HUB IQ

Inteligencia Artificial & Machine Learning

Applicaciones en movilidad

Dr. Iván S. Razo Zapata

EngineeringX
Founded by the Royal Academy of Engineering and Lloyd's Register Foundation

Agrupamiento por Densidad

DBSCAN

Agrupamiento por densidad

- A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise - DBSCAN

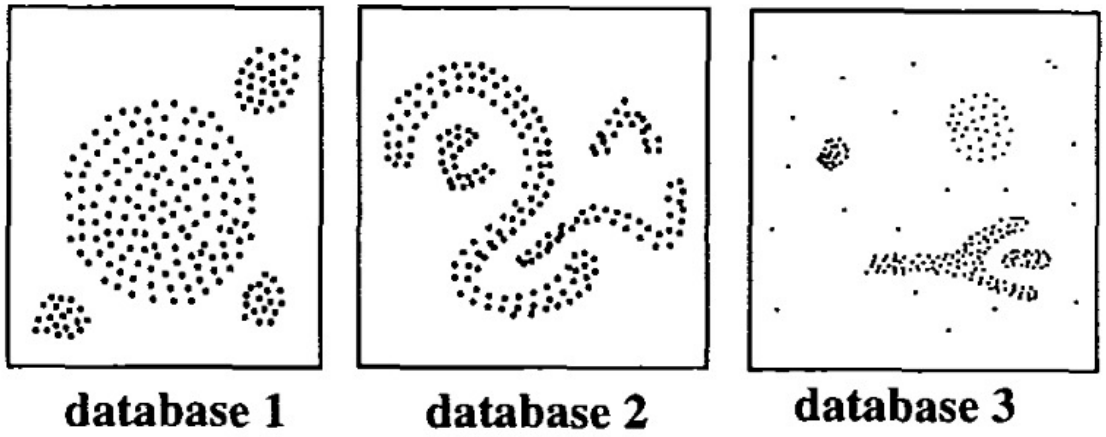


figure 1: Sample databases

A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise

Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu

Institute for Computer Science, University of Munich
Oettingenstr. 67, D-80538 München, Germany
{ester | kriegel | sander | xwxu}@informatik.uni-muenchen.de

Agrupamiento por densidad

- A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise - DBSCAN

2014 SIGKDD TEST OF TIME AWARD

Aug 18 2014 /

2014 SIGKDD Test of Time Award:

The SIGKDD Test of Time award recognizes outstanding papers from past KDD Conferences beyond the last decade that have had an important impact on the data mining research community.

The 2014 Test of Time award recognizes the following influential contributions to SIGKDD that have withstood the test of time:

A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise [KDD 1996]

A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise

Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu

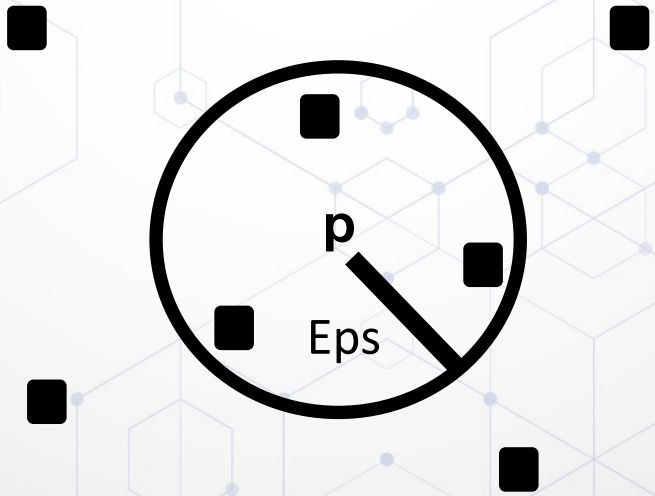
Institute for Computer Science, University of Munich
Oettingenstr. 67, D-80538 München, Germany
{ester | kriegel | sander | xwxu}@informatik.uni-muenchen.de

Idea general

Agrupamiento por densidad – Conceptos básicos

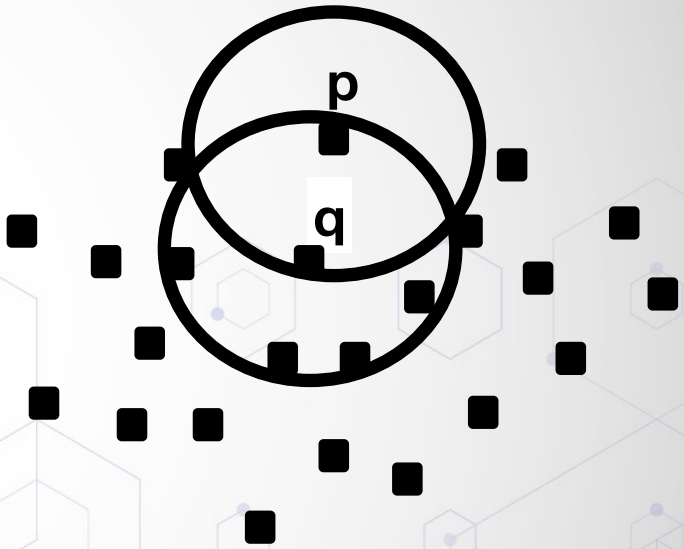
- Vecindad

Definition 1: (Eps-neighborhood of a point) The *Eps-neighborhood* of a point p, denoted by $N_{Eps}(p)$, is defined by $N_{Eps}(p) = \{q \in D \mid dist(p,q) \leq Eps\}$.



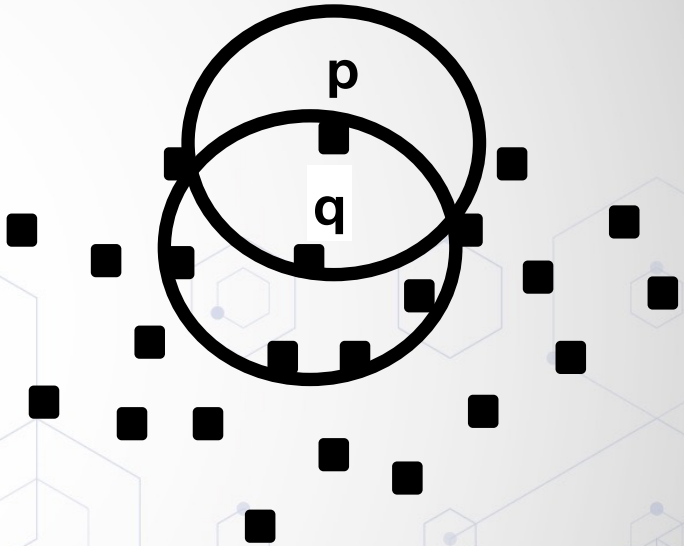
Core points and border points

- There are two kinds of points in a cluster:
- Points **inside of the cluster** (core points)
- Points **on the border** of the cluster (border points)



Densamente alcanzable de forma directa

- Si p es parte de la vecindad de q, y
- Si la vecindad de q contiene al menos MinPts … i.e. q es un core point

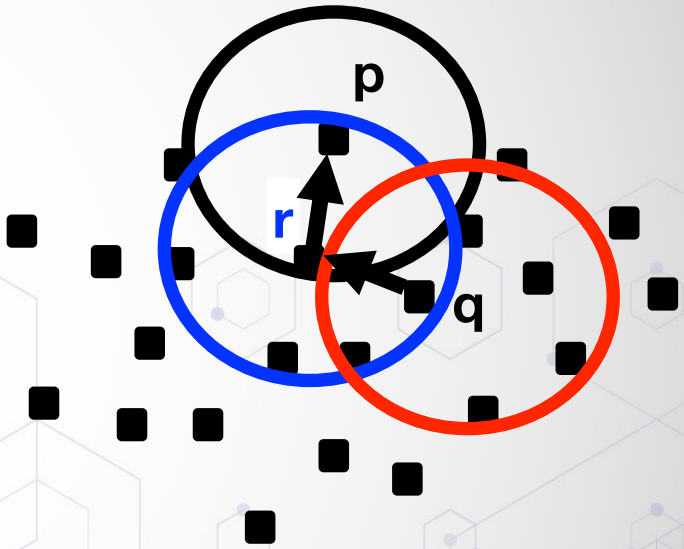


Definition 2: (directly density-reachable) A point p is *directly density-reachable* from a point q wrt. Eps, MinPts if

1) $p \in N_{Eps}(q)$ and
2) $|N_{Eps}(q)| \geq$ MinPts (core point condition).

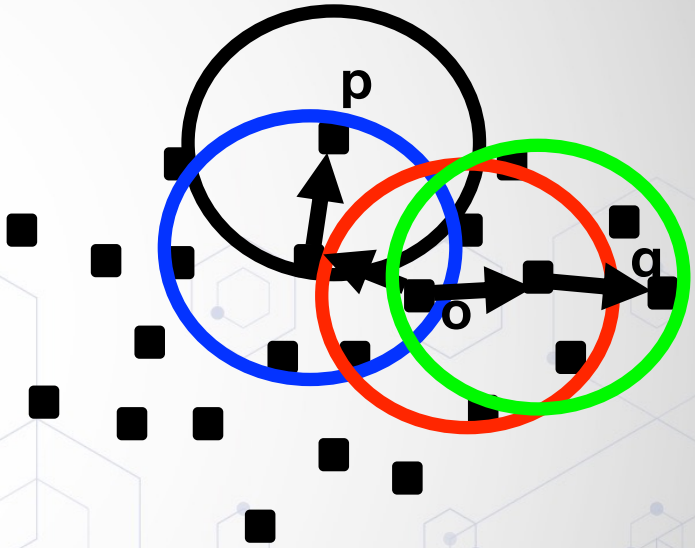
Densamente alcanzable

Definition 3: (density-reachable) A point p is *density-reachable* from a point q wrt. Eps and MinPts if there is a chain of points $p_1, ..., p_n, p_1 = q, p_n = p$ such that p_{i+1} is directly density-reachable from p_i.



Densamente conectado

Definition 4: (density-connected) A point p is *density-connected* to a point q wrt. Eps and MinPts if there is a point o such that both, p and q are density-reachable from o wrt. Eps and MinPts.



Agrupamiento por densidad

- Cluster
  - Puntos: Densamente alcanzables & Densamente conectados
- Ruido

Definition 5: (cluster) Let D be a database of points. A *cluster* C wrt. Eps and MinPts is a non-empty subset of D satisfying the following conditions:

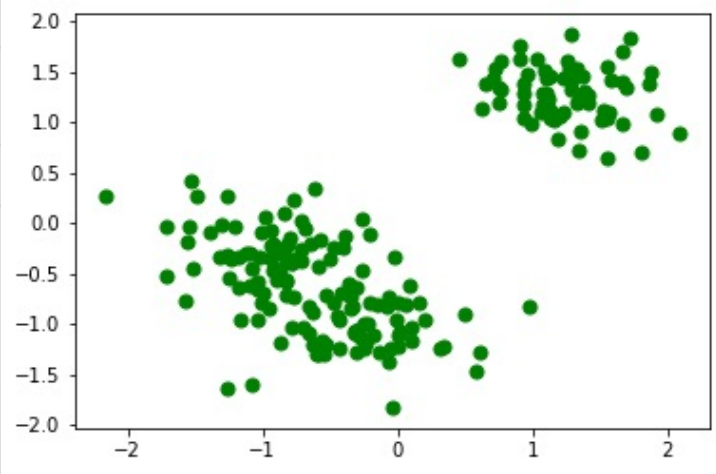
1) $\forall$ p, q: if p $\in$ C and q is density-reachable from p wrt. Eps and MinPts, then q $\in$ C. (Maximality)

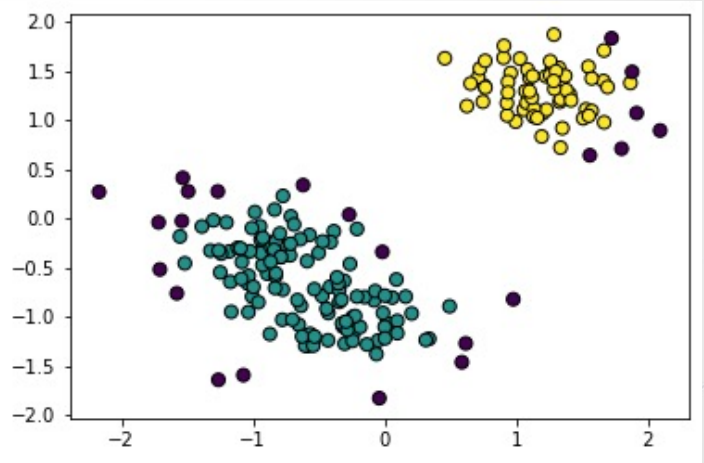
2) $\forall$ p, q $\in$ C: p is density-connected to q wrt. EPS and MinPts. (Connectivity)

Definition 6: (noise) Let $C_1, \ldots, C_k$ be the clusters of the database D wrt. parameters Eps_i and $MinPts_i$, i = 1, . . ., k. Then we define the *noise* as the set of points in the database D not belonging to any cluster C_i, i.e. noise = {p $\in D$ | $\forall$ i: p $\notin C_i$}.

Agrupamiento por densidad

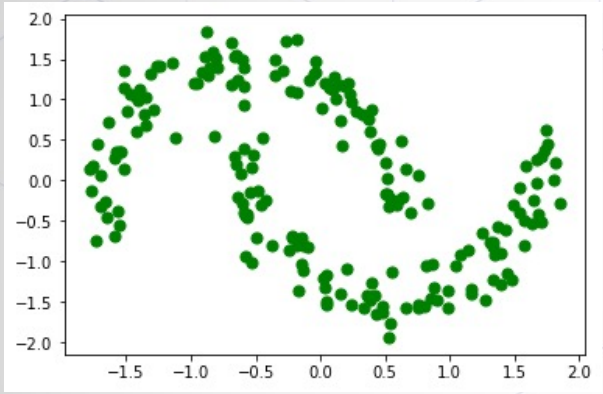
- Ejemplo



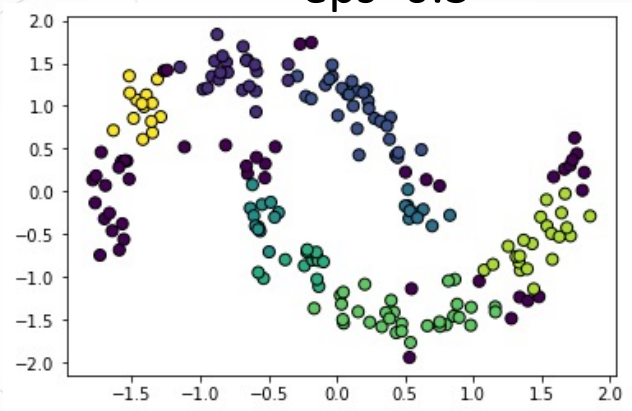


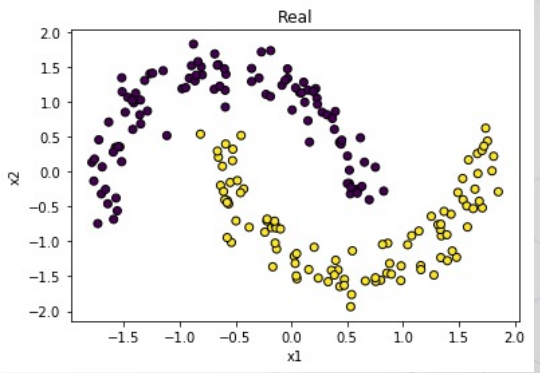
Agrupamiento por densidad

Ejemplo 2

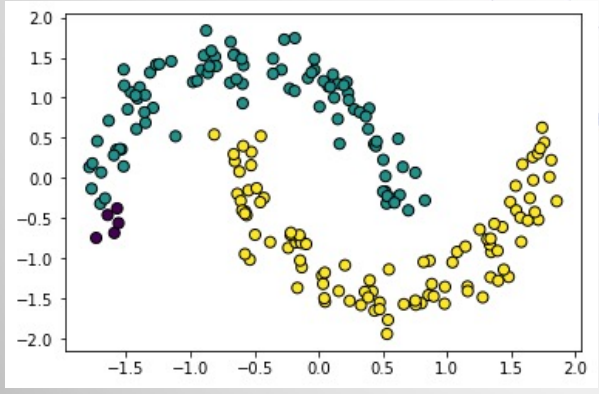


eps=0.3

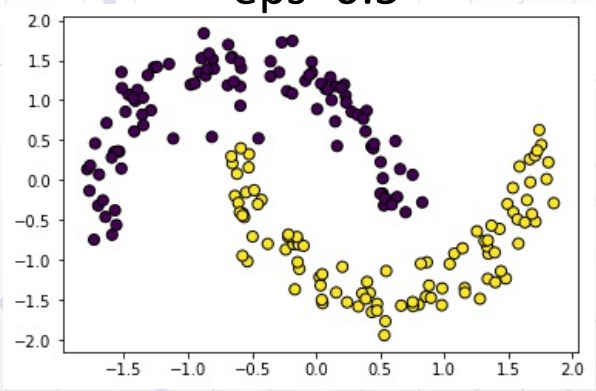




eps=0.4



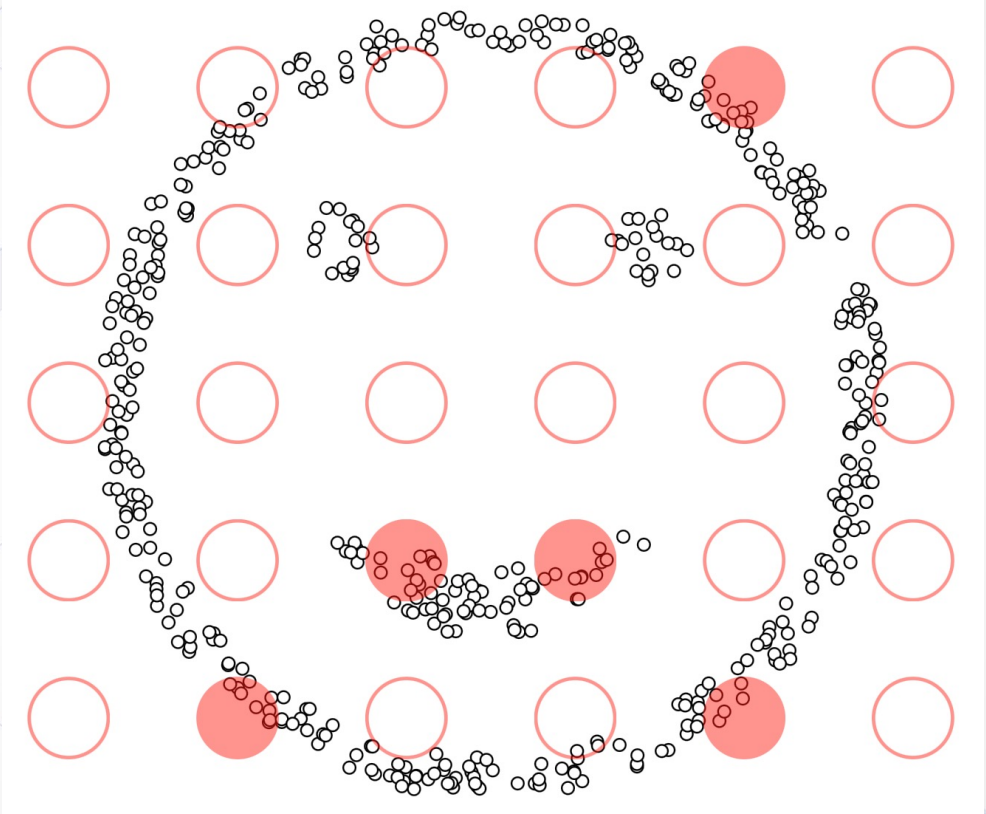
eps=0.5



Código

Agrupamiento por densidad

Ejemplo 3



https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/

EngineeringX

Founded by the Royal Academy of Engineering
and Lloyd's Register Foundation

GRACIAS

https://hubiq.mx/

HUBIQRO HUBIQ HUBIQRO