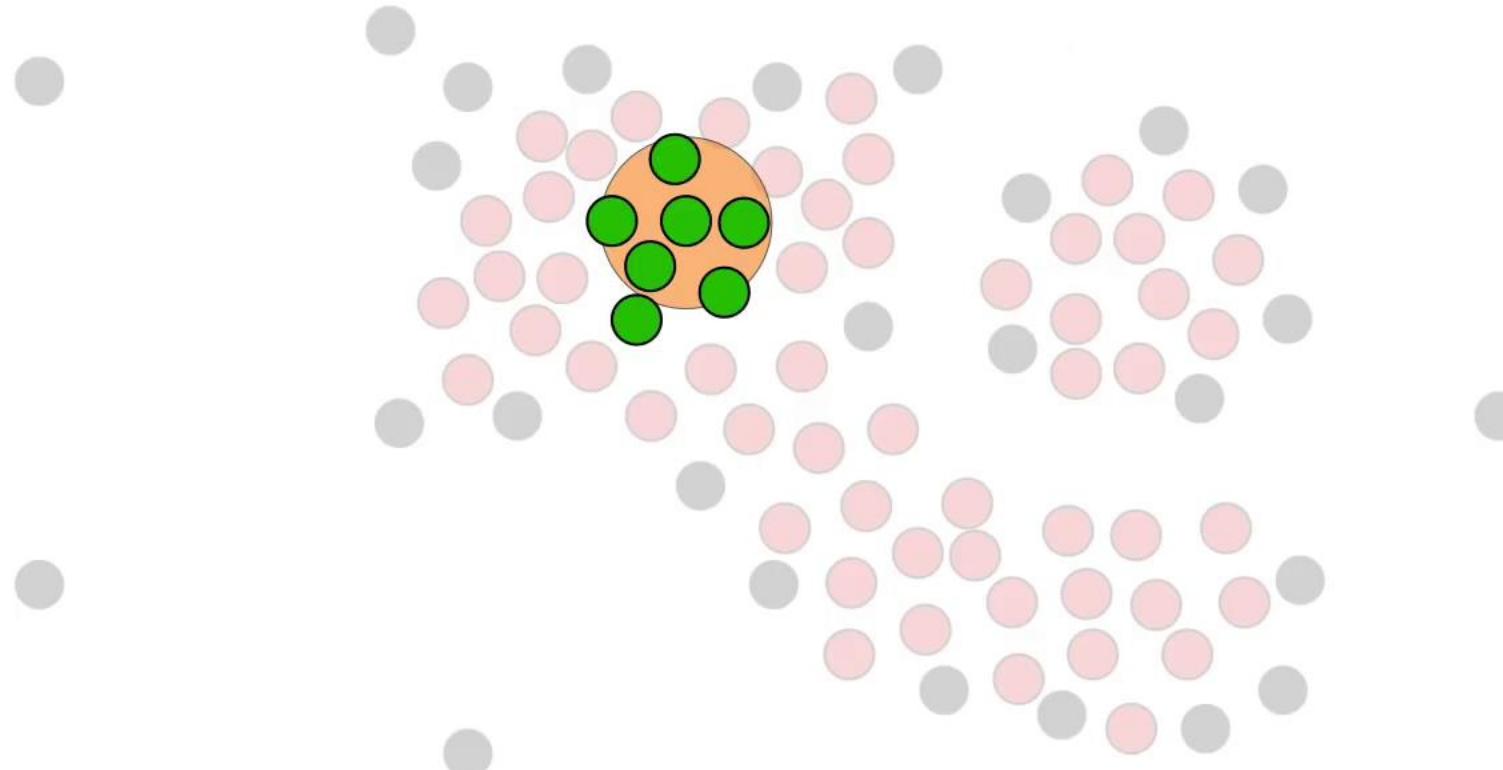


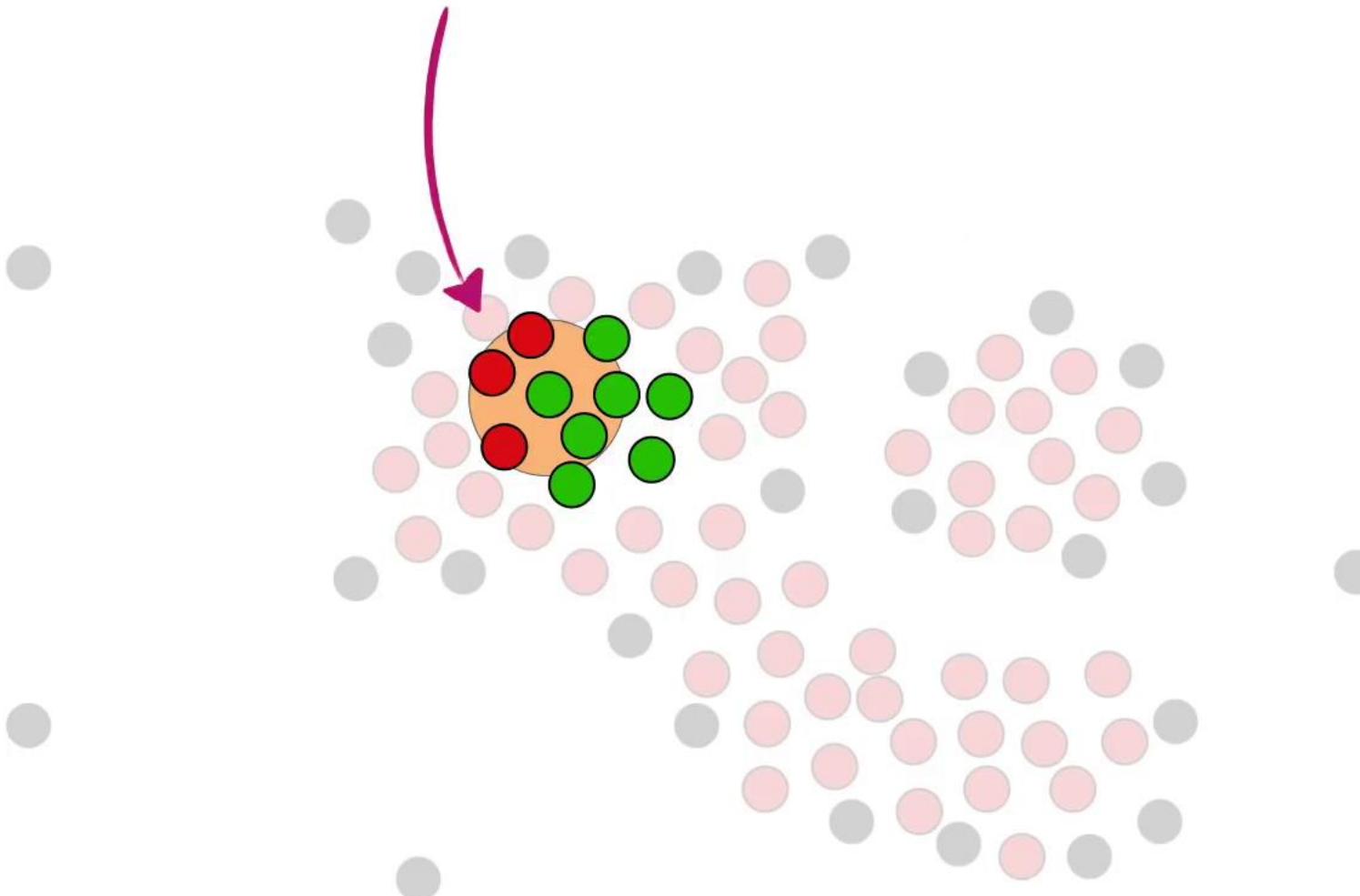
نختار نقطة بشكل عشوائي ونحددها بـ First Cluster

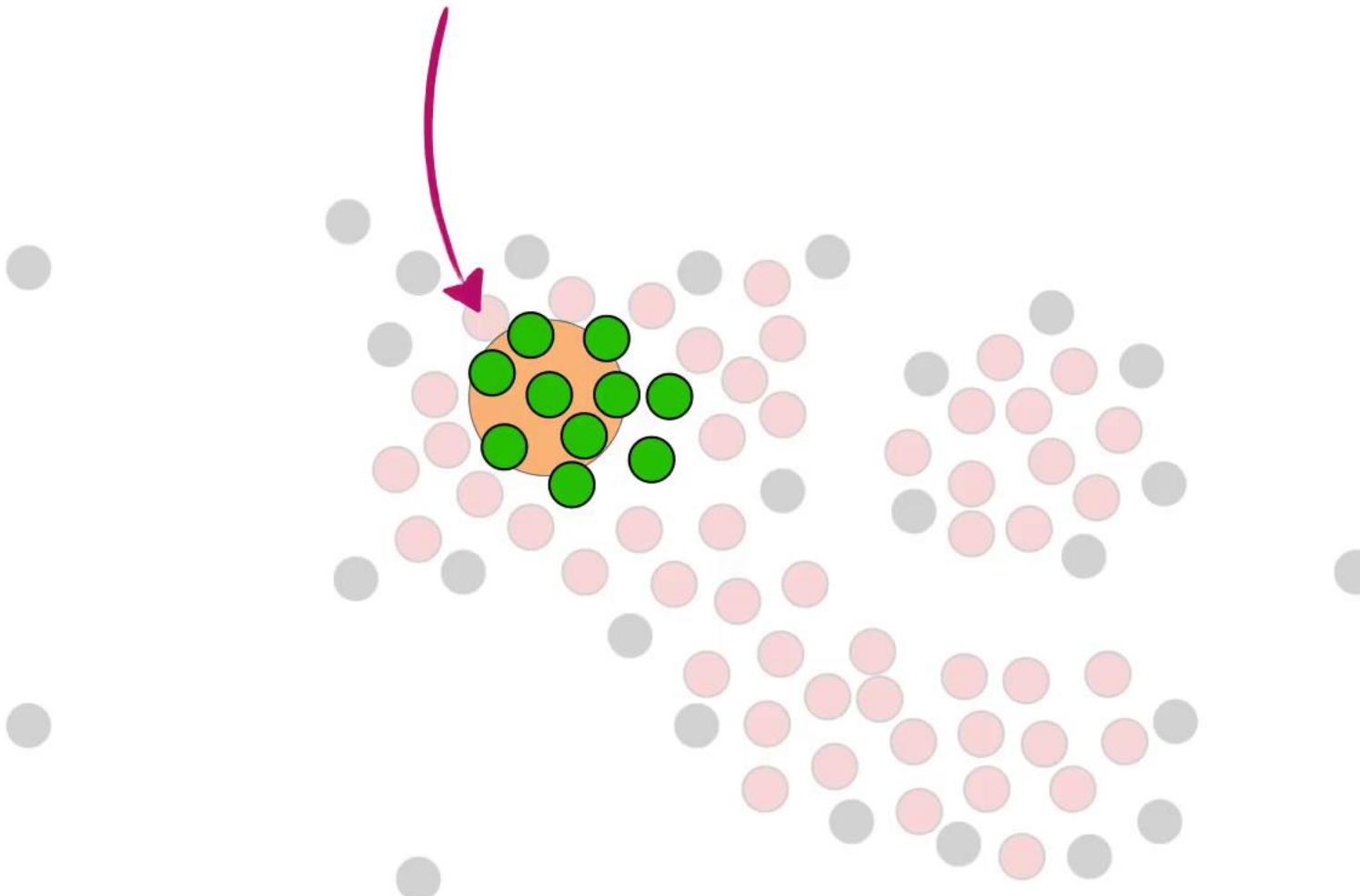


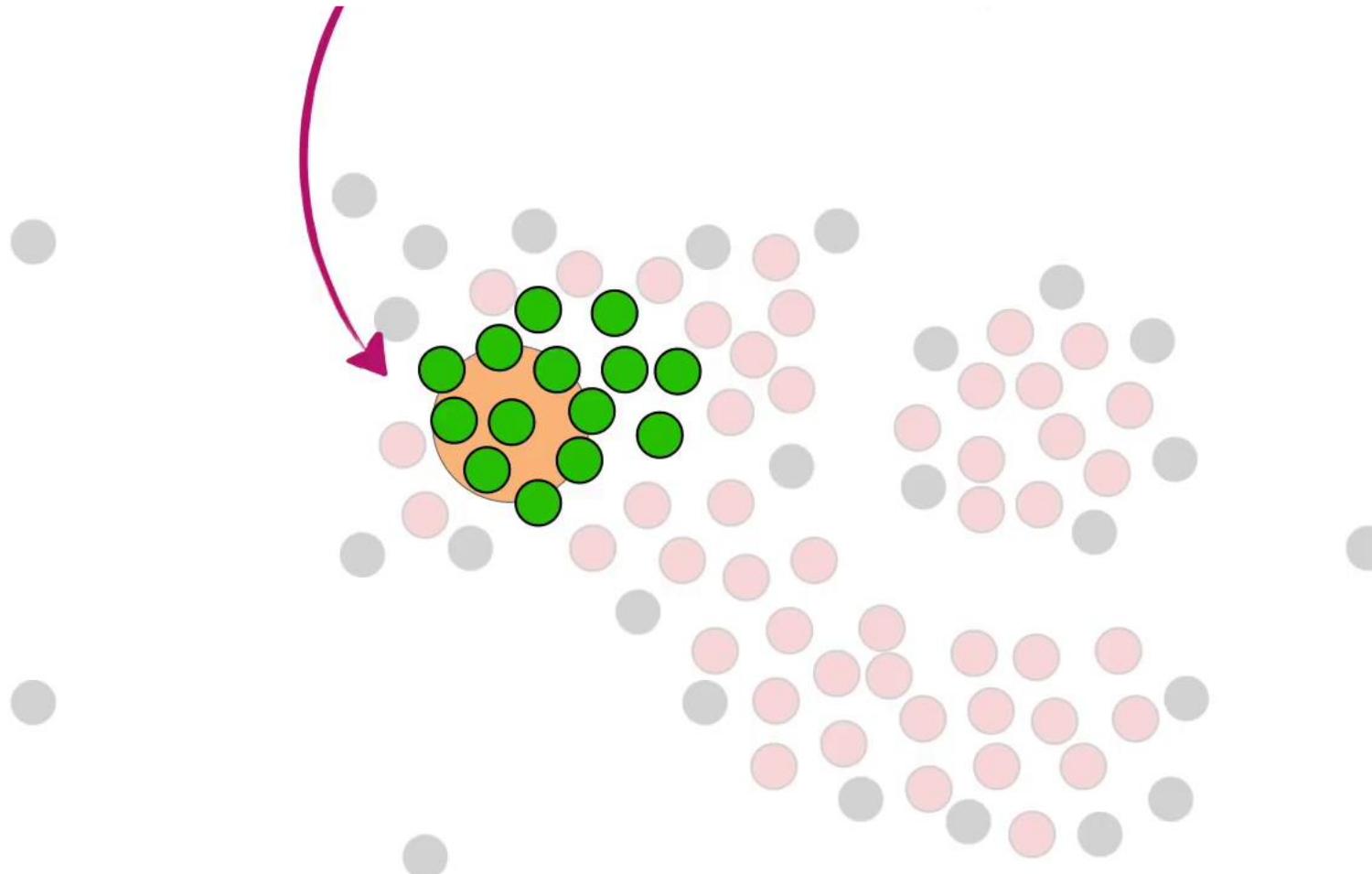
أي نقطة مصنفة لأي نوع، النقاط التي تحيط بها تنضم لها في الـ Cluster



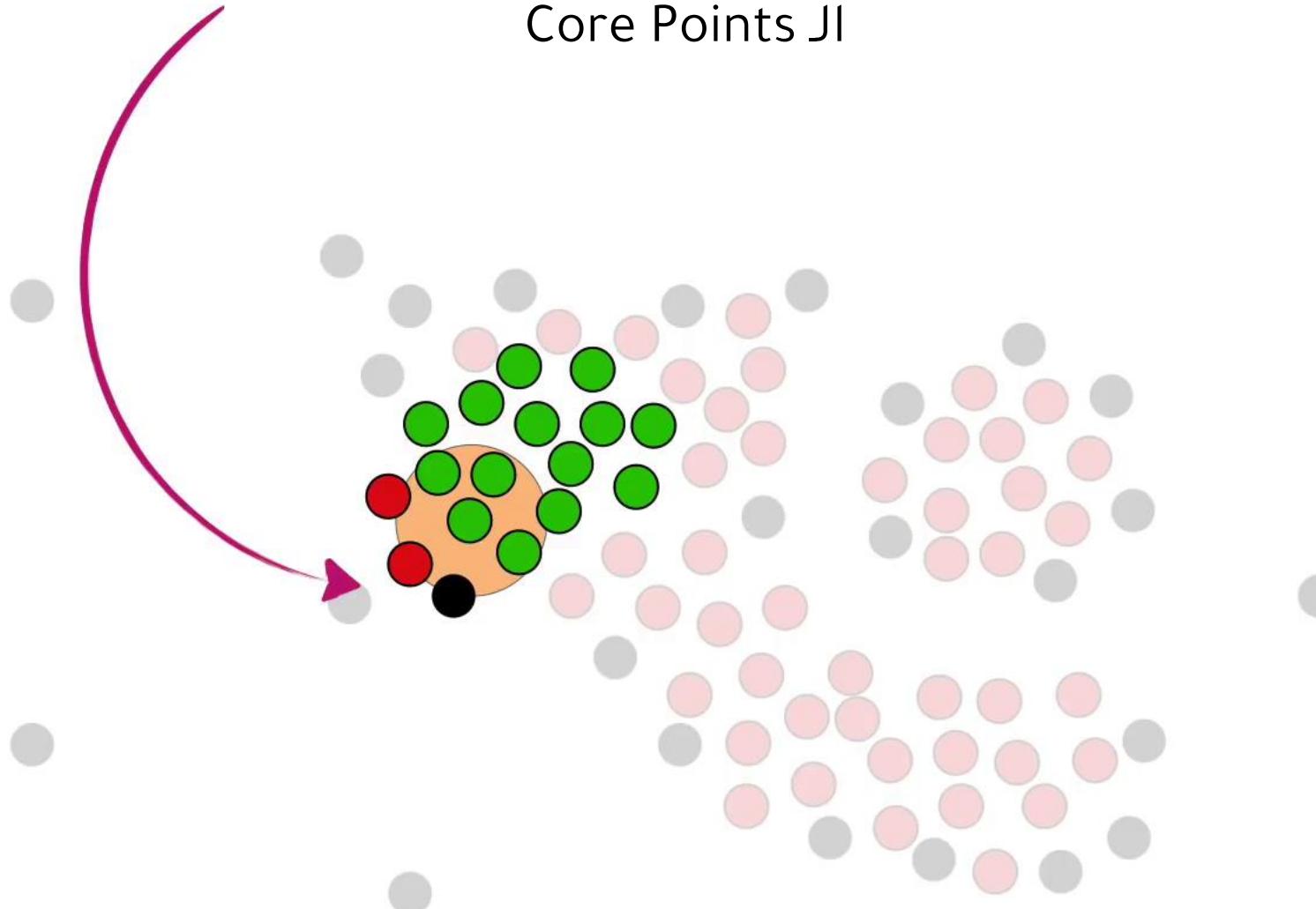


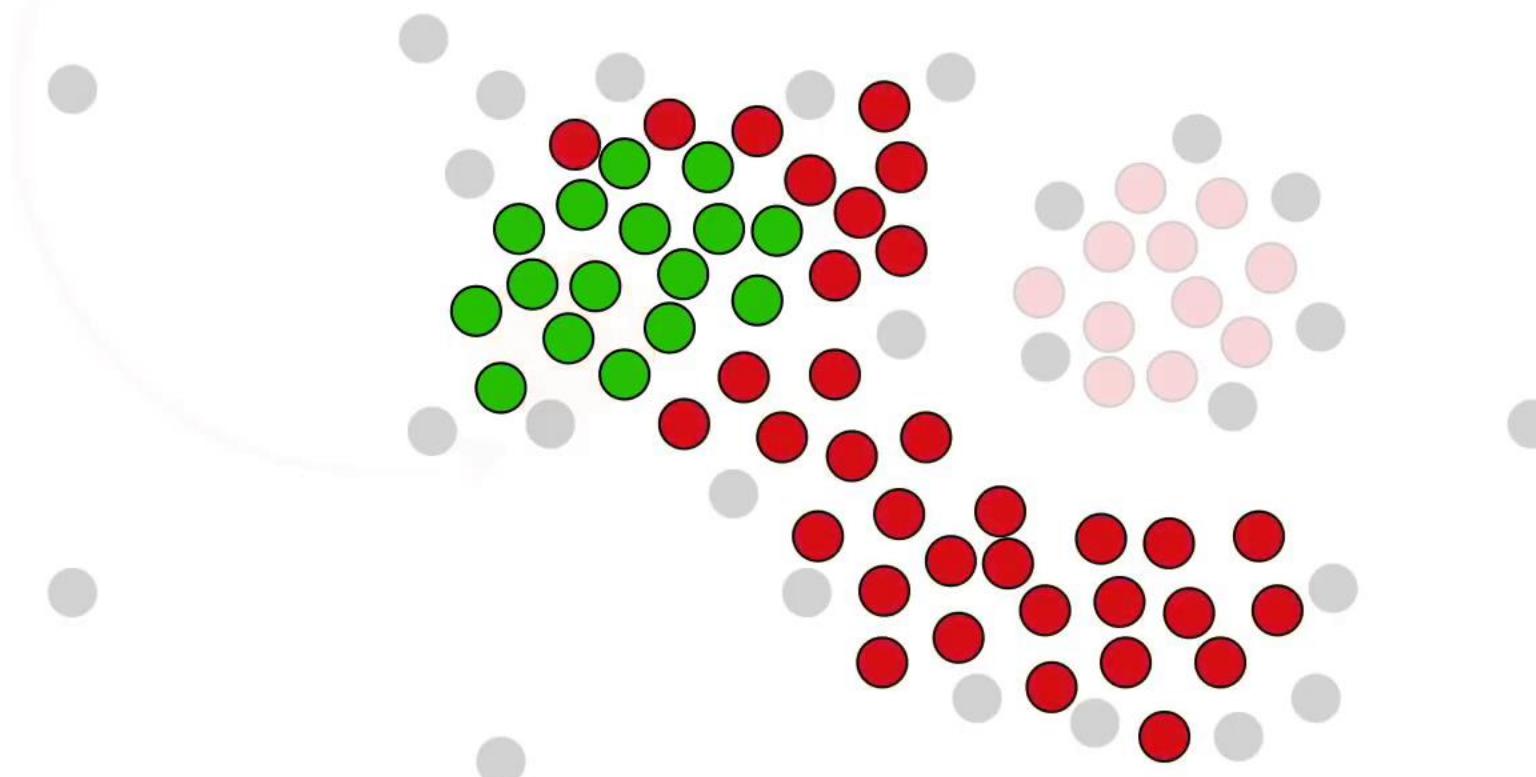


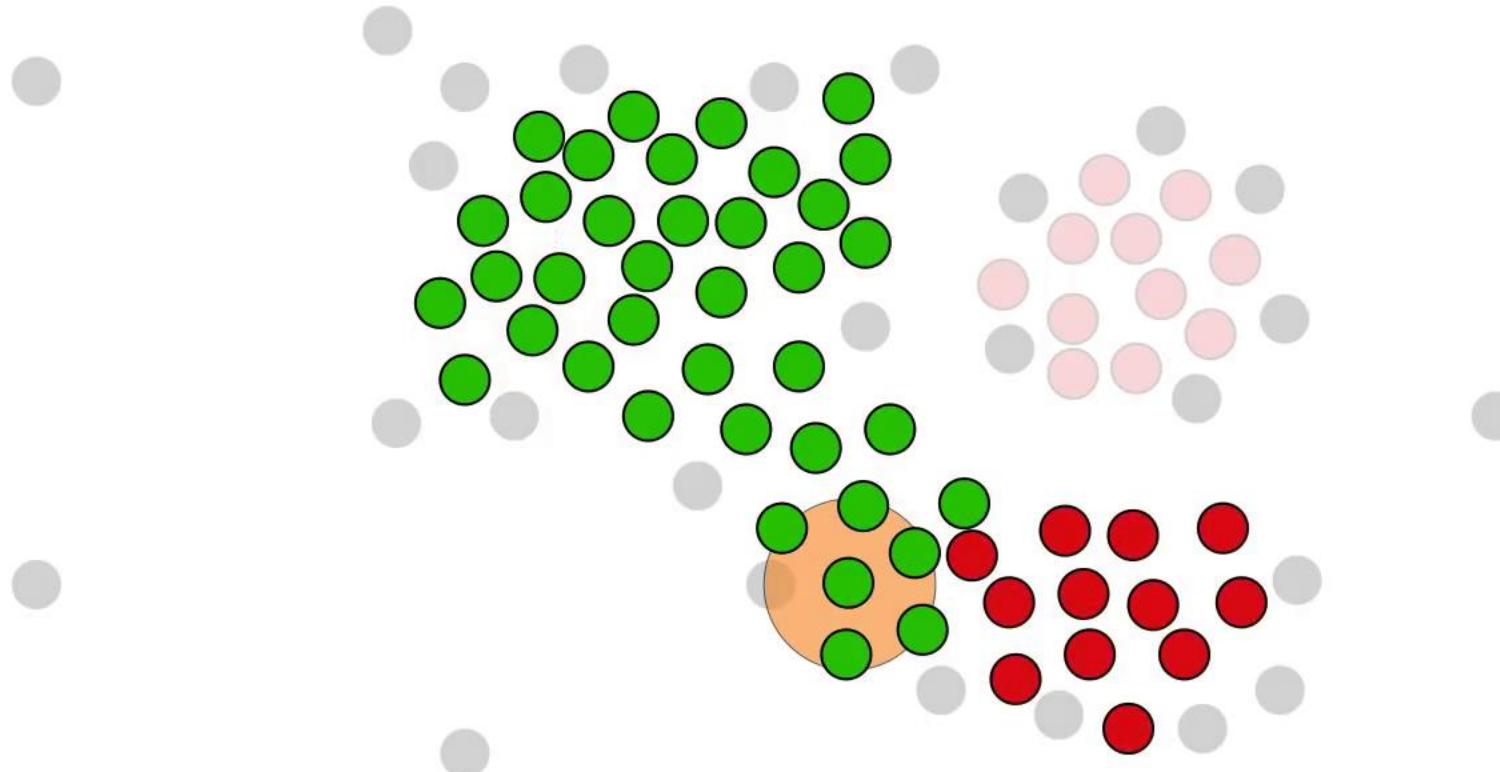


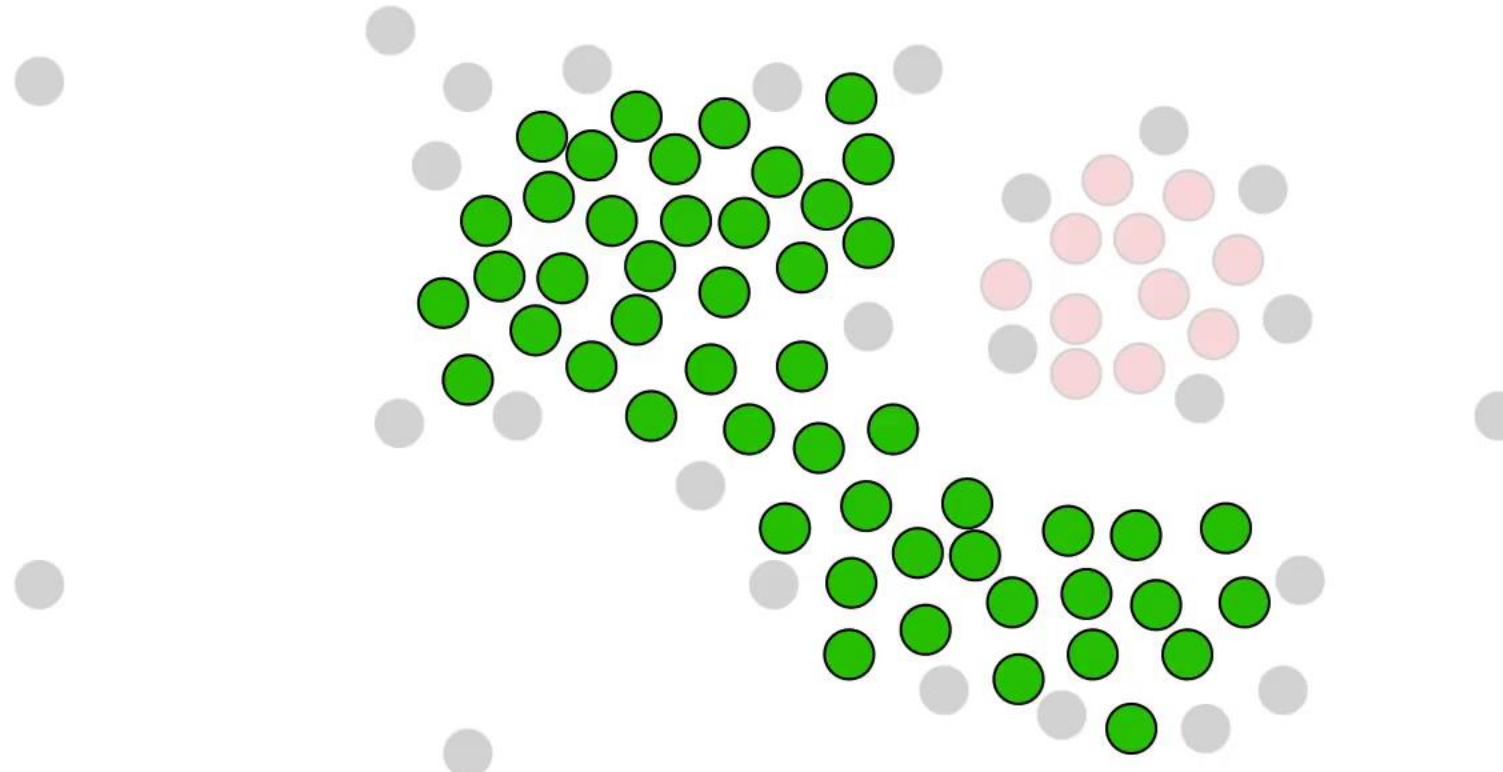


في هذه المرحلة نضم لـ Cluster الـ Core Points فقط



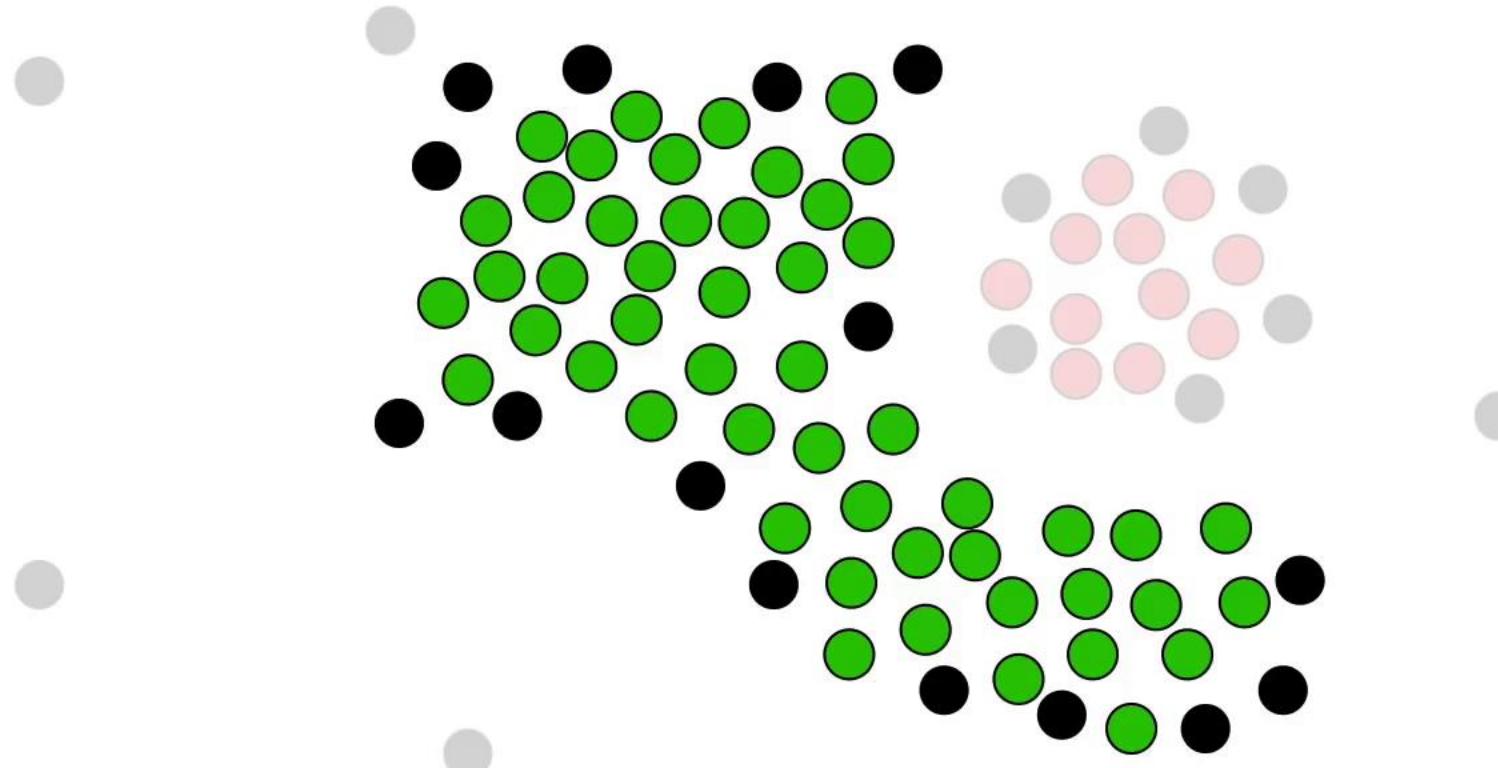






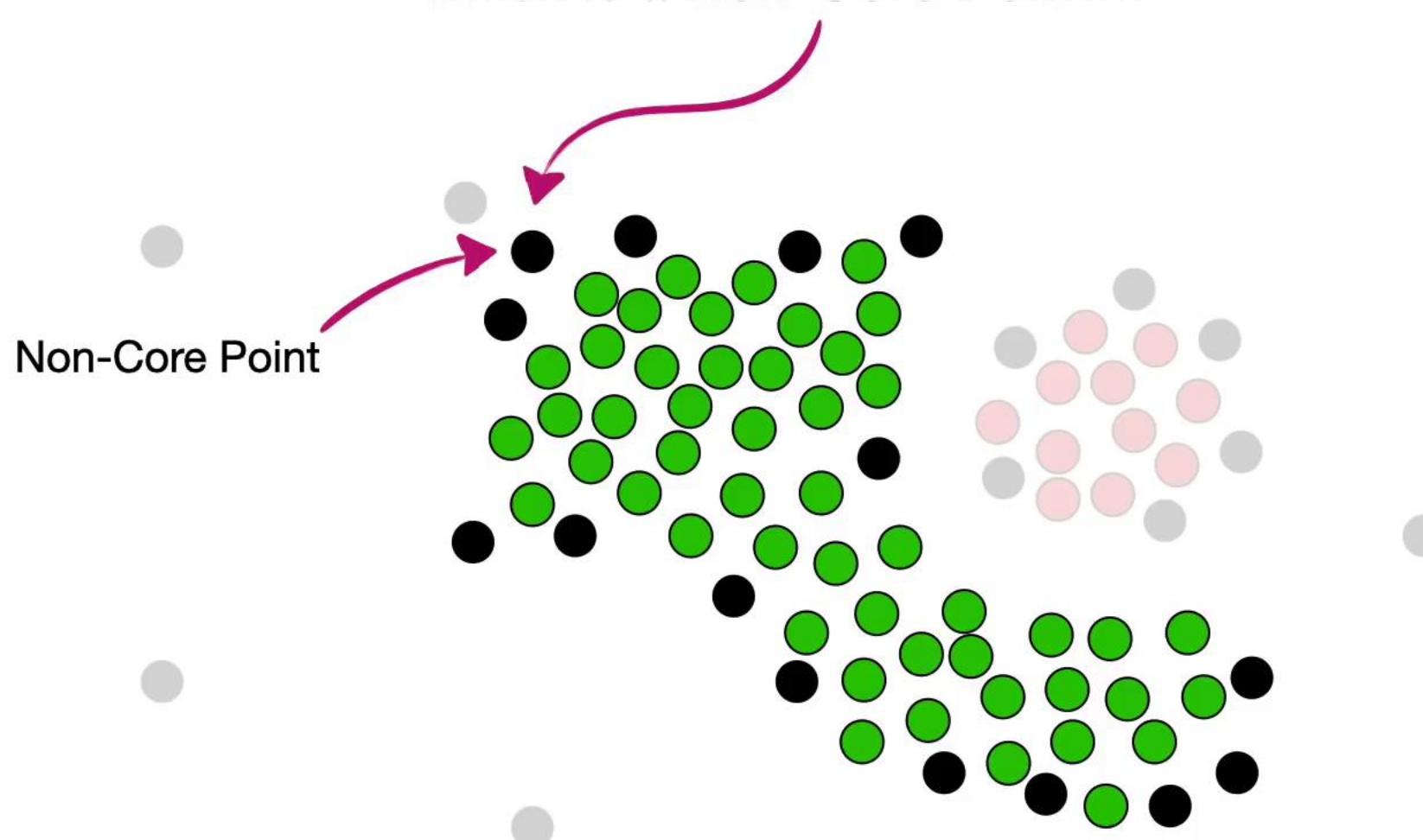


انتهينا من المرحلة الأولى لتصنيف الـ
Core Points



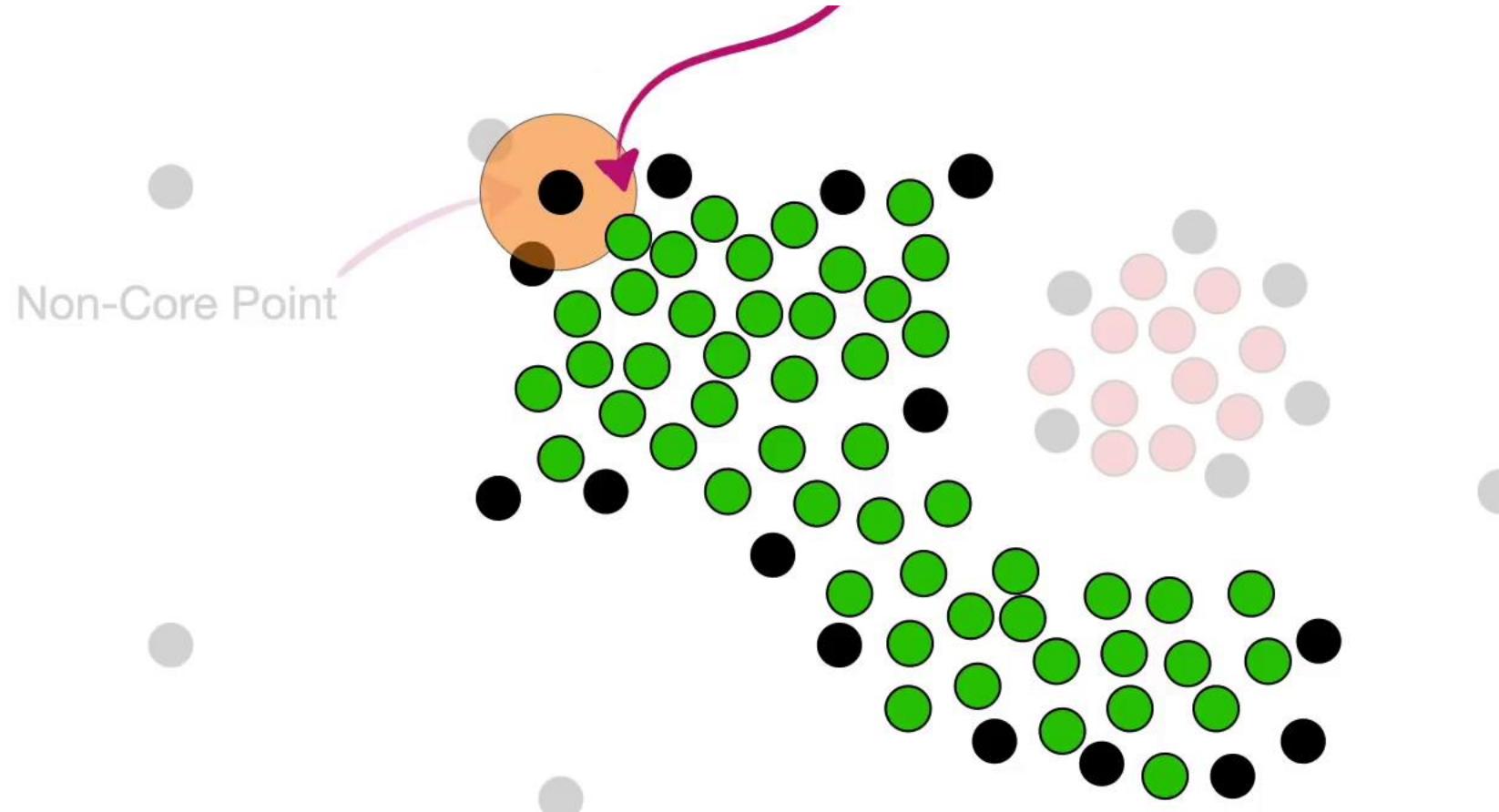


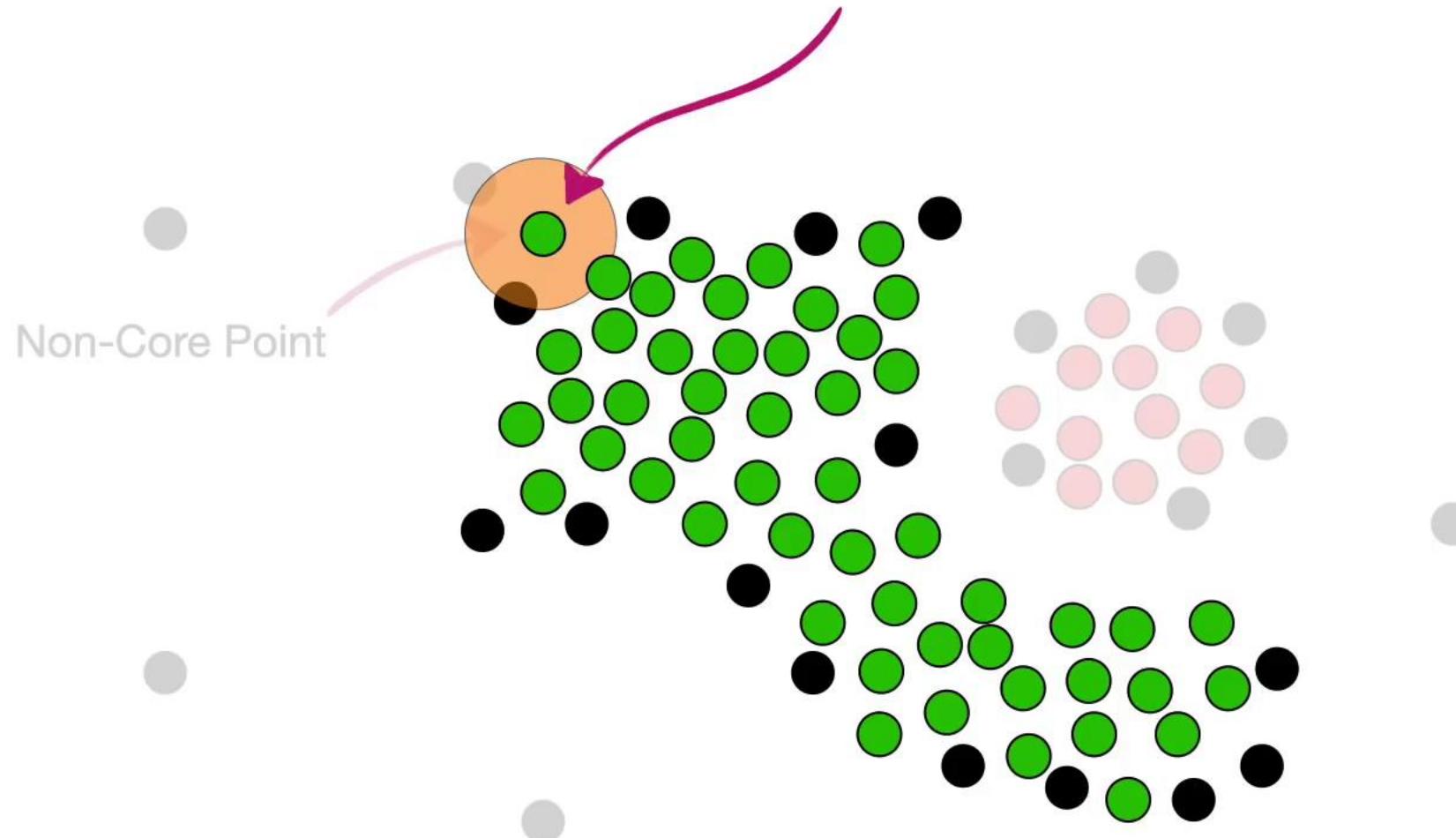
المرحلة التي تليها تصنیف الـ Core Points



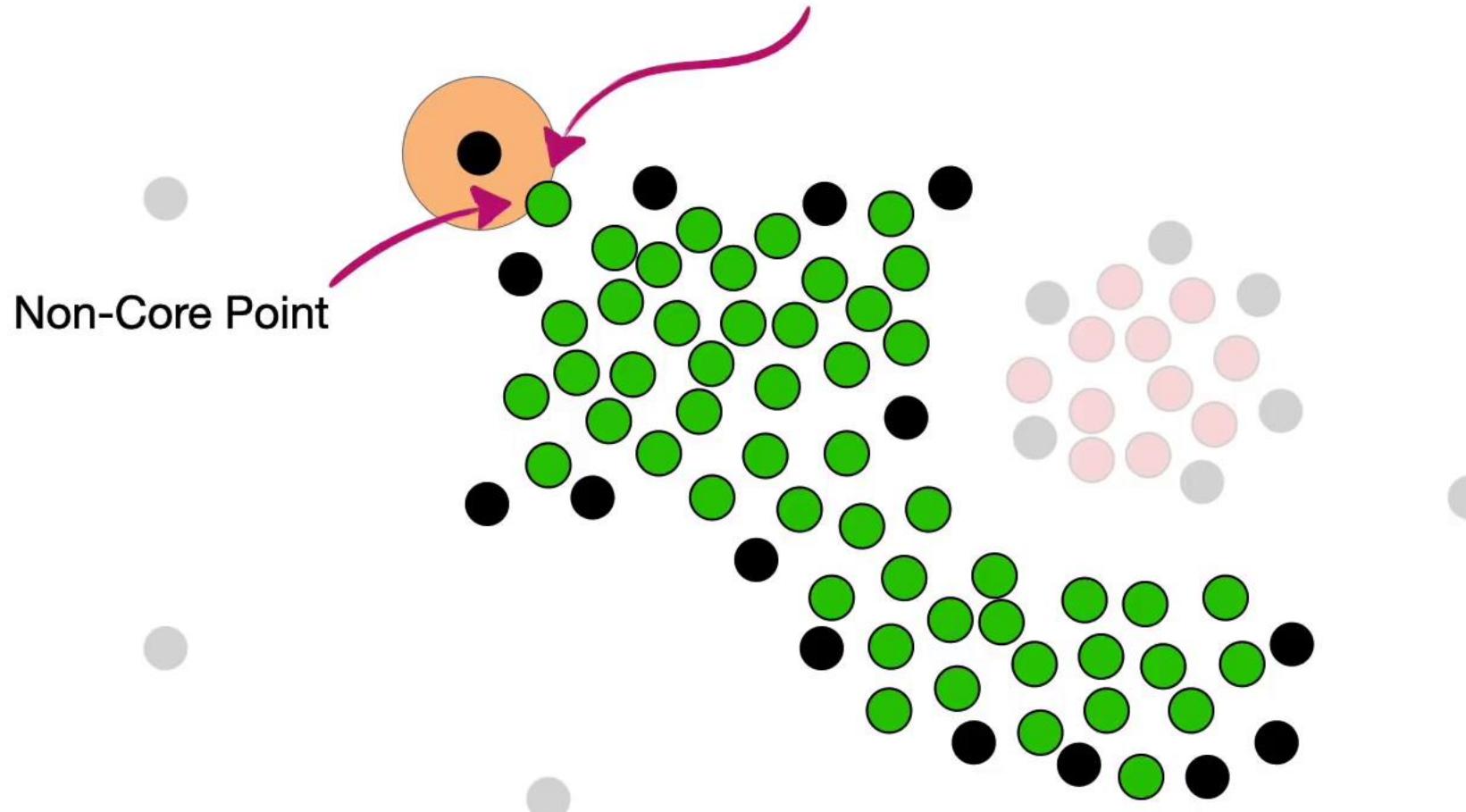


نرسم الدائرة البرتقالية المستخدمة
سابقاً، أقرب دائرة مصنفة ستنتضم
لها في ال Cluster



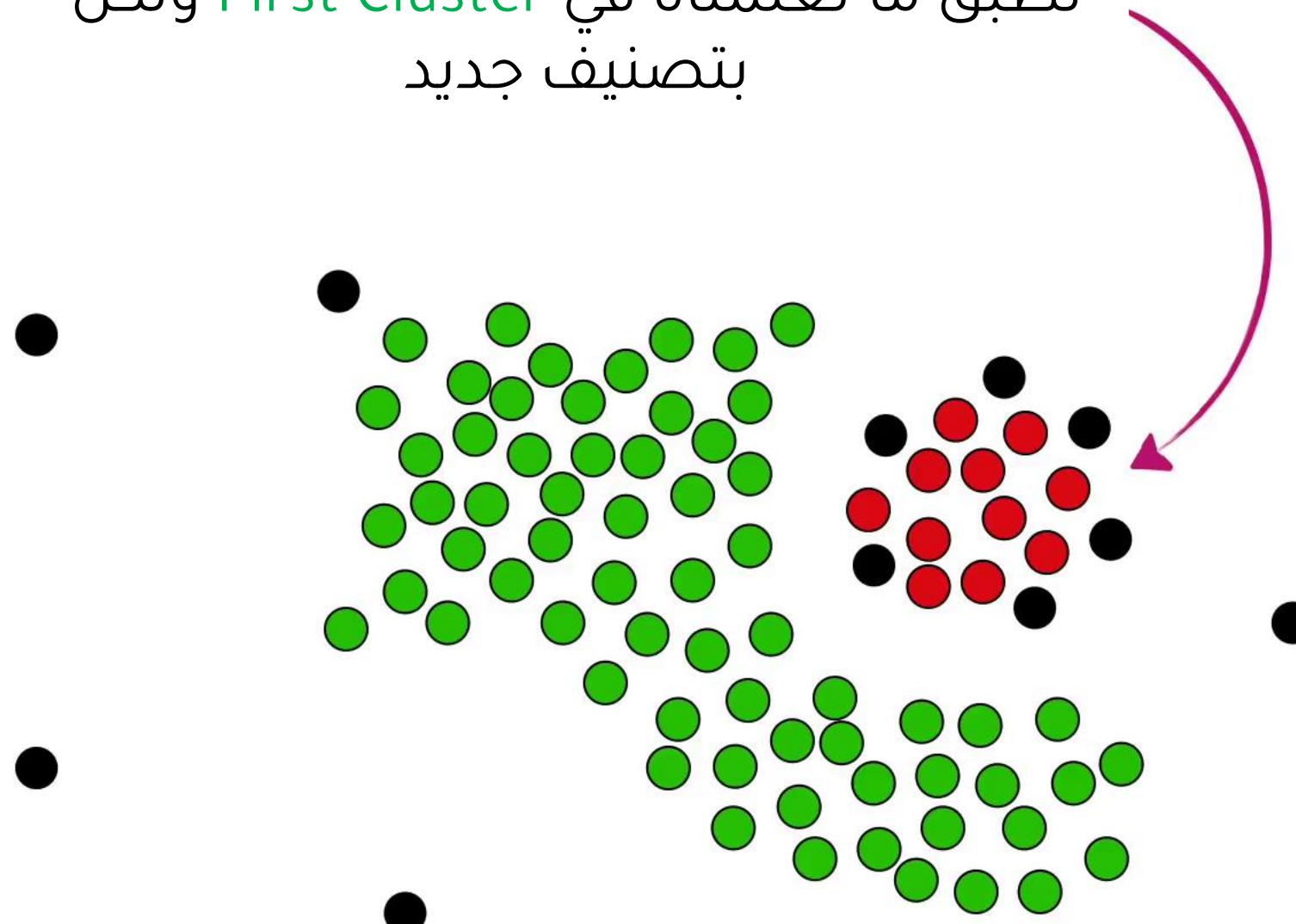


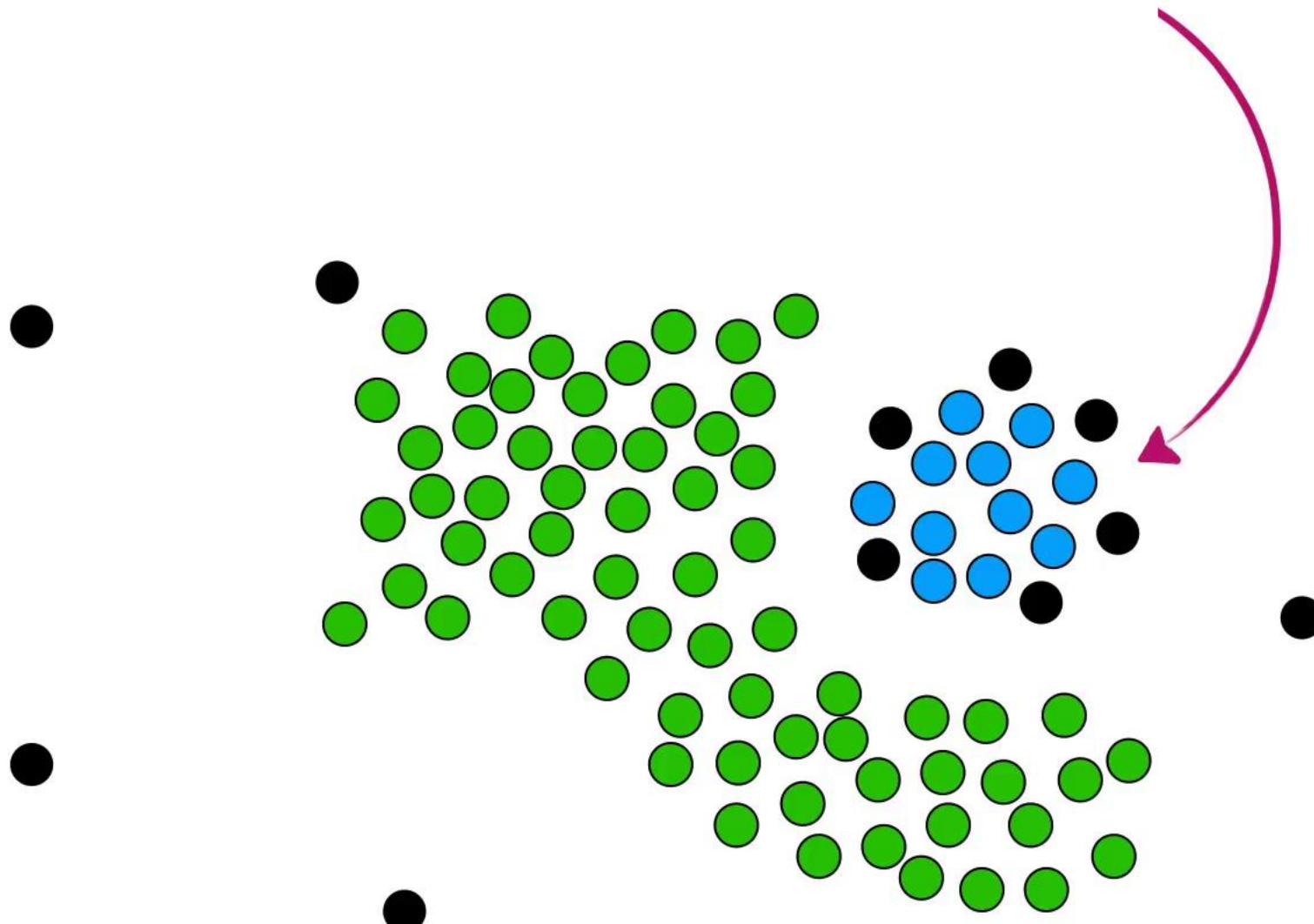
Non-Core Point أي معلومة معماة: مصنفة لا يمكن أن تضم نقطة أخرى لها.

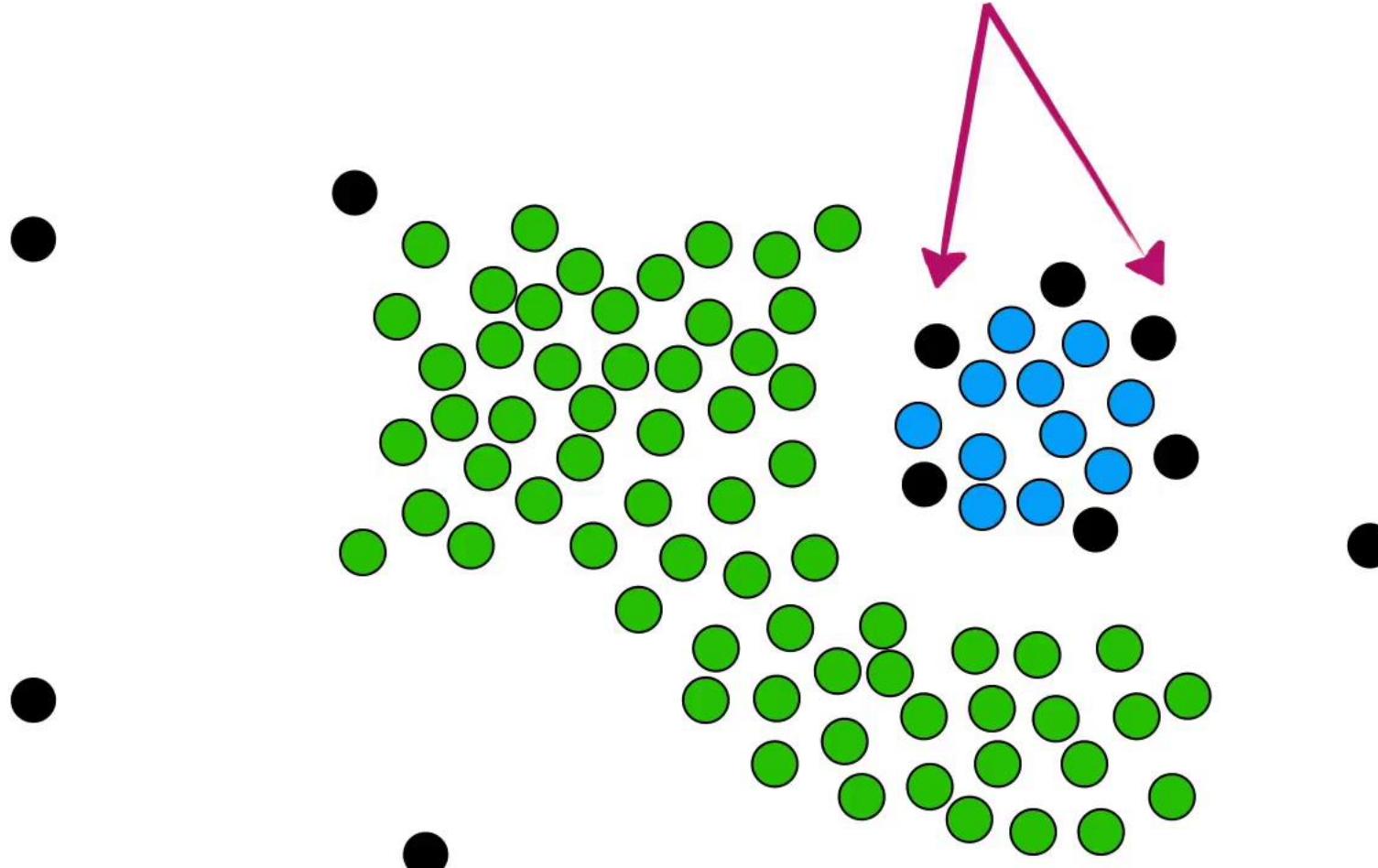


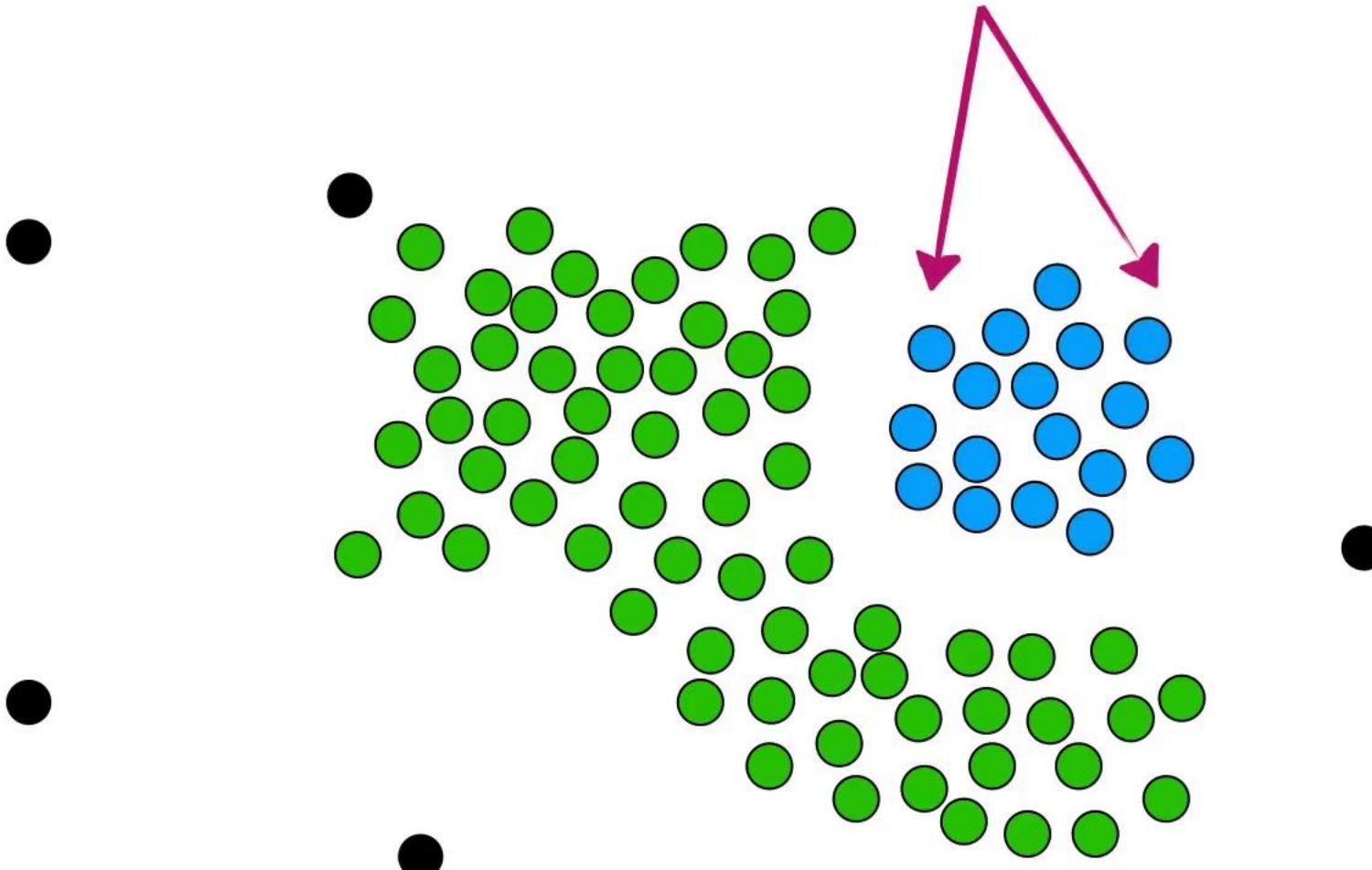


نطبق ما تعلمناه في First Cluster ولكن
بتصنیف جديد

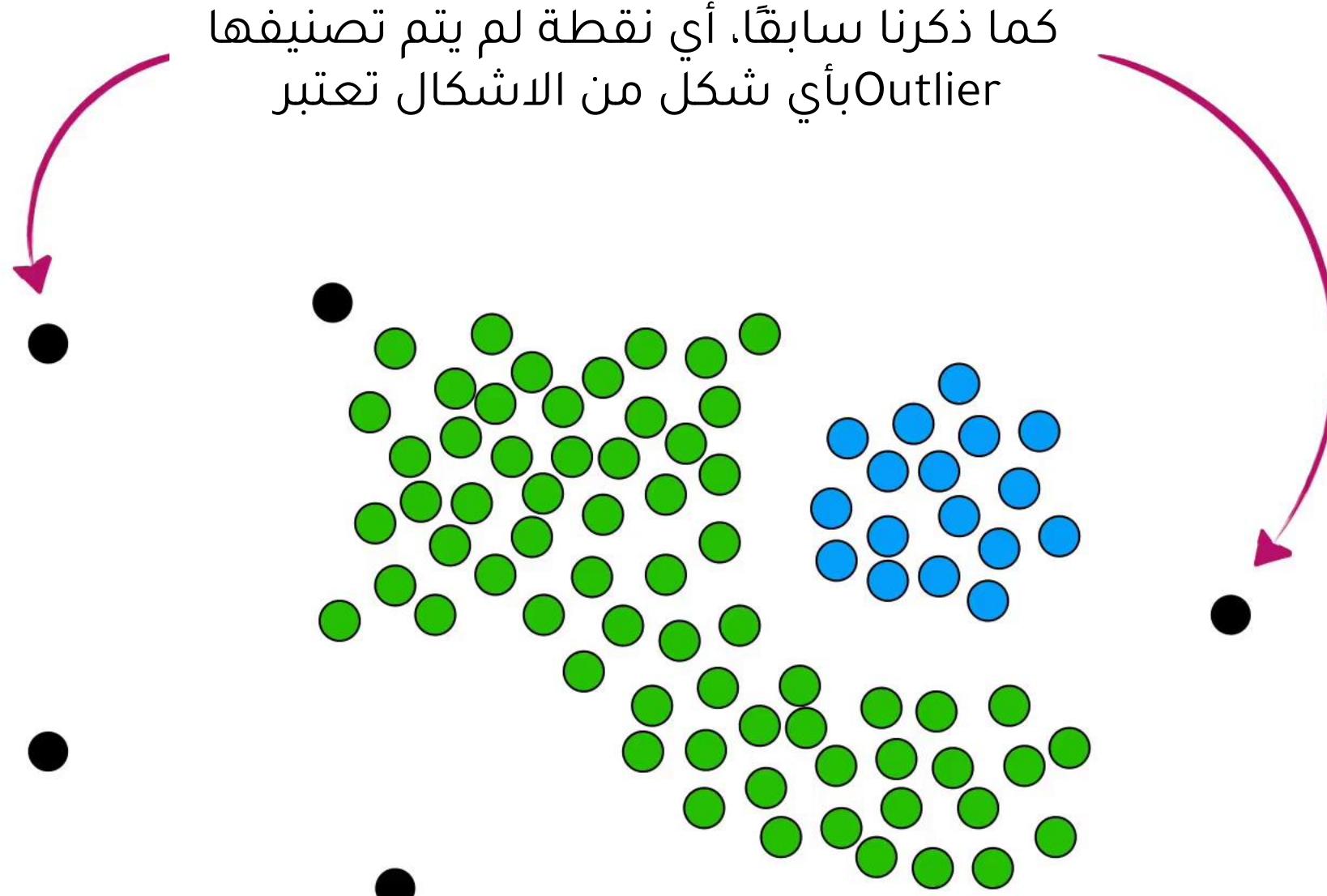






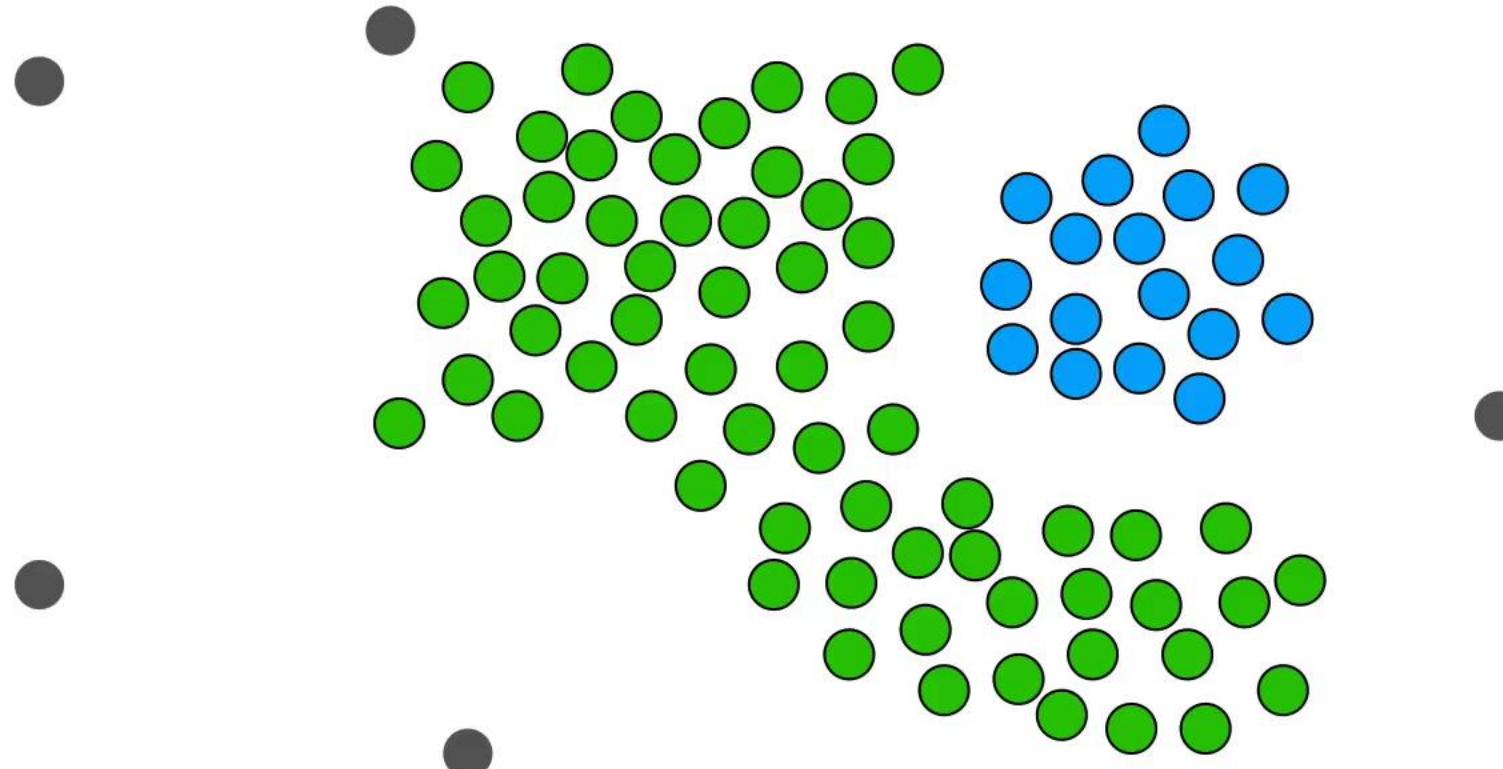


طريقة عمل خوارزمية DBScan





وهذا بشكل مبسط طريقة عمل الـ dbscan والشكل النهائي
Clustering للـ



خلونا نروح نشوف مثال على الخوارزمية





<https://quizizz.com/join?gc=draobhsaDevil=ecruos&601423>



<https://www.youtube.com/watch?v=4b5d3muPQmA>

<https://www.youtube.com/watch?v=RDZUdRSDOok>

<https://www.analyticsvidhya.com/blog/2021/03/everything-you-need-to-know-about-machine-learning/>

<https://developers.google.com/machine-learning/clustering/clustering-algorithms>